Nonperturbative Aspects of Strong Interactions

Uwe-Jens Wiese
Center for Theoretical Physics,
Laboratory for Nuclear Science, and Department of Physics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139, U.S.A.

March 12, 2009

Abstract

These lectures are an introduction to the theory of strong interactions. First we discuss the constituent quark model, in which the quarks are described as heavy quasiparticles. In this model some properties of strongly interacting particles (hadrons) can be understood based on the group theory of the flavor symmetry group. The parton model, in which hadrons are viewed as a collection of parallel moving free partons, is used to describe high energy scattering processes. The fundamental theory of strong interactions — quantum chromodynamics (QCD) — identifies the partons with quarks and gluons. QCD is an $SU(3)$ non-Abelian gauge theory with a coupling that becomes strong at low energies. This leads to permanent confinement of quarks and gluons within hadrons, and to the spontaneous breakdown of the chiral symmetries of the quarks. The Goldstone bosons of the spontaneously broken chiral symmetry are the pions. At low energies the chiral properties of QCD are described by a pion effective Lagrangian. The nontrivial topology of the gluon field gives rise to Euclidean instanton solutions, and leads to a solution of the axial $U(1)$ problem of the large η' meson mass. In the heavy quark limit QCD can again be described by an effective theory, and in the case of quarkonia even nonrelativistic potential models are adequate descriptions. The fully nonperturbative aspects of QCD can be explored by regularizing the theory on a space-time lattice. In the strong coupling limit lattice QCD explains confinement. At high temperatures confinement disappears and a hot quark-gluon plasma emerges from a deconfinement phase transitions. The implications of this phase transition in the early universe are discussed at the end of this course.

Contents

Chapter 1

Introduction

Based on our present understanding of Nature there are four fundamental interactions. All matter is subject to gravity, the weakest among the four. The weak interaction, which for example is responsible for the radioactive decay of atomic nuclei, is the next stronger, followed by the electromagnetic interaction, in which all charged particles as well as the photon participate. The strongest interaction known at present is responsible not only for nuclear binding, but also for the formation of the nuclear constituents (protons, neutrons and pions) themselves. All strongly interacting particles — the so-called hadrons — consist of quarks and gluons. The quarks are massive spin 1/2 fermions, while the gluons are massless spin 1 vector bosons. Quarks and gluons are the degrees of freedom of quantum chromodynamics (QCD) — the fundamental theory of strong interactions. Still, quarks and gluons are not particles in the usual sense because a single quark or gluon has never been observed despite numerous experimental efforts. In fact, it seems that the strong interactions are so strong that quarks and gluons are permanently confined within hadrons.

All we presently know about particle physics is summarized in the so-called Standard model, which QCD is part of. The Standard model is a relativistic quantum field theory based on the principles of special relativity and quantum mechanics. It is a gauge theory with a local $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ symmetry. Here $SU(3)_C$ is the color gauge group of QCD, while $SU(2)_L \otimes U(1)_Y$ is the symmetry that governs the electromagnetic

Generation	Quarks	Leptons
1.	u, d	e, ν_e
2.	c, s	μ, ν_μ
3.	t, b	τ, ν_τ

Table 1.1: *The three generations of quarks and leptons.*

and weak interactions. The Standard model describes the interactions of quarks and leptons via the exchange of vector bosons and scalar Higgs particles. There are three generations of quarks and leptons, which differ only in the particle masses but not in other quantum numbers. The generations are listed in table 1.1. The light quarks u, d and s (up, down and strange) have mass terms below the typical QCD energy scale $\Lambda_{QCD} \approx 0.2$ GeV, while the quarks c, b and t (charm, bottom and top) are much heavier. These quarks do not play a role in QCD at low energies. Still, they give rise to interesting heavy quark physics. The gauge bosons of the Standard model are eight gluons in the adjoint representation of $SU(3)_C$, as well as two massive charged W-bosons, a massive neutral Z-boson and the massless neutral photon in the adjoint representation of $SU(2)_L \otimes U(1)_Y$. Here we are interested in the strong interactions. Then we can approximately neglect the other much weaker interactions and restrict ourselves to QCD — the $SU(3)_C$ gauge theory of quarks and gluons. The leptons, photons, W, Z and Higgs-bosons do not participate in the strong interactions. Still, leptons can be used as electromagnetic or weak probes to investigate the complex interior of hadrons.

Since the interaction is very strong a perturbative solution of QCD is not possible at low energies. Nonperturbative methods also do not lead to an analytic understanding of the complex QCD dynamics. Therefore phenomenological models have been developed, which describe only certain aspects of the strong interactions, but are simpler to handle analytically. The simplest of these models is the constituent quark model, in which the hadrons are composed of a few constituent quarks. Baryons consist of three quarks, mesons consist of a quark-antiquark pair. The constituent quarks are massive, charged quasiparticles with spin 1/2 and flavor quantum numbers u, d and s, which cannot be directly identified with the quarks of QCD.

Instead the constituent quarks themselves should be viewed as complicated conglomerates of quarks, antiquarks and gluons. Even if the mass terms of the fundamental quarks would be zero, the constituent quarks would still have a mass $M_q \approx \Lambda_{QCD}$. The combination of flavor and spin quantum numbers of the quarks leads to a classification scheme for the hadrons that yields results in good agreement with experiments. To lowest order the baryons have a mass $3M_q$ and the meson mass is $2M_q$. The nonvanishing mass terms of the fundamental quarks give rise to mass splittings generally consistent with the observed ones.

While the constituent quark model describes static hadron properties rather well, it fails to describe dynamical processes at high energies as, for example, deep-inelastic lepton-hadron scattering. These processes are well understood in the parton model, in which hadrons consist of free, pointlike partons. In QCD the partons are identified with quarks and gluons. This is possible because QCD is asymptotically free, i.e. the interaction becomes weaker and weaker at higher and higher energies.

When the mass terms of the light u and d quarks are neglected, the QCD Lagrangian has a global $U(2)_L \otimes U(2)_R$ 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 $U(2)_{L=R}$ 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.

When a $U(2)_L \otimes U(2)_R$ symmetry breaks spontaneously to a $U(2)_{L=R}$ symmetry, the Goldstone theorem predicts four Goldstone bosons. Still, one observes only three light pions, while the flavor singlet η'-meson with a mass of 0.958 GeV is much too heavy to be a fourth Goldstone boson. This is the so-called axial $U(1)$-problem of QCD. It gets resolved when one studies the vacuum structure of a non-Abelian gauge theory in detail. Due to the nontrivial topology of the space of all gauge transformations

Euclidean instanton configurations exist. They describe tunneling processes between different classical vacua and give rise to a θ-vacuum structure. The instantons induce a quantum mechanical explicit symmetry breaking of the $U(2)_L \otimes U(2)_R$ symmetry to an $SU(2)_L \otimes SU(2)_R \otimes U(1)_{L=R}$ symmetry. The axial $U(1)$ symmetry is explicitly broken by the Adler-Bell-Jackiw anomaly and is therefore no symmetry of QCD. This explains why there is no fourth Goldstone Boson.

QCD simplifies not only in the limit of light quarks but also for heavy quarks, especially for mesons that contain a heavy c or b quark. The strongly interacting cloud of light quarks and gluons that surrounds the heavy quark is blind to the heavy quark's spin and flavor. This leads to an $SU(4)$ symmetry and to heavy quark symmetry relations that agree well with experiments. If both quarks of a meson are heavy — as it is the case for the quarkonia J/Ψ and Υ — one can use simple nonrelativistic potential models to arrive at quantitative predictions. In these models confinement is put in by hand as a linear rising potential.

Unfortunately, not all aspects of strong interactions can be understood in terms of symmetries or simple models. In fact, QCD confronts us with a highly complicated nonperturbative dynamical problem, and it is a great challenge in theoretical physics to solve this problem. Ultimately, we will only be convinced that QCD describes the strong interactions correctly when we are able to derive from it the hadron spectrum, the confinement potential for heavy quarks, hadronic matrix elements and other low energy quantities, and if these quantities agree with experimental observations. At present, lattice QCD is the most promising method for attacking these difficult questions from first principles. Lattice QCD should be viewed not as an approximation to QCD, but as a nonperturbative definition of the QCD Feynman path integral. The path integral is regularized by defining quark and gluon fields on a Euclidean space-time lattice. The system is then analogous to one of classical statistical mechanics, and the continuum limit of vanishing lattice spacing corresponds to an approach to a second order phase transition. Away from the phase transition a strong coupling expansion leads to confinement with a linear quark-antiquark potential. Closer to the continuum limit numerical simulations also show confinement. The representation of the quark's chiral symmetries is problematical on the lattice. In fact, the fermion species doubling problem can be traced back to the Adler-Bell-Jackiw anomaly. The numerical simulation of dynamical

quarks is very tedious. Still, present day lattice QCD calculations agree already quite well with experimental results.

At high temperatures interesting effects arise in the strong interactions. Due to asymptotic freedom quarks and gluons behave as free particles at high energies. Hence confinement gets lost and chiral symmetry gets restored at high temperatures. This corresponds to a phase transition from ordinary hadronic matter to a high temperature quark-gluon plasma. In heavy ion collisions one tries to create a hot quark-gluon bubble for a short amount of time. Also in the early universe the quark gluon plasma must have existed for the first 10^{-5} seconds after the big bang. At that time the temperature was around Λ_{QCD}. Then a phase transition occurred to today's low temperature phase, which may have influenced the nucleosynthesis of light atomic nuclei during the first few minutes after the bang. In a universe with very massive quarks the phase transition would be first order, and with two massless quarks it would be second order. In our universe with two light and one intermediate mass quark it is still unclear if the phase transition is first or second order or if there is merely a crossover.

Chapter 2

The Constituent Quark Model

2.1 Isospin Symmetry

The history of strong interactions began in the thirties of this century with the measurement of the magnetic moment of the proton. The Dirac theory had very successfully predicted the magnetic moment of the electron

$$\vec{\mu}_e = -\frac{e}{2m_e}\vec{\sigma}. \tag{2.1.1}$$

Here m_e and e are mass and charge of the electron and $\vec{\sigma}$ is a vector of spin 1/2 Pauli matrices. (Throughout these lectures we will use units in which $c = \hbar = 1$.) Analogously, for the proton one has

$$\vec{\mu}_p = g_p\frac{e}{2M_p}\vec{\sigma}, \tag{2.1.2}$$

where M_p is the proton mass. For an elementary spin 1/2 particle one would expect $g_p = 1$, while the measured value is

$$g_p = 2.793 \tag{2.1.3}$$

One can view this result as a first indication that the proton has some complicated inner structure, while the electron appears structureless. About the same time Chadwick discovered the neutron. Proton and neutron have almost the same masses

$$M_p = 0.938 \text{ GeV}, \; M_n = 0.940 \text{ GeV}. \tag{2.1.4}$$

While the proton seems to be absolutely stable, a free neutron decays radioactively into a proton, an electron and an electron-antineutrino $n \rightarrow p + e + \bar{\nu}_e$. Also the neutron has a magnetic moment, which is experimentally given by

$$g_n = -1.913, \tag{2.1.5}$$

indicating that it also has some internal degrees of freedom. Protons and neutrons (the nucleons) are the constituents of atomic nuclei. In the mid thirties Yukawa postulated a light particle mediating the interaction between protons and neutrons. This π-meson or pion was discovered at the end of the forties. Pions are bosons with spin 0, which exist in three charge states π^+, π^0 and π^-. The corresponding masses are

$$M_{\pi^+} = M_{\pi^-} = 0.140 \text{ GeV}, \; M_{\pi^0} = 0.135 \text{ GeV}. \tag{2.1.6}$$

In pion-nucleon scattering a resonance occurs in the total cross section as a function of the pion-nucleon center of mass energy $\sqrt{s}$. This is illustrated in fig.2.1. The resonance energy is interpreted as the mass of an unstable

Figure 2.1: *Resonance in the total cross section of pion-proton scattering as a function of the center of mass energy.*

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 2.1: *The isospin classification of hadrons.*

particle — the so-called Δ-isobar. One may 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

$$M_{\Delta^{++}} \approx M_{\Delta^{+}} \approx M_{\Delta^{0}} \approx M_{\Delta^{-}} \approx 1.232 \text{ GeV} \tag{2.1.7}$$

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

$$M_{\rho^{+}} \approx M_{\rho^{0}} \approx M_{\rho^{-}} \approx 0.768 \text{ GeV}. \tag{2.1.8}$$

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 $S = 0, 1/2, 1, 3/2, ...$ is associated with an irreducible representation of the $SU(2)_S$ rotation group containing $2S + 1$ states distinguished by their spin projection

$$S_z = -S, -S + 1, ..., S - 1, S. \tag{2.1.9}$$

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

$$I_3 = -I, -I + 1, ..., I - 1, I. \tag{2.1.10}$$

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 2.1. 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 $p + \pi \to \Delta$ is consistent with isospin symmetry because the coupling of the isospin representations of nucleon and pion

$$\{2\} \otimes \{3\} = \{2\} \oplus \{4\} \tag{2.1.11}$$

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. Later we will introduce a generalized charge — the so-called color charge — that is indeed responsible for the strong interactions. It turns out that all hadrons are color neutral.

2.2 up and down Quarks

The anomalous magnetic moments of proton and neutron indicate that these particles have a complicated internal structure. Also the fact that the Δ-resonance can be viewed as an excited state of the proton points to the substructure of hadrons. 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 u ($I_3 = 1/2$) and d ($I_3 = -1/2$). Since the Δ-isobar has isospin 3/2 it contains at least three constituent quarks. We couple

$$\{2\} \otimes \{2\} \otimes \{2\} = (\{1\} \oplus \{3\}) \otimes \{2\} = \{2\} \oplus \{2\} \oplus \{4\}, \tag{2.2.1}$$

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

$$Q = I_3 + \frac{1}{2} = \sum_{q=1}^{3} (I_{3q} + \frac{1}{6}) = \sum_{q=1}^{3} Q_q, \tag{2.2.2}$$

and hence we obtain for the charges of the quarks

$$Q_q = I_{3q} + \frac{1}{6},\ Q_u = \frac{1}{2} + \frac{1}{6} = \frac{2}{3},\ Q_d = -\frac{1}{2} + \frac{1}{6} = -\frac{1}{3}. \tag{2.2.3}$$

The quarks have fractional electric charges. Using Clebsch-Gordon coefficients of $SU(2)$ one finds

$$\begin{aligned}
&\boxed{1}\boxed{2}\boxed{3}_{\,3/2} = uuu \equiv \Delta^{++},\\
&\boxed{1}\boxed{2}\boxed{3}_{\,1/2} = \frac{1}{\sqrt{3}}(uud + udu + duu) \equiv \Delta^{+},\\
&\boxed{1}\boxed{2}\boxed{3}_{\,-1/2} = \frac{1}{\sqrt{3}}(udd + dud + ddu) \equiv \Delta^{0},\\
&\boxed{1}\boxed{2}\boxed{3}_{\,-3/2} = ddd \equiv \Delta^{-}.
\end{aligned} \tag{2.2.4}$$

These isospin states are completely symmetric against permutations of the constituent quarks. We have used a Young tableaux notation that we will discuss in more detail in the next section.

2.3 The Permutation Group S_N

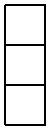
Let us consider the permutation symmetry of N objects — in our case the constituent quarks. Their permutations form the group S_N. The permutation group has $N!$ elements — all permutations of N objects. The group S_2 has two elements: the identity and the pair permutation. The representations of S_2 are represented by Young tableaux

$$\begin{aligned}
&\boxed{}\boxed{} \quad \text{1-dimensional symmetric representation,}\\
&\begin{array}{|c|}\hline\\\hline\\\hline\end{array} \quad \text{1-dimensional antisymmetric representation.}
\end{aligned} \tag{2.3.1}$$

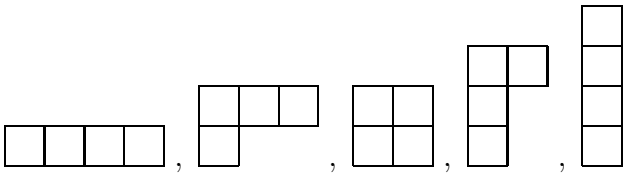
To describe the permutation properties of three objects — for example the three constituent quarks of the Δ-isobar — we need the group S_3. It has $3! = 6$ elements: the identity, 3 pair permutations and 2 cyclic permutations. The group S_3 has three irreducible representations

$$\boxed{}\boxed{}\boxed{} \quad \text{1-dimensional symmetric representation,}$$

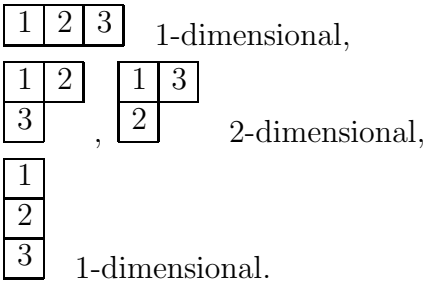

2-dimensional representation of mixed symmetry,


1-dimensional antisymmetric representation. (2.3.2)

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


(2.3.3)

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. (2.3.4)

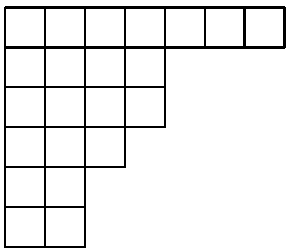
The squares of the dimensions of all representations add up to the order of the group, i.e.

$$\sum_{\Gamma} d_{\Gamma}^2 = N! \ . \tag{2.3.5}$$

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

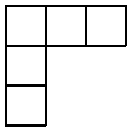


(2.3.6)

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

$$d_{m_1,m_2,\ldots,m_n} = N! \frac{\prod_{i<k}(l_i - l_k)}{l_1! l_2! \ldots l_n!}, \; l_i = m_i + n - i. \tag{2.3.7}$$

Applying this formula to the following Young tableau from S_5



(2.3.8)

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

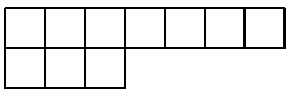
$$d_{3,1,1} = 5! \frac{(l_1 - l_2)(l_1 - l_3)(l_2 - l_3)}{l_1! l_2! l_3!} = 5! \frac{3 \cdot 4 \cdot 1}{5!2!1!} = 6. \tag{2.3.9}$$

It is a good exercise to construct all representations of S_4 and S_5, and determine their dimensions by the enumeration method of eq.2.3.4 and independently from eq.2.3.7. One can check the results using eq.2.3.5.

In our case the permuted objects are the constituent quarks of the hadrons, i.e. fundamental representations of the isospin group. Hence we identify

 $= \{2\}.$ (2.3.10)

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.



(2.3.11)

has $m_1 = 7$ and $m_2 = 3$. The corresponding $SU(2)$ representation has

$$I = \frac{1}{2}(m_1 - m_2), \tag{2.3.12}$$

which is also denoted by $\{m_1 - m_2 + 1\}$. The above Young tableau hence represents $I = 2$ — an isospin quintet $\{5\}$. Young tableaux with more than two rows have no realization in $SU(2)$ since among just two distinguishable objects (in our case u and d) no more than two can be combined antisymmetrically.

2.4 Nucleon and Δ-Isobar in the Constituent Quark Model

After this short excursion to the group theory of the permutation group let us come back to the three constituent quarks of the Δ-isobar. We now write the coupling of the three quarks as

$$\boxed{1} \otimes \boxed{2} \otimes \boxed{3} = \boxed{1}\boxed{2}\boxed{3} \oplus \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array} \oplus \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array} \oplus \begin{array}{|c|}\hline 1 \\ \hline 2 \\ \hline 3 \\ \cline{1-1}\end{array}. \tag{2.4.1}$$

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

$$\{2\} \otimes \{2\} \otimes \{2\} = \{4\} \oplus \{2\} \oplus \{2\} \oplus \{0\}. \tag{2.4.2}$$

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 $\boxed{1}\boxed{2}\boxed{3}_{I_3}$.

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

$$\begin{aligned}
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}\,_{3/2} &= \uparrow\uparrow\uparrow,\\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}\,_{1/2} &= \frac{1}{\sqrt{3}}(\uparrow\uparrow\downarrow \;+\; \uparrow\downarrow\uparrow \;+\; \downarrow\uparrow\uparrow),\\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}\,_{-1/2} &= \frac{1}{\sqrt{3}}(\uparrow\downarrow\downarrow \;+\; \downarrow\uparrow\downarrow \;+\; \downarrow\downarrow\uparrow),\\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}\,_{-3/2} &= \downarrow\downarrow\downarrow\,. \qquad (2.4.3)
\end{aligned}$$

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

$$|\Delta I_3 S_z\rangle = \begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}\,_{I_3}\;\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}\,_{S_z}. \qquad (2.4.4)$$

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 $S_z = 1/2$

$$\begin{aligned}
|\Delta\tfrac{1}{2}\tfrac{1}{2}\rangle \;&=\; \frac{1}{3}(u\uparrow u\uparrow d\downarrow + u\uparrow u\downarrow d\uparrow + u\downarrow u\uparrow d\uparrow\\
&+\; u\uparrow d\uparrow u\downarrow + u\uparrow d\downarrow u\uparrow + u\downarrow d\uparrow u\uparrow\\
&+\; d\uparrow u\uparrow u\downarrow + d\uparrow u\downarrow u\uparrow + d\downarrow u\uparrow u\uparrow). \qquad (2.4.5)
\end{aligned}$$

One sees explicitly that this state is totally symmetric.

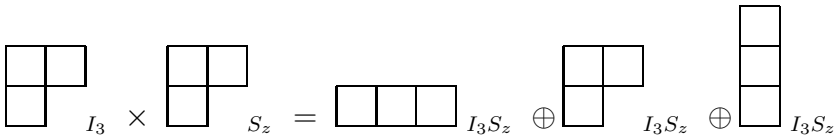
As we have seen, the Young tableau $\begin{array}{|c|c|}\hline\multicolumn{1}{|c}{}&\multicolumn{1}{c|}{}\\\hline\multicolumn{1}{|c}{}&\multicolumn{1}{c|}{}\\\cline{1-1}\end{array}$ is associated with the isodoublet $\{2\}$. Hence, it is natural to expect that the nucleon state can be constructed from 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

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

$$\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{1/2} = \frac{1}{\sqrt{2}}(udu - duu),$$

$$\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{-1/2} = \frac{1}{\sqrt{2}}(udd - dud). \qquad (2.4.6)$$

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



$$\qquad (2.4.7)$$

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

$$|NI_3S_z\rangle = \frac{1}{\sqrt{2}}\Big(\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{I_3}\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{S_z} + \begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{I_3}\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{S_z}\Big). \qquad (2.4.8)$$

Now we want to compute the magnetic moments of proton and neutron, assuming that the quarks are elementary fermions, i.e. that they have no anomalous magnetic moment

$$\vec{\mu}_q = \frac{Q_q}{2M_q}\vec{\sigma}_q,\ \vec{\mu}_B = \sum_{q=1}^{3}\frac{Q_q}{2M_q}\vec{\sigma}_q. \qquad (2.4.9)$$

The magnetic moment of the baryons is given by the following matrix element

$$\begin{aligned}\mu_{B_{I_3}} &= \langle BI_3S|\mu_{B_3}|BI_3S\rangle \\ &= \langle BI_3S|\frac{1}{M_q}\sum_{q=1}^{3}(I_{3q}+\frac{1}{6})S_{zq}|BI_3S\rangle\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{M_q}\langle BI_3S|\sum_{q=1}^{3} I_{3q}S_{zq}|BI_3S\rangle + \frac{1}{M_q}\frac{1}{6}S \\
&= \frac{1}{M_q}3\langle BI_3S|I_{33}S_{z3}|BI_3S\rangle + \frac{1}{M_q}\frac{1}{6}S. \qquad (2.4.10)
\end{aligned}
$$

One finds

$$
\begin{aligned}
I_{33}\,\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{1/2} &= \frac{1}{2}\Big(-\frac{2\sqrt{2}}{3}\,\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{1/2} - \frac{1}{3}\,\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{1/2}\Big), \\
I_{33}\,\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{-1/2} &= \frac{1}{2}\Big(-\frac{2\sqrt{2}}{3}\,\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{-1/2} + \frac{1}{3}\,\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{-1/2}\Big), \\
I_{33}\,\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{1/2} &= \frac{1}{2}\,\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{1/2}, \\
I_{33}\,\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{-1/2} &= -\frac{1}{2}\,\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{-1/2}. \qquad (2.4.11)
\end{aligned}
$$

Analogous relations hold for the spin states. For the proton we obtain

$$
\begin{aligned}
&\langle N\tfrac{1}{2}\tfrac{1}{2}|I_{33}S_{z3}|N\tfrac{1}{2}\tfrac{1}{2}\rangle \\
&= \frac{1}{\sqrt{2}}\Big(\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{1/2}\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{1/2} + \begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{1/2}\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{1/2}\Big)I_{33}S_{z3} \\
&\times\frac{1}{\sqrt{2}}\Big(\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{1/2}\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{1/2} + \begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{1/2}\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{1/2}\Big) \\
&= \frac{1}{2}\frac{1}{2}\frac{1}{2}((-\frac{1}{3})(-\frac{1}{3}) + 1\cdot 1) = \frac{5}{36}. \qquad (2.4.12)
\end{aligned}
$$

Similarly, for the neutron one finds

$$
\begin{aligned}
&\langle N - \tfrac{1}{2}\tfrac{1}{2}|I_{33}S_{z3}|N - \tfrac{1}{2}\tfrac{1}{2}\rangle \\
&= \frac{1}{\sqrt{2}}\Big(\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{-1/2}\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{1/2} + \begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{1/2}\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{-1/2}\Big)I_{33}S_{z3}
\end{aligned}
$$

$$\times\frac{1}{\sqrt{2}}(\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{\,-1/2}\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array}_{\,1/2}+\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{\,-1/2}\begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array}_{\,1/2})$$
$$=\frac{1}{2}\frac{1}{2}\frac{1}{2}(\frac{1}{3}(-\frac{1}{3})+(-1)1)=-\frac{5}{36}, \tag{2.4.13}$$

and hence

$$\begin{aligned}\mu_p &= \frac{1}{M_q}(3\frac{5}{36}+\frac{1}{12})=\frac{1}{M_q}\frac{1}{2},\\ \mu_n &= \frac{1}{M_q}(-3\frac{5}{36}+\frac{1}{12})=-\frac{1}{M_q}\frac{1}{3}.\end{aligned} \tag{2.4.14}$$

On the other hand we have

$$\begin{aligned}\mu_p &= \langle N\frac{1}{2}\frac{1}{2}|\mu_{Pz}|N\frac{1}{2}\frac{1}{2}\rangle=\langle N\frac{1}{2}\frac{1}{2}|\frac{g_p}{M_N}S_z|N\frac{1}{2}\frac{1}{2}\rangle=\frac{g_p}{2M_N},\\ \mu_n &= \langle N-\frac{1}{2}\frac{1}{2}|\mu_{Pz}|N-\frac{1}{2}\frac{1}{2}\rangle\\ &= \langle N-\frac{1}{2}\frac{1}{2}|\frac{g_n}{M_N}S_z|N-\frac{1}{2}\frac{1}{2}\rangle=\frac{g_n}{2M_N}.\end{aligned} \tag{2.4.15}$$

Until now we have not fixed the constituent quark mass M_q. We do this by requiring

$$\mu_p=\frac{1}{M_q}\frac{1}{2}=\frac{g_p}{2M_N}, \tag{2.4.16}$$

such that

$$M_q=\frac{M_N}{g_p}=\frac{0.939\text{GeV}}{2.793}=0.336\text{GeV}. \tag{2.4.17}$$

It is consistent that the constituent quark mass is roughly a third of the nucleon mass. Having fixed the only free parameter, we can now predict the magnetic moment of the neutron

$$\mu_n=-\frac{1}{M_q}\frac{1}{3}=-\frac{g_p}{3M_N}, \tag{2.4.18}$$

and hence

$$g_n=-\frac{2}{3}g_p=-1.86. \tag{2.4.19}$$

This is in good agreement with the experimental value $g_n=-1.91$. Probably the result is even better than the constituent quark model itself. We

should not forget, that we have drastically oversimplified the complex dynamics of the strong interactions. It is a good exercise to compute the magnetic moments of the four Δ-isobar states as well.

In our calculations of magnetic moments we have only considered the spin, but not the orbital angular momentum of the constituent quarks. Hence, we have implicitly assumed that the orbital angular momentum 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 $\square\square\square$. 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 the 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 we postulate a new symmetry of the quarks, that allows to obtain a totally antisymmetric wave function. This new symmetry cannot be $SU(2)$ again, because $\begin{array}{|c|}\hline\ \\\hline\ \\\hline\ \\\hline\end{array}$ is an empty representation in $SU(2)$. This is not the case for the group $SU(3)$, where $\begin{array}{|c|}\hline\ \\\hline\ \\\hline\ \\\hline\end{array}$ corresponds to a singlet representation. The new $SU(3)_C$ symmetry is the color group of QCD. Each quark appears in three 'colors' (sometimes labelled r (red), g (green) and b(blue)). Since we have three colors we can now completely antisymmetrize three quarks

$$\begin{array}{|c|}\hline\ \\\hline\ \\\hline\ \\\hline\end{array} = \frac{1}{\sqrt{6}}(rgb - rbg + gbr - grb + brg - bgr). \tag{2.4.20}$$

As we will see later, 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 — the gauge symmetry of QCD.

2.4.1 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

$$\begin{aligned} &UU^+ = U^+U = 1,\ \mathrm{det}U = 1. \\ &\mathrm{det}UV = \mathrm{det}U\mathrm{det}V = 1. \end{aligned} \tag{2.4.21}$$

Associativity ($(UV)W = U(VW)$) holds for all matrices, a unit element 1 exists (the unit matrix), the inverse is $U^{-1} = U^+$, and finally the group property

$$(UV)^+UV = V^+U^+UV = 1,\ UV(UV)^+ = UVV^+U^+ = 1 \tag{2.4.22}$$

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

$$U = \exp(iH), \tag{2.4.23}$$

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

$$H = \alpha_i\eta_i,\ \alpha_i \in R. \tag{2.4.24}$$

The structure of the algebra results from the commutation relations

$$[\eta_i, \eta_j] = 2ic_{ijk}\eta_k, \tag{2.4.25}$$

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 $SU(2)_I$ and r, g and b 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	$n+1$			
$n-3$	$n-2$	$n-1$				
$n-4$	$n-3$					
$n-5$	$n-4$					

. (2.4.26)

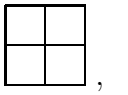
The dimension of the $SU(n)$ representation is given as the product of all factors divided by $N!$ and multiplied with the S_N dimension $d_{m1,m_2,\ldots,m_n}$ of the Young tableau

$$\begin{aligned} D^n_{m1,m_2,\ldots,m_n} &= \frac{(n+m_1-1)!}{(n-1)!}\frac{(n+m_2-2)!}{(n-2)!}\ldots\frac{m_n!}{0!}\frac{1}{N!}N!\frac{\prod_{i<k}(l_i-l_k)}{l_1!l_2!\ldots l_n!} \\ &= \frac{\prod_{i<k}(m_i-m_k-i+k)}{(n-1)!(n-2)!\ldots 0!}. \end{aligned} \tag{2.4.27}$$

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

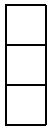
$$D^2_{m_1,m_2} = \frac{m_1-m_2-1+2}{1!0!} = m_1 - m_2 + 1 = q_1 + 1 \tag{2.4.28}$$

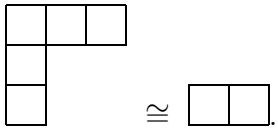
in agreement with our previous result. For a rectangular Young tableau with n rows, e.g. in $SU(2)$ for


, (2.4.29)

all $q_i = 0$, and we obtain

$$D^n_{m,m,\ldots,m} = \frac{\prod_{i<k}(m_i - m_k - i + k)}{(n-1)!(n-2)!\ldots 0!} = \frac{(n-1)!(n-2)!\ldots 0!}{(n-1)!(n-2)!\ldots 0!} = 1, \quad (2.4.30)$$

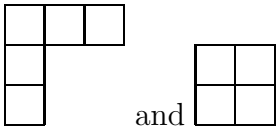
and therefore a singlet. This confirms that in $SU(3)$  corresponds to a singlet. It also explains why the dimension of an $SU(n)$ representation depends only on the differences q_i. Without changing the dimension we can couple a representation with a singlet, and hence we can always add a rectangular Young tableau with n rows to any $SU(n)$ representation. For example in $SU(3)$

 (2.4.31)

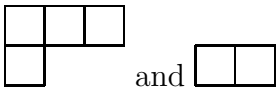
We want to associate an anti-representation with each representation by replacing m_i and q_i with

$$\bar{m}_i = m_1 - m_{n-i+1}, \ \bar{q}_i = \bar{m}_i - \bar{m}_{i+1} = m_{n-i} - m_{n-i+1} = q_{n-i}. \quad (2.4.32)$$

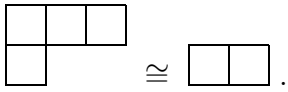
Geometrically the Young tableau of a representation and its anti-representation (after rotation) fit together to form a rectangular Young tableau with n rows. For example, in $SU(3)$

 (2.4.33)

are anti-representations of one another. In $SU(2)$ each representation is its own anti-representation. For example

 (2.4.34)

are anti-representations of one another, but

 (2.4.35)

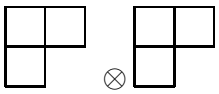
This is not the case for higher n. The dimension of a representation and its anti-representation are identical

$$D^n_{\bar{m}_1,\bar{m}_2,\ldots,\bar{m}_n} = D^n_{m_1,m_2,\ldots,m_n}. \tag{2.4.36}$$

It is an instructive exercise to prove this. 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. Again it is instructive to show that the dimension of the adjoint representation is

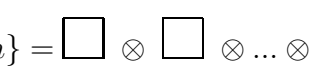
$$D^n_{2,1,1,\ldots,1,0} = n^2 - 1. \tag{2.4.37}$$

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 labelled with 'a', the boxes of the second row with 'b', etc. Then the boxes labelled with 'a' are added to the first Young tableau in all possible ways that lead to new allowed Young tableau. 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. It is instructive to couple

 (2.4.38)

in this way. As a check one can translate the result into $SU(2)$ and $SU(3)$ language and confirm that the product of the dimensions of the coupled representations agrees with the sum of dimensions of the representations in the reduction.

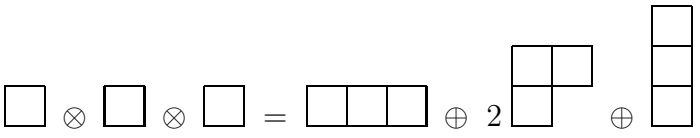
We now want to couple N fundamental representations of $SU(n)$. In Young tableau language this reads

$\{n\} \otimes \{n\} \otimes \ldots \otimes \{n\} =$  . (2.4.39)

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. $d_{m_1,m_2,...,m_n}$ times. Hence we can write

$$\{n\} \otimes \{n\} \otimes ... \otimes \{n\} = \sum_{\Gamma} d_{m_1,m_2,...,m_n} \{D^n_{m_1,m_2,...,m_n}\}. \tag{2.4.40}$$

The sum goes over all Young tableaux with N boxes. For example



(2.4.41)

Translated into $SU(n)$ language this reads

$$\begin{aligned} \{n\} \otimes \{n\} \otimes \{n\} &= \{\frac{n(n+1)(n+2)}{6}\} \oplus 2\{\frac{(n-1)n(n+1)}{3}\} \\ &\oplus \{\frac{(n-2)(n-1)n}{6}\}. \end{aligned} \tag{2.4.42}$$

The dimensions test

$$\frac{n(n+1)(n+2)}{6} + 2\frac{(n-1)n(n+1)}{3} + \frac{(n-2)(n-1)n}{6} = n^3 \tag{2.4.43}$$

confirms this result. In $SU(2)$ this corresponds to

$$\{2\} \otimes \{2\} \otimes \{2\} = \{4\} \oplus 2\{2\} \oplus \{0\}, \tag{2.4.44}$$

and in $SU(3)$

$$\{3\} \otimes \{3\} \otimes \{3\} = \{10\} \oplus 2\{8\} \oplus \{1\}. \tag{2.4.45}$$

Hence the color triplets of three constituent quarks can not only be coupled to an $SU(3)_C$ singlet but also to octets and a decuplet. Such states (like the quark triplet states themselves) have never been observed in Nature. Therefore, we think that all observable particles are color singlets (neutral with respect to the color charge). This is the confinement hypothesis that could up to now not been proved rigorously from QCD first principles.

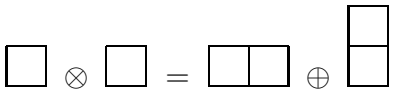
As an exercise using the above technique one can couple the fundamental representation with its anti-representation, and show that one obtains a singlet and the adjoint representation for any $SU(n)$, i.e.

$$\{\bar{n}\} \otimes \{n\} = \{1\} \oplus \{n^2 - 1\}. \tag{2.4.46}$$

In $SU(2)$ this reads $\{2\} \otimes \{2\} = \{1\} \oplus \{3\}$ and in $SU(3)$ we have $\{\bar{3}\} \otimes \{3\} = \{1\} \oplus \{8\}$.

2.5 Antiquarks and Mesons

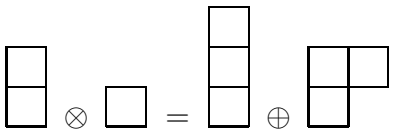
We have seen that the baryons (nucleon and Δ) consist of three constituent quarks (isospin doublets, spin doublets, color triplets). Now we want to construct the mesons (pion and ρ) in a similar manner. Since these particles have spin 0 and 1 respectively, they must contain an even number of constituent 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



$$\{3\} \otimes \{3\} = \{6\} \oplus \{\bar{3}\}, \tag{2.5.1}$$

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

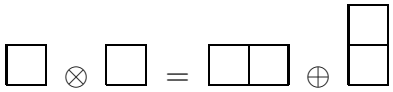
We have seen already that a representation together with its anti-representation can always be coupled to a singlet. In $SU(3)$ this corresponds to



$$\{\bar{3}\} \otimes \{3\} = \{1\} \oplus \{8\}, \tag{2.5.2}$$

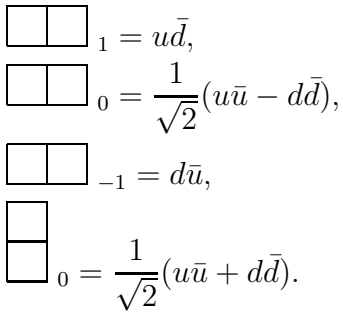
Hence it is natural to work with antiquarks. Antiquarks are isospin doublets, spin doublets and color anti-triplets. We have quarks $\bar{u}$ and $\bar{d}$ with electric charges $Q_{\bar{u}} = -2/3$ and $Q_{\bar{d}} = 1/3$. Now we consider combinations of quark

and anti-quark u$\bar{d}$, u$\bar{u}$, d$\bar{d}$ and d$\bar{u}$, which have charges 1, 0 and -1 as we need them for the mesons. First we couple the isospin wave function



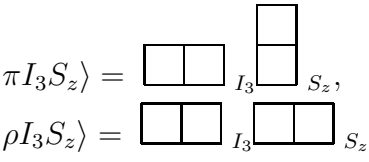
$$\{2\} \otimes \{2\} = \{3\} \oplus \{1\}, \tag{2.5.3}$$

and we obtain



$$= u\bar{d},$$
$$= \frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d}),$$
$$= d\bar{u},$$
$$= \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d}). \tag{2.5.4}$$

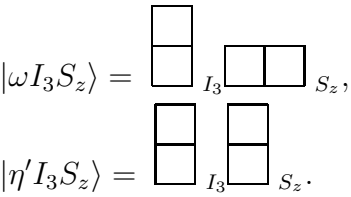
We proceed analogously for the spin and we obtain



$$|\pi I_3 S_z\rangle =$$
$$|\rho I_3 S_z\rangle = \tag{2.5.5}$$

Since quarks and antiquarks 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



$$|\omega I_3 S_z\rangle =$$
$$|\eta' I_3 S_z\rangle = \tag{2.5.6}$$

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

In the constituent quark model one would expect that the mesons have roughly two times the constituent quark mass, i.e. $2M_q = 0.672\text{GeV}$. For the vector mesons ($S = 1$) ρ and ω this works reasonably well. For the scalar mesons ($S = 0$) π and η', on the other hand, it does not. Here we see the limitations of the constituent quark model. The pion is much lighter than naively expected. This is because it is the pseudo Goldstone boson of the spontaneously broken chiral symmetry of QCD, and therefore it is almost massless. Later we will also discuss the $U(1)$-problem of the large η' mass.

2.6 Strange Hadrons and the s Quark

Up to now we have considered hadrons that consist of u and d quarks and their anti-particles. However, one also observes other so-called strange hadrons. In the meson sector their spectrum is given in fig.2.2. The masses of the scalar ($S = 0$) mesons are given by

$$M_\pi = 0.138\text{GeV}, M_K = 0.496\text{GeV}, M_\eta = 0.549\text{GeV}, M_{\eta'} = 0.958\text{GeV}, \tag{2.6.1}$$

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

$$M_\rho = 0.770\text{GeV}, M_\omega = 0.783\text{GeV}, M_{K^*} = 0.892\text{GeV}, M_\varphi = 1.020\text{GeV}. \tag{2.6.2}$$

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

$$\{\bar{2}\} \otimes \{2\} = \{1\} \oplus \{3\}. \tag{2.6.3}$$

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

$$\{\bar{3}\} \otimes \{3\} = \{1\} \oplus \{8\}. \tag{2.6.4}$$

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

Figure 2.2: *The spectrum of scalar and vector mesons.*

strange quark. Like u and d quarks the s quark comes in three colors (r, g and b).

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

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix},$$
$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{2.6.5}$$

Two of the generators commute with each other $[\lambda_3, \lambda_8] = 0$. We say that

Figure 2.3: *The states of an $SU(3)_F$ triplet and anti-triplet.*

the group $SU(3)$ has rank 2. One can now identify the generators of the isospin subgroup $SU(2)_I$

$$I_1 = \frac{1}{2}\lambda_1,\ I_2 = \frac{1}{2}\lambda_2,\ I_3 = \frac{1}{2}\lambda_3. \tag{2.6.6}$$

Also it is convenient to introduce the so-called hypercharge

$$Y = \frac{1}{\sqrt{3}}\lambda_8. \tag{2.6.7}$$

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

$$\begin{aligned} I^2 u &= \frac{1}{2}(\frac{1}{2}+1)u = \frac{3}{4}u,\ I_3 u = \frac{1}{2}u,\ Yu = \frac{1}{3}u, \\ I^2 d &= \frac{1}{2}(\frac{1}{2}+1)d = \frac{3}{4}d,\ I_3 d = -\frac{1}{2}d,\ Yd = \frac{1}{3}d, \\ I^2 s &= 0,\ I_3 s = 0,\ Ys = -\frac{2}{3}s. \end{aligned} \tag{2.6.8}$$

In fig.2.3 the states are represented in an I_3-Y diagram. Similarly the states of the scalar and vector mesons are displayed in fig.2.4 and fig.2.5.

Figure 2.4: *The scalar meson states in an I_3-Y diagram.*

Figure 2.5: *The vector meson states in an I_3-Y diagram.*

The electric charge is now given by

$$Q = I_3 + \frac{1}{2}Y, \tag{2.6.9}$$

such that

$$Q_u = \frac{2}{3},\ Q_d = -\frac{1}{3},\ Q_s = -\frac{1}{3}, \tag{2.6.10}$$

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

$$\{3\} \otimes \{3\} \otimes \{3\} = \{10\} \oplus 2\{8\} \oplus \{1\} \tag{2.6.11}$$

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

$$\{2\} \otimes \{2\} \otimes \{2\} = \{4\} \oplus 2\{2\} \oplus \{0\}. \tag{2.6.12}$$

Indeed one observes more baryons than just nucleon and Δ-isobar. The observed baryon spectrum is depicted in fig.2.6. The baryon masses for the spin 1/2 baryons are

$$M_N = 0.939\text{GeV},\ M_\Lambda = 1.116\text{GeV},\ M_\Sigma = 1.193\text{GeV},\ M_\Xi = 1.318\text{GeV}, \tag{2.6.13}$$

while the spin 3/2 baryon masses are

$$M_\Delta = 1.232\text{GeV},\ M_{\Sigma^*} = 1.385\text{GeV},\ M_{\Xi^*} = 1.530\text{GeV},\ M_\Omega = 1.672\text{GeV}. \tag{2.6.14}$$

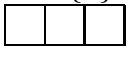
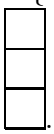
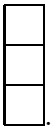
Proton and neutron are part of an octet:  is $\{2\}$ in $SU(2)_I$ and $\{8\}$ in $SU(3)_F$. The Δ-isobar is part of a decouplet:  is $\{4\}$ in $SU(2)_I$ and $\{10\}$ in $SU(3)_F$. One does not find an $SU(3)_F$ singlet  . This is because a spatially symmetric color singlet wave function is totally antisymmetric. To obtain a totally antisymmetric wave function also the spin part should

Figure 2.6: *The baryon spectrum.*

Figure 2.7: *The baryon octet and decouplet.*

transform as  . Of course, in $SU(2)_S$ this is impossible. We can arrange the baryons in I_3-Y diagrams as shown in fig.2.7.

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

$$M_{\Sigma^*} - M_\Delta = M_{\Xi^*} - M_{\Sigma^*} = M_\Omega - M_{\Xi^*} = M_s - M_q. \tag{2.6.15}$$

In fact one finds experimentally

$$M_{\Sigma^*} - M_\Delta = 0.153\text{GeV},\ M_{\Xi^*} - M_{\Sigma^*} = 0.145\text{GeV},\ M_\Omega - M_{\Xi^*} = 0.142\text{GeV}. \tag{2.6.16}$$

Averaging over the various mass differences one obtains

$$M_s = M_q + 0.147\text{GeV}, \tag{2.6.17}$$

i.e. the s quark is about 0.15 GeV heavier than the light u and d quarks.

2.7 The Gellman-Okubo 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, which we will discuss in detail later. Here we want to use very basic QCD physics together with group theory to describe patterns in the hadron spectrum. QCD is the non-Abelian $SU(3)_C$ gauge theory of quarks and gluons. Let us restrict ourselves to the quarks u, d and s. 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

$$\mathcal{M} = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_d & 0 \\ 0 & 0 & m_s \end{pmatrix}. \tag{2.7.1}$$

We want to assume that u and d quark have the same mass m_q, while the s quark is heavier ($m_s > m_q$). These quark masses are so-called current masses. They are not identical with the constituent quark masses M_q and M_s. The quark mass matrix can be written as

$$\begin{aligned} \mathcal{M} &= \frac{2m_q + m_s}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{m_q - m_s}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \\ &= \frac{2m_q + m_s}{3} 1 + \frac{m_q - m_s}{\sqrt{3}} \lambda_8. \end{aligned} \tag{2.7.2}$$

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

$$H_{QCD} = H_1 + H_8. \tag{2.7.3}$$

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 $|B_1 Y I I_3\rangle$

$$H_1|B_1 Y I I_3\rangle = M_{B_1}|B_1 Y I I_3\rangle. \tag{2.7.4}$$

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

$$M_B = M_{B_1} + \langle B_1 Y I I_3|H_8|B_1 Y I I_3\rangle. \tag{2.7.5}$$

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 $\{10\}$. It is instructive to perform this coupling using the Young tableaux method. Indeed the decouplet appears in the reduction. Using the Wigner-Eckart theorem we obtain

$$\langle B_1 Y I I_3|H_8|B_1 Y I I_3\rangle = \langle B_1||H_8||B_1\rangle\langle\{10\} Y I I_3|\{8\}000\{10\} Y I I_3\rangle, \tag{2.7.6}$$

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. These are tabulated (e.g. in J. J. de Swart, Rev. Mod. Phys. 35 (1963) 916) and are given by

$$\langle\{10\} Y I I_3|\{8\}000\{10\} Y I I_3\rangle = \frac{Y}{\sqrt{8}}. \tag{2.7.7}$$

Then we obtain for the baryon masses in the decouplet

$$M_B = M_{B_1} + \langle B_1||H_8||B_1\rangle\frac{Y}{\sqrt{8}}, \tag{2.7.8}$$

and hence

$$M_{\Sigma^*} - M_\Delta = M_{\Xi^*} - M_{\Sigma^*} = M_\Omega - M_{\Xi^*} = -\frac{1}{\sqrt{8}}\langle B_1||H_8||B_1\rangle. \tag{2.7.9}$$

Indeed, as we saw before, the three mass differences are almost identical. In view of the fact that we have 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

$$\{8\} \otimes \{8\} = \{27\} \oplus \{10\} \oplus \{\bar{10}\} \oplus 2\{8\} \oplus \{1\}. \tag{2.7.10}$$

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

$$\begin{aligned} \langle B_1 Y I I_3 | H_8 | B_1 Y I I_3 \rangle &= \langle B_1 || H_8 || B_1 \rangle_s \langle \{8\} Y I I_3 | \{8\} 000 \{8\} Y I I_3 \rangle_s \\ &+ \langle B_1 || H_8 || B_1 \rangle_a \langle \{8\} Y I I_3 | \{8\} 000 \{8\} Y I I_3 \rangle_a. \end{aligned} \tag{2.7.11}$$

The Clebsch-Gordon coefficients are given by

$$\begin{aligned} \langle \{8\} Y I I_3 | \{8\} 000 \{8\} Y I I_3 \rangle_s &= \frac{1}{\sqrt{5}} (I(I+1) - \frac{1}{4} Y^2 - 1), \\ \langle \{8\} Y I I_3 | \{8\} 000 \{8\} Y I I_3 \rangle_a &= \sqrt{\frac{3}{4}} Y, \end{aligned} \tag{2.7.12}$$

and we obtain for the baryon octet

$$M_B = M_{B_1} + \langle B_1 || H_8 || B_1 \rangle_s \frac{1}{\sqrt{5}} (I(I+1) - \frac{1}{4} Y^2 - 1) + \langle B_1 || H_8 || B_1 \rangle_a \sqrt{\frac{3}{4}} Y. \tag{2.7.13}$$

The formulas for the baryon masses are named after Gellman and Okubo. From the octet formula one obtains

$$\begin{aligned} 2M_N + 2M_\Xi &= 4M_{B_1} + \langle B_1 || H_8 || B_1 \rangle_s \frac{4}{\sqrt{5}} (\frac{3}{4} - \frac{1}{4} - 1), \\ M_\Sigma + 3M_\Lambda &= 4M_{B_1} + \langle B_1 || H_8 || B_1 \rangle_s \frac{1}{\sqrt{5}} ((2-1) + 3(-1)), \\ 2M_N + 2M_\Xi &= M_\Sigma + 3M_\Lambda. \end{aligned} \tag{2.7.14}$$

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

2.8 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

$$H_1|M_1 Y I I_3\rangle = M_{M_1}|M_1 Y I I_3\rangle. \tag{2.8.1}$$

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

$$M_M = M_{M_1} + \langle M_1 Y I I_3|H_8|M_1 Y I I_3\rangle. \tag{2.8.2}$$

In complete analogy to the baryon octet we obtain

$$M_M = M_{M_1} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(I(I+1) - \frac{1}{4}Y^2 - 1) + \langle M_1||H_8||M_1\rangle_a \sqrt{\frac{3}{4}}Y. \tag{2.8.3}$$

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

$$\begin{aligned} M_{K^+} &= M_{M_1} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(\frac{3}{4} - \frac{1}{4} - 1) + \langle M_1||H_8||M_1\rangle_a \sqrt{\frac{3}{4}}, \\ M_{K^-} &= M_{M_1} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(\frac{3}{4} - \frac{1}{4} - 1) - \langle M_1||H_8||M_1\rangle_a \sqrt{\frac{3}{4}}. \end{aligned} \tag{2.8.4}$$

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

$$\langle M_1||H_8||M_1\rangle_a = 0. \tag{2.8.5}$$

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

$$\begin{aligned}\langle\omega_1|H_8|\omega_1\rangle &= 0,\\ \langle\omega_8|H_8|\omega_8\rangle &= \langle M_1||H_8||M_1\rangle_s\langle\{8\}000|\{8\}000\{8\}000\rangle_s\\ &= \langle M_1||H_8||M_1\rangle_s(-\frac{1}{\sqrt{5}}).\end{aligned} \tag{2.8.6}$$

The actual meson masses are the eigenvalues of the matrix

$$\mathcal{M} = \begin{pmatrix} M_{\omega_1} & \langle\omega_1|H_8|\omega_8\rangle \\ \langle\omega_8|H_8|\omega_1\rangle & M_{\omega_8} - \langle M_1||H_8||M_1\rangle_s\frac{1}{\sqrt{5}}\end{pmatrix}. \tag{2.8.7}$$

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

$$\begin{aligned}|\varphi\rangle &= \cos\theta|\omega_1\rangle - \sin\theta|\omega_8\rangle,\\ |\omega\rangle &= \sin\theta|\omega_1\rangle + \cos\theta|\omega_8\rangle.\end{aligned} \tag{2.8.8}$$

Here θ is the meson mixing angle. One obtains

$$\begin{aligned}M_\varphi + M_\omega &= M_{\omega_1} + M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s,\\ M_\varphi M_\omega &= M_{\omega_1}(M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s) - |\langle\omega_1|H_8|\omega_8\rangle|^2.\end{aligned} \tag{2.8.9}$$

Also we have

$$\begin{aligned}M_\rho &= M_{\omega_8} + \langle M_1||H_8||M_1\rangle_s\frac{1}{\sqrt{5}}(2-1),\\ M_{K^*} &= M_{\omega_8} + \langle M_1||H_8||M_1\rangle_s\frac{1}{\sqrt{5}}(\frac{3}{4} - \frac{1}{4} - 1),\end{aligned} \tag{2.8.10}$$

and hence

$$\begin{aligned}\frac{4}{3}M_{K^*} - \frac{1}{3}M_\rho &= M_{\omega_8} + \langle M_1||H_8||M_1\rangle_s\frac{1}{\sqrt{5}}(\frac{4}{3}(-\frac{1}{2}) - \frac{1}{3})\\ &= M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s,\end{aligned} \tag{2.8.11}$$

such that

$$\begin{aligned}M_{\omega_1} &= M_\varphi + M_\omega - \frac{4}{3}M_{K^*} + \frac{1}{3}M_\rho = 0.870\text{GeV},\\ |\langle\omega_1|H_8|\omega_8\rangle|^2 &= M_{\omega_1}(\frac{4}{3}M_{K^*} - \frac{1}{3}M_\rho) - M_\varphi M_\omega = (0.113\text{GeV})^2.\end{aligned} \tag{2.8.12}$$

The mixing angle is now determined from

$$M_{\omega_1}\cos\theta - \langle\omega_1|H_8|\omega_8\rangle\sin\theta = M_\varphi\cos\theta,$$
$$\langle\omega_8|H_8|\omega_1\rangle\cos\theta - (M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s)\sin\theta = -M_\varphi\sin\theta. \tag{2.8.13}$$

and we obtain

$$(M_{\omega_1} + M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s)\sin\theta\cos\theta - \langle\omega_1|H_8|\omega_8\rangle =$$
$$2M_\varphi\sin\theta\cos\theta, \tag{2.8.14}$$

and hence

$$\frac{1}{2}\sin(2\theta) = \pm\frac{\sqrt{(M_\varphi + M_\omega - \frac{4}{3}M_{K^*} + \frac{1}{3}M_\rho)(\frac{4}{3}M_{K^*} - \frac{1}{3}M_\rho) - M_\varphi M_\omega}}{M_\varphi - M_\omega}. \tag{2.8.15}$$

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

$$\begin{aligned} |\varphi\rangle &\approx s\bar{s} \text{ or } \frac{1}{3}(2u\bar{u} + 2d\bar{d} - s\bar{s}), \\ |\omega\rangle &\approx \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d}) \text{ or } -\frac{1}{\sqrt{18}}(u\bar{u} + d\bar{d} + 4s\bar{s}). \end{aligned} \tag{2.8.16}$$

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

$$|\varphi\rangle \approx s\bar{s}, \quad |\omega\rangle \approx \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d}). \tag{2.8.17}$$

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

Chapter 3

The Parton Model

3.1 Hadrons as a Collection of Partons

As we have seen, the constituent quark model provides a rather good description of static hadron properties like mass, spin, charge or magnetic moment. We have viewed the constituent quarks as nonrelativistic quasiparticles, which are composed of fundamental quarks, antiquarks and gluons in a complicated way. Now we want to study the dynamics of hadrons in the extreme high energy regime, in which nonrelativistic concepts break down. In this regime one can use the so-called parton model, which describes a fast moving hadron as a collection of parallel moving partons sharing the total momentum of the hadron. We will later identify the partons with the quarks and gluons of QCD. Certainly, they have not much in common with the constituent quarks, although they have the same spin and flavor. Some typical hard — i.e. high energy — processes are electron-positron annihilation into hadrons ($e^+ + e^- \to X$) and deep-inelastic lepton-nucleon scattering ($e^- + N \to e^- + X$). In the parton model a hadron is described as a bundle of partons as illustrated in fig.3.1. The parton structure is determined by the strong interactions and is parametrized by so-called structure functions.

The internal structure of hadrons can be studied by electromagnetically or weakly interacting probes like electrons or neutrinos, which exchange a

Figure 3.1: *A hadron as a bundle of partons.*

photon γ or a $W^{\pm}$ or Z^0 boson with a parton inside the hadron. Various processes of this kind are illustrated in fig.3.2. The elementary parton processes can be computed perturbatively in the Standard model. Then one sums over all parton processes.

In the final state of a hadronic reaction due to confinement one never finds single partons but only hadrons. Each parton therefore hadronizes into a jet, and can therefore in turn be viewed as a bundle of hadrons. This is illustrated in fig.3.3. We will now study some hard processes in more detail, trying to derive the properties of the partons from the observed properties of hadrons.

3.2 Electron Positron Annihilation into Hadrons

Electron-positron annihilation into hadrons has been studied in detail at e^+e^- accelerators like PETRA at DESY in Hamburg and PEP at SLAC in Stanford. The energy released in the annihilation materializes again as particles. Of course, not only strongly interacting hadrons are created but also lepton pairs (e.g. muons). The elementary process that turns an electron-positron pair into a muon-antimuon pair is described by the Feyn-

Figure 3.2: *Various deep-inelastic scattering processes.*

man diagram shown in fig.3.4. The annihilation into hadrons is described by a corresponding fundamental process involving a quark-antiquark pair as illustrated in fig.3.5. The quark-antiquark pair then hadronizes in a complicated way involving nonperturbative strong interaction processes as symbolized in fig.3.6.

We want to compute the ratio R of the total cross sections for electron-positron annihilation into hadrons and muons at high energies

$$R = \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)} = \sum_{\text{partons q}} \frac{\sigma(e^+e^- \to q\bar{q})}{\sigma(e^+e^- \to \mu^+\mu^-)}. \qquad (3.2.1)$$

Let us consider the kinematics of the elementary process illustrated in fig.3.7. Before the annihilation we have the following four-momenta

Figure 3.3: *A parton as a jet of hadrons.*

$$p_{e^-} = (\sqrt{m_e^2 + |\vec{p}|^2}, \vec{p}),\ p_{e^+} = (\sqrt{m_e^2 + |\vec{p}|^2}, -\vec{p}), \tag{3.2.2}$$

and after the annihilation we have

$$p_{\mu^-} = (\sqrt{m_\mu^2 + |\vec{p}'|^2}, \vec{p}'),\ p_{\mu^+} = (\sqrt{m_\mu^2 + |\vec{p}'|^2}, -\vec{p}'). \tag{3.2.3}$$

Energy conservation implies

$$2\sqrt{m_e^2 + |\vec{p}|^2} = 2\sqrt{m_\mu^2 + |\vec{p}'|^2}. \tag{3.2.4}$$

The square of the invariant mass of the system is denoted by

$$s = (p_{e^-} + p_{e^+})^2 = (p_{\mu^-} + p_{\mu^+})^2 = 4(m_e^2 + |\vec{p}|^2) = 4(m_\mu^2 + |\vec{p}'|^2). \tag{3.2.5}$$

The cross section σ measures the number of muons that are created per unit of time, normalized to the incoming flux of electron-positron pairs. The dimension of the cross section is length squared. The cross section can be computed from the S-matrix element corresponding to the Feynman diagram of fig.3.4. Here we will not derive the result in this way, but simply state it. We restrict ourselves to high energies, i.e. $m_e^2, m_\mu^2 \ll s$, and find

$$\sigma(e^+e^- \to \mu^+\mu^-) = \frac{4\pi\alpha^2}{3s}. \tag{3.2.6}$$

Here $\alpha = e^2/4\pi = 1/137$ is the fine structure constant. It is easy to understand this result qualitatively. First the two vertices in the Feynman diagram each contribute a factor e to the S-matrix element. The cross section

Figure 3.4: *Feynman diagram for electron-positron annihilation into a muon-antimuon pair.*

is proportional to the S-matrix element squared, and hence proportional to e^4. In the high energy limit

$$p^2 = p'^2 = \frac{s}{4} \tag{3.2.7}$$

is the only dimensionful quantity in the problem. Hence σ must be proportional to $1/s$. Since in the following we are interested only in the ratio of cross sections the numerical proportionality factor drops out anyways. The relevant parton process is illustrated in fig.3.5. Again we consider the high energy limit $m_e^2, m_q^2 \ll s$. One of the two vertices is still given by the charge of the electron e. The other vertex, however, contributes the charge $Q_q e$ of the parton. We want to assume that the partons are elementary fermions just like the muon. Then we obtain

$$\sigma(e^+e^- \to q\bar{q}) = \frac{4\pi\alpha^2}{3s} Q_q^2, \tag{3.2.8}$$

and hence

$$R = \sum_{\text{partons q}} \frac{\sigma(e^+e^- \to q\bar{q})}{\sigma(e^+e^- \to \mu^+\mu^-)} = \sum_{\text{partons q}} Q_q^2. \tag{3.2.9}$$

Experimental results for R are illustrated in figs.3.8 and 3.9. First one realizes that for large $\sqrt{s} > 2\text{GeV}$ the ratio R becomes essentially energy

Figure 3.5: *Feynman diagram for electron-positron annihilation into a quark-antiquark pair.*

independent, and approaches the value 2. Up to now we have worked with three quarks u, d and s with electric charges

$$Q_u = \frac{2}{3},\ Q_d = -\frac{1}{3},\ Q_s = -\frac{1}{3}. \tag{3.2.10}$$

Summing over these three flavors we obtain

$$\sum_{f=u,d,s} Q_f^2 = (\frac{2}{3})^2 + (\frac{1}{3})^2 + (\frac{1}{3})^2 = \frac{2}{3}. \tag{3.2.11}$$

There is a factor 3 missing compared to the experimental result. However, in the constituent quark model quarks where introduced with three colors (r, g and b). Taking this into account we find indeed

$$\sum_{\text{partons q}} Q_q^2 = \sum_{c=r,g,b} \sum_{f=u,d,s} Q_f^2 = 3\frac{2}{3} = 2. \tag{3.2.12}$$

This can be viewed as an experimental verification of the color degree of freedom.

From the experimental data we see that for $\sqrt{s} > 3$GeV the ratio R increases beyond 2. This points to the existence of another quark with a mass around 1.5GeV, which cannot be created at lower energies. This so-called charm or c quark leads to the J/Ψ resonance ($c\bar{c}$ state) and has the

Figure 3.6: *Hadronization of created quark-antiquark pair.*

Figure 3.7: *Kinematics of electron-positron annihilation into muons.*

charge $Q_c = 2/3$ just like the u quark. In the higher energy range we then obtain

$$3 \sum_{f=u,d,s,c} Q_f^2 = 2 + 3(\frac{2}{3})^2 = \frac{10}{3}. \tag{3.2.13}$$

The same phenomenon occurs again when we reach the mass of the bottom or b quark at about 5GeV. The b quark gives rise to the Υ resonance and has $Q_b = -1/3$ like the d and s quarks. We then get

$$3 \sum_{f=u,d,s,c,b} Q_f^2 = \frac{10}{3} + 3(\frac{1}{3})^2 = \frac{11}{3} \tag{3.2.14}$$

in reasonable agreement with the experiments. The remaining small deviations can be understood in QCD perturbation theory.

Figure 3.8: *Data for R at center of mass energies less than 3.5 GeV. The prediction of the quark parton model (QPM) for this energy regime is 2.*

The last quark in the Standard model is the top or t quark that was recently discovered at Fermilab with a mass of about 180GeV and with $Q_t = 2/3$. Naively one would then expect that for $\sqrt{s} > 360$GeV the ratio should be

$$3 \sum_{f=u,d,s,c,b,t} Q_f^2 = \frac{11}{3} + 3(\frac{2}{3})^2 = 5. \tag{3.2.15}$$

This is, however, not so because below the t quark threshold there are the W bosons as other charged particles that can be pair created. Since they are vector bosons (not elementary fermions like the quarks) their inclusion would require the evaluation of another Feynman diagram. Also the annihilation could then proceed via other channels than just the photon. For example, the electron-positron pair can then annihilate into a Z or if energetically possible into a Higgs boson. Due to its large Yukawa coupling the top quark couples particularly strongly to the Higgs particle, such that the above purely electromagnetic determination of R would certainly fail at these very high energies.

Figure 3.9: *Experimental data for R for center of mass energies between 8 GeV and 45 GeV. The prediction of the quark parton model (QPM) for energies above 12 GeV is 11/3.*

3.3 Deep-inelastic Electron-Nucleon Scattering

To study the interior of hadrons we now bombard them with electromagnetic or weakly interacting probes like electrons or neutrinos. Let us begin with electron-nucleon scattering $e+N \to e+X$ at high energies. To leading order in α the corresponding diagram in the parton model is shown in fig.3.10. We are interested in reactions at high momentum transfers q. In a hard reaction the nucleon will in general be destroyed, and we will find a variety of hadrons X in the final state. Among those will certainly be another baryon, because the strong interactions conserve the total baryon number. Since other hadrons are created as well the reaction is inelastic. The experiments have been done such that the scattered electron is analyzed for scattering angle and energy while the emerging hadrons are left undetected. Let us first look at the invariant mass W of the hadronic final state

$$W^2 = (p+q)^2 = p^2 + 2pq + q^2 = M_N^2 + 2M_N\nu - Q^2. \tag{3.3.1}$$

Figure 3.10: *Deep-inelastic electron-nucleon scattering.*

We have used the nucleon mass $M_N^2 = p^2$, the energy transfer in the rest frame of the nucleon

$$\nu = \frac{pq}{M_N}, \tag{3.3.2}$$

as well as the four-momentum transfer squared

$$Q^2 = -q^2. \tag{3.3.3}$$

In the rest frame of the nucleon one has

$$p = (M_N, 0),\ k = (E, \vec{k}),\ k' = (E', \vec{k}') \tag{3.3.4}$$

such that $pk = M_N E$, $pk' = M_N E'$ and hence

$$\nu = E - E' = \frac{p}{M_N}(k - k') = \frac{pq}{M_N}. \tag{3.3.5}$$

The scattering angle of the electron (in the rest frame of the nucleon) results from

$$\cos\theta = \frac{\vec{k} \cdot \vec{k}'}{|\vec{k}||\vec{k}'|}. \tag{3.3.6}$$

At high energies we can neglect the electron mass, i.e.

$$E = |\vec{k}|,\ E' = |\vec{k}'|, \tag{3.3.7}$$

such that

$$1 - 2\sin^2\frac{\theta}{2} = \frac{EE' - kk'}{EE'},\ kk' = 2EE'\sin^2\frac{\theta}{2}. \tag{3.3.8}$$

Hence one obtains

$$Q^2 = -q^2 = -(k-k')^2 = 2kk' = 4EE'\sin^2\frac{\theta}{2}. \tag{3.3.9}$$

Due to baryon number conservation the hadronic final state contains at least one baryon. Therefore the invariant mass must at least be the nucleon mass

$$W^2 = M_N^2 + 2M_N\nu - Q^2 \geq M_N^2, \tag{3.3.10}$$

such that the so-called Bjorken scaling variable

$$x = \frac{Q^2}{2M_N\nu} \leq 1. \tag{3.3.11}$$

is always less than 1. Since $E \geq E'$ (i.e. $\nu \geq 0$) we also have $x \geq 0$. In the so-called Bjorken limit both the energy and the momentum transfer (ν and Q^2) go to infinity, but such that their ratio x remains constant.

The S-matrix element for our reaction can now be written as

$$\begin{aligned}&\langle e(k')X(p')|S|e(k)N(p)\rangle = \\ &-i(2\pi)^4\delta(p'+k'-p-k)4\pi\alpha\bar{e}(k')\gamma^\mu e(k)\frac{1}{q^2}\langle X(p')|J_\mu|N(p)\rangle.\end{aligned} \tag{3.3.12}$$

Here $\alpha = e^2/4\pi$ is the fine structure constant, the δ-function ensures four-momentum conservation, $\bar{e}(k')\gamma^\mu e(k)$ is the electromagnetic current of the electron, $1/q^2$ is the photon propagator and J_μ is operator for the hadronic electromagnetic current. In QCD the current J_μ is expressed in terms of quark fields

$$J_\mu = \sum_{f=u,d,s,c,b,t} Q_f\bar{\Psi}_f\gamma_\mu\Psi_f. \tag{3.3.13}$$

To evaluate the differential cross section one squares the S-matrix element

$$\frac{d^2\sigma}{dE'd\Omega} = \frac{\alpha^2}{Q^4}\frac{E'}{E}\frac{1}{2}\sum_{\text{spins}}\bar{e}(k')\gamma^\mu e(k)[\bar{e}(k')\gamma^\nu e(k)]^*$$

$$\times \sum_X \frac{1}{M_N}(2\pi)^3\delta(p'-p-q)\frac{1}{2}\sum_{\text{spins}}\langle N(p)|J_\mu|X(p')\rangle\langle X(p')|J_\nu|N(p)\rangle$$
$$= \frac{\alpha^2}{Q^4}\frac{E'}{E}L^{\mu\nu}W_{\mu\nu}. \tag{3.3.14}$$

Averaging over the unobserved electron spin polarizations one obtains for the leptonic tensor

$$L^{\mu\nu} = \frac{1}{2}\sum_{\text{spins}}\bar{e}(k')\gamma^\mu e(k)[\bar{e}(k')\gamma^\nu e(k)]^* = 2(k^\mu k'^\nu + k^\nu k'^\mu - kk'g^{\mu\nu}). \tag{3.3.15}$$

We have also introduced a hadronic tensor

$$W_{\mu\nu} = \sum_X \frac{1}{M_N}(2\pi)^3\delta(p'-p-q)\frac{1}{2}\sum_{\text{spins}}\langle N(p)|J_\mu|X(p')\rangle\langle X(p')|J_\nu|N(p)\rangle. \tag{3.3.16}$$

The hadronic tensor contains all the complicated dynamics of the strong interactions, while the leptonic tensor is simple. Still, symmetries restrict the form of the hadronic tensor. First of all, $W_{\mu\nu}$ only depends on the four-vectors p_μ and q_μ. From those we can build three Lorentz scalars

$$p^2 = p^\mu p_\mu = M_N^2,\ pq = p^\mu q_\mu = \nu M_N,\ q^2 = q^\mu q_\mu = -Q^2. \tag{3.3.17}$$

The independent scalars can be chosen to be ν and Q^2. Next we ask which tensors we can form. One has $p_\mu p_\nu$, $p_\mu q_\nu$, $q_\mu p_\nu$ and $q_\mu q_\nu$. Besides these there is the metric tensor $g_{\mu\nu}$ and the totally antisymmetric tensor $\varepsilon_{\mu\nu\rho\sigma}$ from which we can construct $\varepsilon_{\mu\nu\rho\sigma}p^\rho q^\sigma$. For symmetry reasons we can hence write

$$W_{\mu\nu} = Ap_\mu p_\nu + Bp_\mu q_\nu + Cq_\mu p_\nu + Dq_\mu q_\nu + Eg_{\mu\nu} + F\varepsilon_{\mu\nu\rho\sigma}p_\rho q_\sigma. \tag{3.3.18}$$

Here $A, B, ..., F$ are scalar functions of the variables ν and Q^2. We want to restrict $W_{\mu\nu}$ further by using additional properties of the electromagnetic current operator J_μ. First of all, the operator is Hermitean such that

$$W_{\nu\mu} = W^*_{\mu\nu}. \tag{3.3.19}$$

Hence A, D, E are real, F is purely imaginary and $B^* = C$. The electromagnetic interaction conserves parity (in contrast to the weak interactions).

Therefore the parity odd $\varepsilon_{\mu\nu\rho\sigma}$ term cannot appear and $F = 0$. Finally, the electromagnetic current is conserved, i.e.

$$\partial^\mu J_\mu = 0 \;\Rightarrow\; q^\mu W_{\mu\nu} = q^\nu W_{\mu\nu} = 0. \tag{3.3.20}$$

For arbitrary p and q we then obtain

$$\begin{aligned}
&q^\mu W_{\mu\nu} = A\nu M_N p_\nu + B\nu M_N q_\nu - CQ^2 p_\nu - DQ^2 q_\nu + Eq_\nu = 0,\\
&q^\nu W_{\mu\nu} = A\nu M_N p_\mu - BQ^2 p_\mu + C\nu M_N q_\mu - DQ^2 q_\mu + Eq_\mu = 0 \;\Rightarrow\\
&A\nu M_N - CQ^2 = 0,\; B\nu M_N - DQ^2 + E = 0,\\
&A\nu M_N - BQ^2 = 0,\; C\nu M_N - DQ^2 + E = 0 \;\Rightarrow\; B = C,
\end{aligned} \tag{3.3.21}$$

such that B and C are also real. Now we obtain

$$B = \frac{\nu M_N}{Q^2} A,\; E = DQ^2 - B\nu M_N = DQ^2 - \frac{M_N^2 \nu^2}{Q^2} A, \tag{3.3.22}$$

and hence

$$\begin{aligned}
W_{\mu\nu} &= A(p_\mu p_\nu + \frac{M_N \nu}{Q^2}(p_\mu q_\nu + p_\nu q_\mu - M_N \nu g_{\mu\nu})) + D(q_\mu q_\nu + Q^2 g_{\mu\nu})\\
&= A(p_\mu + \frac{M_N \nu}{Q^2} q_\mu)(p_\nu + \frac{M_N \nu}{Q^2} q_\nu)\\
&+ (D - A\frac{\nu^2 M_N^2}{Q^4})(q_\mu q_\nu + Q^2 g_{\mu\nu}).
\end{aligned} \tag{3.3.23}$$

It is common to replace A and D by the quantities

$$\begin{aligned}
W_1^{eN}(\nu, Q^2) &= \frac{\nu^2 M_N^2}{Q^2} A(\nu, Q^2) - Q^2 D(\nu, Q^2),\\
W_2^{eN}(\nu, Q^2) &= M_N^2 A(\nu, Q^2).
\end{aligned} \tag{3.3.24}$$

It is a good exercise to compute the differential cross section by inserting the leptonic and hadronic tensors

$$\begin{aligned}
&\frac{d^2\sigma}{dE' d\Omega} = \frac{\alpha^2}{Q^4}\frac{E'}{E} L^{\mu\nu} W_{\mu\nu}\\
&= \frac{\alpha^2}{Q^4}\frac{E'}{E} 2(k^\mu k'^\nu + k^\nu k'^\mu - kk' g^{\mu\nu})\\
&\times[-W_1^{eN}\frac{1}{Q^2}(q_\mu q_\nu + Q^2 g_{\mu\nu}) + W_2^{eN}\frac{1}{M_N^2}(p_\mu + \frac{M_N\nu}{Q^2} q_\mu)(p_\nu + \frac{M_N \nu}{Q^2} q_\nu)].
\end{aligned} \tag{3.3.25}$$

After some algebra one finds

$$\frac{d^2\sigma}{dE'd\Omega} = 4\frac{\alpha^2 E'^2}{Q^4}[2W_1^{eN}(\nu, Q^2)\sin^2\frac{\theta}{2} + W_2^{eN}(\nu, Q^2)\cos^2\frac{\theta}{2}]. \qquad (3.3.26)$$

The so-called structure functions W_1 and W_2 cannot be determined from symmetry considerations. They contain the information about the nonperturbative dynamics of the strong interactions.

Up to now we have just done kinematics. Next we will describe deep-inelastic scattering in the parton model. For this purpose we first go from the rest frame of the nucleon to a frame in which the nucleon moves very fast. The so-called Breit frame is characterized by the fact that there is only a momentum but no energy transfer, i.e. $q_0 = 0$. Furthermore the momentum transfer $\vec{q}$ is aligned with the negative z-axis, such that

$$q = (0, \vec{q}) = (0, 0, 0, -Q). \qquad (3.3.27)$$

Furthermore

$$p = (\sqrt{P^2 + M_N^2}, 0, 0, P), \qquad (3.3.28)$$

where

$$q^2 = -Q^2,\ pq = PQ = M_N\nu \ \Rightarrow\ P = \frac{M_N\nu}{Q} = \frac{Q}{2x}. \qquad (3.3.29)$$

In the Bjorken limit $Q^2, \nu \to \infty$ with $x = Q^2/2M_N\nu$ fixed we have $P \to \infty$, such that the nucleon moves very fast. In the parton model the nucleon momentum is shared by parallel moving partons of different types i. We write

$$\vec{p_i} = \xi(0, 0, P),\ \xi \in [0, 1], \qquad (3.3.30)$$

i.e. we assume that all partons move parallel to each other (not antiparallel). We now characterize the nucleon by a distribution function $N_i(\xi)$ that counts the number of partons of type i that have momentum fraction ξ. Momentum conservation then implies

$$\sum_i \int_0^1 d\xi\ \xi N_i(\xi) = 1. \qquad (3.3.31)$$

We consider an elementary scattering reaction for a single parton as shown in fig.3.11. The reaction is described by a parton tensor

Figure 3.11: *Elementary electron-parton scattering.*

$$
\begin{aligned}
P^i_{\mu\nu} &= \frac{Q_i^2}{p_i^0}\delta(2p_iq+q^2)\frac{1}{4}\sum_{\text{spins}}\bar{i}(p_i')\gamma_\mu i(p_i)[\bar{i}(p_i')\gamma_\nu i(p_i)]^* \\
&= \frac{Q_i^2}{p_i^0}\delta(2p_iq+q^2)(p_{i\mu}p'_{i\nu}+p_{i\nu}p'_{i\mu}-p_ip'_ig_{\mu\nu}). \qquad (3.3.32)
\end{aligned}
$$

Here we have assumed that the charged partons of type i are elementary fermions carrying the electric charge Q_i. Neutral partons, of course, do not contribute to electron scattering. We decompose the partonic tensor as follows

$$
\begin{aligned}
&p_{i\mu}p'_{i\nu}+p_{i\nu}p'_{i\mu}-p_ip'_ig_{\mu\nu} = \\
&(-g_{\mu\nu}+\frac{q_\mu q_\nu}{q^2})A+(p_{i\mu}-\frac{p_iq}{q^2}q_\mu)(p_{i\nu}-\frac{p_iq}{q^2}q_\nu)B. \qquad (3.3.33)
\end{aligned}
$$

It is a good exercise to determine A and B. One finds

$$
A=-\frac{q^2}{2},\ B=2, \qquad (3.3.34)
$$

and hence the partonic tensor takes the form

$$
\begin{aligned}
&P^i_{\mu\nu}=\frac{Q_i^2}{p_i^0}\delta(2p_iq+q^2) \\
&\times[(-g_{\mu\nu}+\frac{q_\mu q_\nu}{q^2})(-\frac{q^2}{2})+2(p_{i\mu}-\frac{p_iq}{q^2}q_\mu)(p_{i\nu}-\frac{p_iq}{q^2}q_\nu)]. \qquad (3.3.35)
\end{aligned}
$$

In the next step we integrate over the parton momenta and we sum over the different types of partons in order to compare with the hadronic tensor. We use

$$p_i^2 = p_i'^2 = (p_i + q)^2 = 0 \;\Rightarrow\; 2p_iq + q^2 = 0, \tag{3.3.36}$$

as well as

$$p_i = (\xi P, 0, 0, \xi P) \;\Rightarrow\; p_i^0 = \xi P. \tag{3.3.37}$$

At high energies we can neglect the nucleon mass and we obtain $p = (P, 0, 0, P)$ such that

$$p_i = \xi p \;\Rightarrow\; \xi = \frac{p_iq}{pq} = \frac{-q^2}{2M_N\nu} = x. \tag{3.3.38}$$

We average the partonic tensor over all types of partons and we integrate over their momentum distribution

$$\begin{aligned}
&\sum_i \int_0^1 d\xi N_i(\xi) P^i_{\mu\nu} = \\
&\sum_i \int_0^1 d\xi N_i(\xi) \frac{Q_i^2}{\xi P} \delta(2p_iq + q^2) \\
&\times[(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2})(-\frac{q^2}{2}) + 2\xi^2(p_\mu - \frac{pq}{q^2}q_\mu)(p_\nu - \frac{pq}{q^2}q_\nu)] = \\
&\sum_i \int_0^1 d\xi N_i(\xi) \frac{Q_i^2}{\xi P} \frac{1}{2pq} \delta(\xi - x) \\
&\times[(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2})(-\frac{q^2}{2}) + 2\xi^2(p_\mu - \frac{pq}{q^2}q_\mu)(p_\nu - \frac{pq}{q^2}q_\nu)] = \\
&\sum_i N_i(x) Q_i^2 [(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2}) \frac{Q^2}{xP4pq} + (p_\mu - \frac{pq}{q^2}q_\mu)(p_\nu - \frac{pq}{q^2}q_\nu) \frac{2x^2}{xP2pq}] = \\
&\sum_i N_i(x) Q_i^2 [(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2}) \frac{Q}{2M_N\nu} + (p_\mu - \frac{pq}{q^2}q_\mu)(p_\nu - \frac{pq}{q^2}q_\nu) \frac{xQ}{(M_N\nu)^2}].
\end{aligned} \tag{3.3.39}$$

This is to be compared with the hadronic tensor

$$\begin{aligned}
\frac{M_N}{p^0} W_{\mu\nu} &= (-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2}) \frac{Q}{\nu} W_1^{eN}(\nu, Q^2) \\
&+ (p_\mu - \frac{pq}{q^2}q_\mu)(p_\nu - \frac{pq}{q^2}q_\nu) \frac{Q}{\nu M_N^2} W_2^{eN}(\nu, Q^2).
\end{aligned} \tag{3.3.40}$$

Figure 3.12: *Data of the European Muon Collaboration (EMC) for the structure function $\nu W_2^{\mu p}(\nu, Q^2)$ as a function of the Bjorken scaling variable x for various values of Q^2.*

Hence we identify

$$W_1^{eN}(\nu, Q^2) = \frac{1}{2M_N} \sum_i N_i(x) Q_i^2,$$
$$W_2^{eN}(\nu, Q^2) = \frac{1}{\nu} x \sum_i N_i(x) Q_i^2. \tag{3.3.41}$$

The structure functions do not depend on ν and Q^2 individually, but only on the combination

$$x = \frac{Q^2}{2M_N \nu}. \tag{3.3.42}$$

This behavior is known as Bjorken scaling. Indeed scaling is observed in the experimentally determined structure functions. This is illustrated in fig.3.12. Still, there are small violations of scaling, that cannot be understood in the naive parton model. Perturbative QCD, however, provides a systematic expansion about the free parton limit, and predicts successfully the observed corrections to scaling. It is remarkable that our calculation has revealed a new interpretation of Bjorken's scaling variable x. It characterizes the momentum fraction of the scattering parton before it is hit by the hard virtual photon. It is common to define other scaling functions

$$F_1^{eN}(x) = \lim_{\nu, Q^2 \to \infty} 2M_N W_1^{eN}(\nu, Q^2) = \sum_i N_i(x) Q_i^2,$$

$$F_2^{eN}(x) = \lim_{\nu, Q^2 \to \infty} \nu W_2^{eN}(\nu, Q^2) = x \sum_i N_i(x) Q_i^2. \tag{3.3.43}$$

This leads to the Callan-Gross relation

$$F_2^{eN}(x) = x F_1^{eN}(x). \tag{3.3.44}$$

Averaged over x one finds experimentally

$$\langle \frac{F_2^{eN}(x) - x F_1^{eN}(x)}{F_2^{eN}(x)} \rangle_x = -0.01 \pm 0.11. \tag{3.3.45}$$

The Callan-Gross relation would not be satisfied if the charged partons would not have spin 1/2. Hence, the experimental data confirm that the partons of the nucleon are elementary fermions.

3.4 Deep-inelastic Neutrino-Nucleon Scattering

In neutrino-nucleon scattering we are most interested in processes that have an easily detectable charged lepton in the final state and that hence proceed via charged currents (W bosons). In the parton model such a process is illustrated in fig.3.13. The weak interactions violate parity. Therefore the weak current of the quarks contains the matrix γ_5

$$J_\mu^+ = \bar{\Psi}_u \gamma_\mu (1 - \gamma_5) \cos\theta_C \Psi_d + \bar{\Psi}_u \gamma_\mu (1 - \gamma_5) \sin\theta_C \Psi_s. \tag{3.4.1}$$

The Cabbibo angle θ_C describes the mixing of d and s quarks under the weak interactions. We have neglected the contributions of c, b and t quarks. The mixing among all three generations would be described by the Kobayashi-Maskawa matrix, which θ_C is part of. Due to parity violation there is now an additional term $F \varepsilon_{\mu\nu\rho\sigma} p^\rho q^\sigma$ in the hadronic tensor. Correspondingly, there is a third structure function W_3. For the differential cross section one now finds

$$\begin{aligned} \frac{d^2\sigma}{dE' d\Omega} &= \frac{G_F^2 E'^2}{2\pi^2} [2W_1^{\nu N}(\nu, Q^2) \sin^2\frac{\theta}{2} + W_2^{\nu N}(\nu, Q^2) \cos^2\frac{\theta}{2} \\ &- W_3^{\nu N}(\nu, Q^2) \frac{E + E'}{M_N} \sin^2\frac{\theta}{2}]. \end{aligned} \tag{3.4.2}$$

Figure 3.13: *Neutrino-nucleon scattering in the parton model.*

Here $G_F = 1.16 \cdot 10^{-5}\text{GeV}^{-2}$ is the Fermi coupling constant of the weak interactions. The sign of the W_3 term changes under an exchange of neutrino and antineutrino. In analogy to electron-nucleon scattering one finds in the Bjorken limit

$$\begin{aligned}
F_1^{\nu N}(x) &= \lim_{\nu,Q^2\to\infty} 2M_N W_1^{\nu N}(\nu, Q^2) \\
&= 2\cos^2\theta_C N_d(x) + 2\sin^2\theta_C N_s(x) + 2N_{\bar{u}}(x), \\
F_2^{\nu N}(x) &= \lim_{\nu,Q^2\to\infty} \nu W_2^{\nu N}(\nu, Q^2) = x F_1^{\nu N}(x), \\
F_3^{\nu N}(x) &= \lim_{\nu,Q^2\to\infty} \nu W_3^{\nu N}(\nu, Q^2) \\
&= -2\cos^2\theta_C N_d(x) - 2\sin^2\theta_C N_s(x) + 2N_{\bar{u}}(x). \qquad (3.4.3)
\end{aligned}$$

At CERN the fluid freon containing an equal number of protons and neutrons has been investigated using electrons as well as neutrinos. For protons and neutrons one expects

$$\begin{aligned}
F_1^{ep}(x) &= \frac{4}{9}p_u(x) + \frac{1}{9}p_d(x) + \frac{1}{9}p_s(x) + \frac{4}{9}p_{\bar{u}}(x) + \frac{1}{9}p_{\bar{d}}(x) + \frac{1}{9}p_{\bar{s}}(x), \\
F_1^{en}(x) &= \frac{4}{9}n_u(x) + \frac{1}{9}n_d(x) + \frac{1}{9}n_s(x) + \frac{4}{9}n_{\bar{u}}(x) + \frac{1}{9}n_{\bar{d}}(x) + \frac{1}{9}n_{\bar{s}}(x),
\end{aligned}$$

$$F_1^{\nu p} = 2\cos^2\theta_C p_d(x) + 2\sin^2\theta_C p_s(x) + 2p_{\bar{u}}(x),$$
$$F_1^{\nu n} = 2\cos^2\theta_C n_d(x) + 2\sin^2\theta_C n_s(x) + 2n_{\bar{u}}(x). \tag{3.4.4}$$

Let us assume that the probability for finding s quarks in the nucleon is small. In the constituent quark model this would certainly be correct. Here, however, it is an approximation. We use isospin symmetry to conclude

$$p_u(x) = n_d(x),\ p_d(x) = n_u(x),\ p_{\bar{u}}(x) = n_{\bar{d}}(x),\ p_{\bar{d}}(x) = n_{\bar{u}}(x). \tag{3.4.5}$$

Averaging over proton and neutron we then obtain

$$F_1^{ep}(x) + F_1^{en}(x) = \frac{5}{9}p_u(x) + \frac{5}{9}p_d(x) + \frac{5}{9}p_{\bar{u}}(x) + \frac{5}{9}p_{\bar{d}}(x),$$
$$F_1^{\nu p} + F_1^{\nu n} = 2\cos^2\theta_C p_d(x) + 2\cos^2\theta_C p_u(x) + 2p_{\bar{u}}(x) + 2p_{\bar{d}}(x). \tag{3.4.6}$$

Using $\cos^2\theta_C = 0.96 \approx 1$ one obtains

$$\frac{F_2^{\nu p} + F_2^{\nu n}}{F_2^{ep}(x) + F_2^{en}(x)} = \frac{F_1^{\nu p} + F_1^{\nu n}}{F_1^{ep}(x) + F_1^{en}(x)} \approx \frac{18}{5}. \tag{3.4.7}$$

As shown in fig.3.14 this result is in good agreement with experiment, confirming directly the fractional electric charges of the quarks. The experimentally extracted parton momentum distributions are shown in fig.3.15.

3.5 Sum Rules

We can use conserved quantities to derive sum rules for the parton distributions. For example, the charge of the nucleon is given by

$$Q_N = \int_0^1 dx \sum_{f=u,d,s} (Q_f N_f(x) - Q_f N_{\bar{f}}(x)), \tag{3.5.1}$$

and the baryon number is given by

$$1 = \int_0^1 dx \sum_{f=u,d,s} (\frac{1}{3}N_f(x) - \frac{1}{3}N_{\bar{f}}(x)). \tag{3.5.2}$$

Figure 3.14: *The structure function $F_2^{\nu N}$ averaged over proton and neutron compared to $18F_2^{eN}/5$ (solid curve).*

Figure 3.15: *The momentum distribution of u, d and sea quarks ($S = \bar{u}, \bar{d}$).*

Both sum rules are in agreement with experiment. One can also formulate a sum rule for the total momentum

$$1 = \int_0^1 dx\ x \sum_i N_i(x). \tag{3.5.3}$$

Here we have summed over all partons of the nucleon. The contribution of quarks and antiquarks is given by

$$I_q = \int_0^1 dx\ x \sum_{f=u,d,s} (N_f(x) + N_{\bar{f}}(x)). \tag{3.5.4}$$

Again neglecting the Cabbibo angle, i.e. using $\cos^2\theta \approx 1$, $\sin^2\theta \approx 0$, one obtains

$$\begin{aligned} F_1^{eN}(x) &= \frac{4}{9}N_u(x) + \frac{1}{9}N_d(x) + \frac{1}{9}N_s(x) + \frac{4}{9}N_{\bar{u}}(x) + \frac{1}{9}N_{\bar{d}}(x) + \frac{1}{9}N_{\bar{s}}(x), \\ F_1^{\nu N}(x) &= 2N_d(x) + 2N_{\bar{u}}(x). \end{aligned} \tag{3.5.5}$$

For proton and neutron this takes the form

$$\begin{aligned} F_1^{ep}(x) &= \frac{4}{9}p_u(x) + \frac{1}{9}p_d(x) + \frac{1}{9}p_s(x) + \frac{4}{9}p_{\bar{u}}(x) + \frac{1}{9}p_{\bar{d}}(x) + \frac{1}{9}p_{\bar{s}}(x), \\ F_1^{en}(x) &= \frac{4}{9}n_u(x) + \frac{1}{9}n_d(x) + \frac{1}{9}n_s(x) + \frac{4}{9}n_{\bar{u}}(x) + \frac{1}{9}n_{\bar{d}}(x) + \frac{1}{9}n_{\bar{s}}(x) \\ &= \frac{4}{9}p_d(x) + \frac{1}{9}p_u(x) + \frac{1}{9}p_s(x) + \frac{4}{9}p_{\bar{d}}(x) + \frac{1}{9}p_{\bar{u}}(x) + \frac{1}{9}p_{\bar{s}}(x), \\ F_1^{\nu p}(x) &= 2p_d(x) + 2p_{\bar{u}}(x), \\ F_1^{\nu n}(x) &= 2n_d(x) + 2n_{\bar{u}}(x) = 2p_u(x) + 2p_{\bar{d}}(x). \end{aligned} \tag{3.5.6}$$

Now we compute

$$\begin{aligned} &\frac{9}{2}(F_1^{ep}(x) + F_1^{en}(x)) - \frac{3}{4}(F_1^{\nu p}(x) + F_1^{\nu n}(x)) = \\ &\frac{5}{2}p_u(x) + \frac{5}{2}p_d(x) + p_s(x) + \frac{5}{2}p_{\bar{u}}(x) + \frac{5}{2}p_{\bar{d}}(x) + p_{\bar{s}}(x) \\ &-\frac{3}{2}p_u(x) - \frac{3}{2}p_d(x) - \frac{3}{2}p_{\bar{u}}(x) - \frac{3}{2}p_{\bar{d}}(x) = \\ &p_u(x) + p_d(x) + p_s(x) + p_{\bar{u}}(x) + p_{\bar{d}}(x) + p_{\bar{s}}(x), \end{aligned} \tag{3.5.7}$$

and hence

$$I_q = \int_0^1 dx[\frac{9}{2}(F_1^{ep}(x) + F_1^{en}(x)) - \frac{3}{4}(F_1^{\nu p}(x) + F_1^{\nu n}(x))]. \tag{3.5.8}$$

Experimentally one finds $I_q = 0.5$, not 1. The charged partons (quarks) only carry half the momentum of the nucleon. One concludes that there must be other neutral partons carrying the other half of the momentum. In QCD these partons are identified with the gauge bosons of the strong interactions — the gluons. Gluons are indeed electrically neutral, but they are charged under the $SU(3)_C$ gauge group of QCD transforming as color octets. The gluons also carry a substantial fraction of the nucleon spin.

At first it seems paradoxical that, on one hand, quarks and gluons cannot escape from the hadrons as free particles (they are confined), and that, on the other hand, they appear as quasi free partons at high energies. This behavior results from a characteristic property of QCD — its so-called asymptotic freedom. As opposed to QED the color charge is not screened at large distances. Instead it gets enhanced. At short distances, on the other hand, the strong interaction becomes weaker, and quarks and gluons behave like free particles. At large distances the coupling grows and quarks and gluons are confined. In the next chapter we will construct QCD as an $SU(3)_C$ gauge theory, and we will then come back to the property of asymptotic freedom.

Chapter 4

Quantum Chromodynamics

4.1 A Fundamental Theory for Interacting Quarks and Gluons

QCD is a relativistic quantum field theory with a non-Abelian $SU(3)_C$ gauge symmetry, describing the interaction between quarks via gluon exchange. QCD is formulated in terms of quark and gluon fields. Still, the observed particles are not directly quanta of these fields — namely quarks and gluons — but hadrons (confined multi-quark and gluon states). This means that a perturbative treatment of QCD with free quarks and gluons as in and out states is inappropriate. Confinement is a complicated nonperturbative phenomenon that could not yet be derived from QCD first principles. A nonperturbative formulation of QCD is provided by lattice gauge theory. Its numerical treatment (Monte-Carlo simulation) has confirmed various nonperturbative phenomena — among them confinement. In recent years the results have become more quantitative, but still QCD is far from being well tested nonperturbatively.

Fortunately, there are aspects of the strong interactions that can be understood using QCD perturbation theory. This is due to the property of asymptotic freedom — the fact that quarks and gluons behave like free particles at high energies. In this way one can derive the parton model from QCD, identifying the partons with quarks and gluons. Of course,

quarks and gluons do interact as long as the energy is finite. This leads to small violations of Bjorken scaling, that can be understood in perturbative QCD. Performing these calculations goes beyond the scope of this course, in which we concentrate on nonperturbative questions. We will therefore not discuss the perturbative quantization of a non-Abelian gauge theory (e.g. by introducing Faddeev-Popov ghost fields). We restrict ourselves to the construction of the classical QCD Lagrange function, and then we return to the property of asymptotic freedom.

4.2 Construction of the QCD Lagrange function

We have introduced the quarks as elementary fermions with flavor and color quantum numbers. In a relativistic field theory fermions are described by a Dirac spinor field $\Psi(x)$, where x is a space-time point in Minkowski space. The Dirac equation for a hypothetical free quark with flavor $f \in \{u, d, s, c, b, t, \}$ and color $c \in \{r, g, b\}$ is given by

$$(i\gamma^\mu \partial_\mu - m_f)\Psi_{fc}(x) = 0. \tag{4.2.1}$$

Here m_f is the flavor dependent (but color independent) quark mass. The Dirac equation follows from the Lagrange density

$$\mathcal{L}[\bar{\Psi}\Psi] = \sum_{f,c} \bar{\Psi}_{fc}(x)(i\gamma^\mu \partial_\mu - m_f)\Psi_{fc}(x) = \sum_c \bar{\Psi}_c(x)(i\gamma^\mu \partial_\mu - \mathcal{M})\Psi_c(x). \tag{4.2.2}$$

Here we have introduced the quark mass matrix

$$\mathcal{M} = \mathrm{diag}(m_u, m_d, m_s, m_c, m_b, m_t). \tag{4.2.3}$$

Since quarks of different color have the same mass, the above Lagrange function has a large symmetry. We can rotate each flavor f by a matrix U^f in color space

$$\Psi'_{fc'}(x) = U^f_{c'c}\Psi_{fc}(x) \text{ or } \Psi'_f(x) = U^f \Psi_f(x). \tag{4.2.4}$$

One has

$$\bar{\Psi}_f(x) = \Psi^+_f(x)\gamma^0, \tag{4.2.5}$$

and hence

$$\bar{\Psi}'_f(x) = \Psi'^{+}_f(x)\gamma^0 = [U^f \Psi_f(x)]^+ \gamma^0 = \Psi^+_f(x) U^{f+}\gamma^0 = \bar{\Psi}_f(x) U^{f+}. \quad (4.2.6)$$

The Lagrange function remains invariant if

$$U^{f+}U^f = 1 \;\Rightarrow\; U^f \in U(3). \quad (4.2.7)$$

Each flavor can be rotated independently. Hence the symmetry of the free Lagrange function is

$$\otimes_f U(3)_f = U(3)_u \otimes U(3)_d \otimes U(3)_s \otimes U(3)_c \otimes U(3)_b \otimes U(3)_t \quad (4.2.8)$$

The weak interactions mix the flavors, such that they can no longer be rotated independently. In fact, they break the above symmetry explicitly down to

$$\otimes_f U(3)_f \to U(3) \otimes U(1)_{em}. \quad (4.2.9)$$

The remaining $U(3)$ symmetry rotates all flavors in the same way

$$\Psi'_f(x) = U\Psi_f(x). \quad (4.2.10)$$

$U(1)_{em}$ is the symmetry of electromagnetism, it rotates the quark spinors by a complex phase depending on their electric charge. The remaining $U(3)$ symmetry can be decomposed in a $U(1)_B$ and an $SU(3)_C$ symmetry

$$U(3) = U(1)_B \otimes SU(3)_C. \quad (4.2.11)$$

A $U(1)_B$ transformation multiplies all quark spinor by the same phase

$$\Psi'_{fc}(x) = \exp(i\varphi_B)\Psi_{fc}(x). \quad (4.2.12)$$

The invariance of the free Lagrange function leads to a conserved quantum number — the baryon number. When we add the weak interactions, the classical Lagrange density is still $U(1)_B$ invariant. However, baryon number is not conserved in the Standard model, because it does not survive the quantization of the theory. In fact, there is an anomaly in the $U(1)_B$ symmetry, which corresponds to an explicit symmetry breaking. The breaking becomes appreciable only at very high energies in the TeV range, but would then lead to baryon number violating processes in the Standard model. Hence $U(1)_B$ is explicitly broken by the weak interactions and is not an exact symmetry. Only $SU(3)_C$ remains as an exact symmetry of the Standard model even after quantization.

Exact symmetries are very special in physics. Their existence points to something truly fundamental. Approximate symmetries, on the other hand, are 'accidental'. A good example is isospin: the symmetry is due to the fact that u and d quarks have almost the same mass on the typical QCD scale Λ_{QCD}. The values of the quark masses are determined from Yukawa couplings to the Higgs field, which have nothing to do with Λ_{QCD}. Only by accident the u and d quark masses are small compared to it. Exact symmetries, on the other hand, should not be accidental. Instead they should be due to a dynamical principle. Such a dynamical principle is the expectation that exact symmetries should be realized locally. This so-called gauge principle says, that the fundamental laws of Nature should be invariant against symmetry transformations that different observers perform at different points at different times independently. The gauge principle follows from the principle of locality. Gauge symmetries have drastic dynamical consequences. They imply the existence of a gauge field, and hence lead to a nontrivial fundamental interaction. The most well known example is quantum electrodynamics (QED): the requirement that the complex phase of a charged field can be changed locally leads to the existence of the photon, and hence to the electromagnetic interaction itself. Also general relativity is a gauge theory: gravity follows from the requirement that local observers can perform Lorentz transformations independent of one another. Finally, also the weak and strong interactions are associated with an $SU(2)_L$ and an $SU(3)_C$ gauge symmetry.

Now we want to derive the QCD Lagrange density from the gauge principle for the exact $SU(3)_C$ symmetry. For this purpose we consider transformations

$$U(x) \in SU(3)_C,\ U^+(x)U(x) = U(x)U^+(x) = 1,\ \mathrm{det}U(x) = 1, \quad (4.2.13)$$

and we now transform the quark spinors locally, i.e. we perform a gauge transformation

$$\Psi'_f(x) = U(x)\Psi_f(x). \quad (4.2.14)$$

The Lagrange density of the gauge transformed fields is then given by

$$\begin{aligned} \mathcal{L}[\bar{\Psi}', \Psi'] &= \sum_f \bar{\Psi}_f(x)U^+(x)(i\gamma^\mu\partial_\mu - m_f)U(x)\Psi_f(x) \\ &= \sum_f \bar{\Psi}_f(x)(i\gamma^\mu\partial_\mu + i\gamma^\mu U^+(x)\partial_\mu U(x) - m_f)\Psi_f(x). \end{aligned} \quad (4.2.15)$$

Due to the term $i\gamma^\mu U^+(x)\partial_\mu U(x)$ the Lagrange density is not gauge invariant. To investigate this term further we write

$$U(x) = \exp(iH(x)), \tag{4.2.16}$$

where $H(x) \in su(3)_C$ is an element of the algebra, i.e.

$$H^+(x) = H(x),\ \mathrm{Tr}H(x) = 0. \tag{4.2.17}$$

It is instructive to convince oneself that $iU^+(x)\partial_\mu U(x)$ is an element of the algebra. Next we introduce an algebra-valued gauge potential

$$G_\mu(x) = ig_s G^a_\mu(x)\lambda_a,\ G^a_\mu(x) \in R,\ a \in \{1, 2, ..., 8\}, \tag{4.2.18}$$

whose gauge variation is supposed to cancel $iU^+(x)\partial_\mu U(x)$. Here g_s is the dimensionless gauge coupling constant of the strong interactions. We postulate the following behavior under gauge transformations

$$G'_\mu(x) = U(x)(G_\mu(x) + \partial_\mu)U^+(x), \tag{4.2.19}$$

because then the modified Lagrange function

$$\mathcal{L}[\bar{\Psi}, \Psi, G_\mu] = \sum_f \bar{\Psi}_f(x)(i\gamma^\mu(G_\mu + \partial_\mu) - m_f)\Psi_f(x) \tag{4.2.20}$$

is gauge invariant. It is a good exercise to show this explicitly. $G_\mu(x)$ is the non-Abelian analog of the gauge potential $A_\mu(x)$ of electrodynamics. From electrodynamics we know the Lagrange density of the free electromagnetic field

$$\mathcal{L}[A_\mu] = -\frac{1}{4}F^{\mu\nu}(x)F_{\mu\nu}(x), \tag{4.2.21}$$

that is given in terms of the field strength tensor

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x). \tag{4.2.22}$$

The corresponding term is still missing in the QCD Lagrange density. It is remarkable that $\partial_\mu G_\nu(x) - \partial_\nu G_\mu(x)$ has no definite behavior under gauge transformations. Due to the non-Abelian nature the field strength contains an additional term

$$G_{\mu\nu}(x) = \partial_\mu G_\nu(x) - \partial_\nu G_\mu(x) + [G_\mu(x), G_\nu(x)]. \tag{4.2.23}$$

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

$$G'_{\mu\nu}(x) = U(x)G_{\mu\nu}(x)U^{+}(x). \tag{4.2.24}$$

The full Lagrange function of QCD finally takes the form

$$\mathcal{L}_{QCD}[\bar{\Psi},\Psi,G_\mu] = \sum_f \bar{\Psi}_f(x)(i\gamma^\mu(G_\mu+\partial_\mu)-m_f)\Psi_f(x)-\frac{1}{2g_s^2}\mathrm{Tr}G^{\mu\nu}(x)G_{\mu\nu}(x). \tag{4.2.25}$$

It is instructive to show that $\mathcal{L}_{QCD}[\bar{\Psi},\Psi,G_\mu]$ is indeed gauge invariant.

Using gauge invariance as the guiding principle one can construct another term in the QCD Lagrange function

$$\mathcal{L}^{\theta}_{QCD}[\bar{\Psi},\Psi,G_\mu] = \mathcal{L}_{QCD}[\bar{\Psi},\Psi,G_\mu] + \frac{\theta}{32\pi^2}\varepsilon^{\mu\nu\rho\sigma}\mathrm{Tr}G_{\mu\nu}(x)G_{\rho\sigma}. \tag{4.2.26}$$

The role of the prefactor $32\pi^2$ will become clear when we discuss the $U(1)$-problem. The parameter θ is the vacuum angle of QCD. For $\theta \neq 0$ the CP symmetry would be explicitly broken in the strong interactions (as it is in fact the case for the weak interactions). This would lead to a nonvanishing electric dipole moment of the neutron. The corresponding experimental result is compatible with zero, and one concludes that

$$|\theta| < 10^{-9}. \tag{4.2.27}$$

The question arises why θ is compatible with zero. This is the so-called strong CP problem, which is still unsolved, although various explanations have been suggested. The solution of the strong CP problem goes most probably beyond QCD, and perhaps beyond the Standard model. Therefore it is of no concern for us in this course.

If one wants to construct a renormalizable QCD theory in four dimensions, one cannot add any further terms to the Lagrangian. This follows from gauge invariance and the dimensions of quark and gluon fields.

An important difference between Abelian and non-Abelian gauge theories is that in a non-Abelian gauge theory the gauge fields are themselves charged. In fact, they transform in the adjoint representation of the gauge group — the $SU(3)_C$ octet in case of QCD. Correspondingly, there are eight

gluons $G^a_\mu(x)$, $a \in \{1, 2, ..., 8\}$. 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 $G_\mu(x)$ give rise to the free gluon propagator. Due to the commutator term, however, there are also terms cubic and quartic in $G_\mu(x)$, that lead to the gluon self interaction vertices. This is illustrated in fig.4.1. Correspondingly, there is a free

Figure 4.1: *The gluon propagator and the three and four gluon vertex, as well as the quark propagator and the quark-gluon vertex and the ghost propagator and the ghost-gluon vertex.*

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 $\partial^\mu G_\mu(x) = 0$ 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.

4.3 Asymptotic Freedom

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

$$Z = \int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi\mathcal{D}G \exp(-i \int d^4x \; \mathcal{L}_{QCD}[\bar{\Psi}, \Psi, G_\mu]) \tag{4.3.1}$$

for the QCD path integral is undefined, i.e. divergent, until it is regularized, i.e. made finite, 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 divergencies quarks and gluon fields are renormalized

$$\Psi(x) = Z_\Psi(\varepsilon)^{1/2}\Psi^r(x), \; G_\mu(x) = Z_G(\varepsilon)^{1/2}G^r_\mu(x), \tag{4.3.2}$$

and also the coupling constant is renormalized via

$$g_s = \frac{Z(\varepsilon)}{Z_\Psi(\varepsilon)Z_G(\varepsilon)^{1/2}} g^r_s. \tag{4.3.3}$$

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

$$\Gamma^r_{n_\Psi, n_G}(k_i, p_j) = \lim_{\varepsilon \to 0} Z_\Psi(\varepsilon)^{n_\Psi/2} Z_G(\varepsilon)^{n_G/2} \Gamma_{n_\Psi, n_G}(k_i, p_j, \varepsilon). \tag{4.3.4}$$

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 so-called renormalization conditions. In QCD this can be done using the

Figure 4.2: *Leading Feynman diagrams for the gluon propagator, the quark propagator, and the quark-gluon vertex.*

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. To leading order in the gauge coupling the corresponding Feynman diagrams are shown in fig.4.2. 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

$$\begin{aligned}
&\Gamma^r_{0,2}(p,-p)^{\mu\nu}_{ab}|_{p^2=-\mathcal{M}^2} = i(-g_{\mu\nu}p^2 + p^\mu p^\nu)\delta_{ab},\\
&\Gamma^r_{2,0}(k,k)|_{k^2=-\mathcal{M}^2} = i\gamma^\mu k_\mu,\\
&\Gamma^r_{2,1}(k,k,k)^\mu_a|_{k^2=-\mathcal{M}^2} = -ig^r_s\frac{\lambda_a}{2}\gamma^\mu.
\end{aligned} \tag{4.3.5}$$

The renormalized vertex functions are functions of the renormalized coupling constant g^r_s 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

$$g^r_s = g^r_s(g,\varepsilon,\mathcal{M}). \tag{4.3.6}$$

This relation defines the so-called β-function

$$\beta(g^r_s) = \lim_{\varepsilon\to 0}\mathcal{M}\frac{\partial}{\partial\mathcal{M}}g^r_s(g,\varepsilon,\mathcal{M}). \tag{4.3.7}$$

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

$$\beta(g_s^r) = -\frac{(g_s^r)^3}{16\pi^2}(11 - \frac{2}{3}N_f). \tag{4.3.8}$$

Here N_f is the number of quark flavors (massless quarks have been assumed). 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

$$11 - \frac{2}{3}N_f > 0 \;\Rightarrow\; N_f \leq 16, \tag{4.3.9}$$

i.e. for not too many flavors, the β-function is negative close to the fixed point, as illustrated in fig.4.3. This behavior is known as asymptotic free-

Figure 4.3: *The QCD β-function.*

dom. 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

$$\begin{aligned}
&\beta(g_s^r) = \mathcal{M}\frac{\partial}{\partial\mathcal{M}}g_s^r = -\frac{(g_s^r)^3}{16\pi^2}(11 - \frac{2}{3}N_f) \;\Rightarrow \\
&\frac{\partial g_s^r}{\partial\mathcal{M}}/(g_s^r)^3 = \frac{1}{2}\frac{\partial(g_s^r)^2}{\partial\mathcal{M}}/(g_s^r)^4 = -\frac{11 - \frac{2}{3}N_f}{16\pi^2}\frac{1}{\mathcal{M}} \;\Rightarrow \\
&\frac{\partial(g_s^r)^2}{(g_s^r)^4} = -\frac{33 - 2N_f}{24\pi^2}\frac{\partial\mathcal{M}}{\mathcal{M}} \;\Rightarrow\; \frac{1}{(g_s^r)^2} = \frac{33 - 2N_f}{24\pi^2}\log\frac{\mathcal{M}}{\Lambda_{QCD}}.
\end{aligned} \tag{4.3.10}$$

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

$$\alpha_s^r = \frac{(g_s^r)^2}{4\pi}, \tag{4.3.11}$$

we obtain

$$\alpha_s^r(\mathcal{M}) = \frac{6\pi}{33 - 2N_f} \frac{1}{\log(\mathcal{M}/\Lambda_{QCD})}. \tag{4.3.12}$$

At high energy scales $\mathcal{M}$ the renormalized coupling constant slowly (i.e. logarithmically) goes to zero. Hence the quarks then behave as free particles. This confirms the central assumption of the parton model, such that we can finally identify the partons as quarks and gluons.

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.

Chapter 5

Spontaneous Chiral Symmetry Breaking

5.1 Chiral Symmetry

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. In the next chapter we 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 so-called 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 so-called chiral perturbation theory.

5.2 Chiral Transformations and Symmetry Breaking Terms

Let us consider the quark part of the QCD Lagrange density

$$\mathcal{L}[\bar{\Psi}, \Psi, G_\mu] = \bar{\Psi}(x)(i\gamma^\mu(G_\mu(x) + \partial_\mu) - \mathcal{M})\Psi(x). \qquad (5.2.1)$$

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

$$\Psi_R(x) = \frac{1}{2}(1+\gamma_5)\Psi(x),\ \Psi_L(x) = \frac{1}{2}(1-\gamma_5)\Psi(x),\ \Psi(x) = \Psi_R(x) + \Psi_L(x). \tag{5.2.2}$$

Here we have used

$$\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3,\ \{\gamma^\mu,\gamma^\nu\} = 2g_{\mu\nu},\ \{\gamma^\mu,\gamma_5\} = 0. \tag{5.2.3}$$

Next we consider the adjoint spinors

$$\begin{aligned}
\bar\Psi_R(x) &= \Psi_R(x)^+\gamma^0 = \Psi(x)^+\frac{1}{2}(1+\gamma_5^+)\gamma^0 = \Psi(x)^+\gamma^0\frac{1}{2}(1-\gamma_5) \\
&= \bar\Psi(x)\frac{1}{2}(1-\gamma_5), \\
\bar\Psi_L(x) &= \Psi_L(x)^+\gamma^0 = \Psi(x)^+\frac{1}{2}(1-\gamma_5^+)\gamma^0 = \Psi(x)^+\gamma^0\frac{1}{2}(1+\gamma_5) \\
&= \bar\Psi(x)\frac{1}{2}(1+\gamma_5).
\end{aligned} \tag{5.2.4}$$

Here we used

$$\gamma^0\gamma_5^+\gamma^0 = -\gamma_5. \tag{5.2.5}$$

Inserting the decomposed spinors in the Lagrange density we obtain

$$\mathcal{L}[\bar\Psi,\Psi,G_\mu] = (\bar\Psi_R(x)+\bar\Psi_L(x))(i\gamma^\mu(G_\mu(x)+\partial_\mu) - \mathcal{M})(\Psi_R(x)+\Psi_L(x)). \tag{5.2.6}$$

First, we investigate the γ^μ term

$$\begin{aligned}
&\bar\Psi_R(x)i\gamma^\mu(G_\mu(x)+\partial_\mu)\Psi_L(x) \\
&= \bar\Psi(x)\frac{1}{2}(1-\gamma_5)i\gamma^\mu(G_\mu(x)+\partial_\mu)\frac{1}{2}(1-\gamma_5)\Psi(x) \\
&= \bar\Psi(x)\frac{1}{4}(1-\gamma_5)(1+\gamma_5)i\gamma^\mu(G_\mu(x)+\partial_\mu)\Psi(x) = 0.
\end{aligned} \tag{5.2.7}$$

On the other hand, for the mass term one finds

$$\begin{aligned}
\bar\Psi_R(x)\mathcal{M}\Psi_R(x) &= \bar\Psi(x)\frac{1}{2}(1-\gamma_5)\mathcal{M}\frac{1}{2}(1+\gamma_5)\Psi(x) \\
&= \bar\Psi(x)\frac{1}{4}(1-\gamma_5)(1+\gamma_5)\mathcal{M}\Psi(x) = 0.
\end{aligned} \tag{5.2.8}$$

Hence, we can write

$$\begin{aligned}\mathcal{L}[\bar{\Psi},\Psi,G_\mu] &= \bar{\Psi}_R(x)i\gamma^\mu(G_\mu(x)+\partial_\mu)\Psi_R(x)\\ &+ \bar{\Psi}_L(x)i\gamma^\mu(G_\mu(x)+\partial_\mu)\Psi_L(x)\\ &- \bar{\Psi}_R(x)\mathcal{M}\Psi_L(x)-\bar{\Psi}_L(x)\mathcal{M}\Psi_R(x). \qquad (5.2.9)\end{aligned}$$

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

$$\begin{aligned}&\Psi'_R(x)=R\Psi_R(x),\ \bar{\Psi}'(x)=\bar{\Psi}_R(x)R^+,\ R\in U(N_f)_R,\\ &\Psi'_L(x)=L\Psi_L(x),\ \bar{\Psi}'(x)=\bar{\Psi}_L(x)L^+,\ L\in U(N_f)_L. \qquad (5.2.10)\end{aligned}$$

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

$$SU(N_f)_L\otimes SU(N_f)_R\otimes U(1)_{L=R}=SU(N_f)_L\otimes SU(N_f)_R\otimes U(1)_B. \quad (5.2.11)$$

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 $\mathcal{M}=m1$, one would have

$$\bar{\Psi}'_R(x)\mathcal{M}\Psi'_L(x)=\bar{\Psi}_R(x)R^+m1L\Psi_L(x)=\bar{\Psi}_R(x)R^+L\mathcal{M}\Psi_L(x). \quad (5.2.12)$$

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

$$SU(N_f)_{L=R}\otimes U(1)_{L=R}=SU(N_f)_F\otimes U(1)_B, \quad (5.2.13)$$

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

$$\otimes_f U(1)_f=U(1)_u\otimes U(1)_d\otimes U(1)_s. \quad (5.2.14)$$

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.

5.3 The Spontaneous Breakdown of Chiral Symmetry

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

$$\begin{aligned} J_\mu^{La}(x) &= \bar{\Psi}_L(x)\gamma_\mu \frac{\sigma^a}{2}\Psi_L(x), \\ J_\mu^{Ra}(x) &= \bar{\Psi}_R(x)\gamma_\mu \frac{\sigma^a}{2}\Psi_R(x), \end{aligned} \tag{5.3.1}$$

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

$$\begin{aligned} V_\mu^a(x) &= J_\mu^{La}(x) + J_\mu^{Ra}(x) \\ &= \bar{\Psi}(x)\frac{1}{2}(1+\gamma_5)\gamma_\mu \frac{\sigma^a}{2}\frac{1}{2}(1-\gamma_5)\Psi(x) \\ &+ \bar{\Psi}(x)\frac{1}{2}(1-\gamma_5)\gamma_\mu \frac{\sigma^a}{2}\frac{1}{2}(1+\gamma_5)\Psi(x) \\ &= \bar{\Psi}(x)\frac{1}{2}(1+\gamma_5)\gamma_\mu \frac{\sigma^a}{2}\Psi(x) + \bar{\Psi}(x)\frac{1}{2}(1-\gamma_5)\gamma_\mu \frac{\sigma^a}{2}\Psi(x) \\ &= \bar{\Psi}(x)\gamma_\mu \frac{\sigma^a}{2}\Psi(x), \\ A_\mu^a(x) &= J_\mu^{La}(x) - J_\mu^{Ra}(x) = \bar{\Psi}(x)\gamma_5\gamma_\mu \frac{\sigma^a}{2}\Psi(x). \end{aligned} \tag{5.3.2}$$

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

$$\langle\Phi|J_\mu^{La}(x)J_\nu^{Rb}(y)|\Phi\rangle = \langle\Phi|J_\mu^{Ra}(x)J_\nu^{Lb}(y)|\Phi\rangle = 0, \tag{5.3.3}$$

and hence

$$\langle\Phi|V_\mu^a(x)V_\nu^b(y)|\Phi\rangle = \langle\Phi|A_\mu^a(x)A_\nu^b(y)|\Phi\rangle. \tag{5.3.4}$$

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

$$\begin{aligned} Q_V^a|0\rangle &= \int d^3x V_0^a(\vec{x},0)|0\rangle, \\ Q_A^a|0\rangle &= \int d^3x A_0^a(\vec{x},0)|0\rangle, \end{aligned} \tag{5.3.5}$$

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

$$Q_V^a|\Phi\rangle = Q_A^a|\Phi\rangle = 0. \tag{5.3.6}$$

The real QCD vacuum is not chirally invariant because

$$Q_A^a|0\rangle \neq 0. \tag{5.3.7}$$

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

$$[H_{QCD}, Q_A^a] = 0. \tag{5.3.8}$$

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

$$H_{QCD}Q_A^a|0\rangle = Q_A^a H_{QCD}|0\rangle = 0 \tag{5.3.9}$$

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

$$\langle\bar{\Psi}\Psi\rangle = \langle 0|\bar{\Psi}(x)\Psi(x)|0\rangle = \langle 0|\bar{\Psi}_R(x)\Psi_L(x) + \bar{\Psi}_L(x)\Psi_R(x)|0\rangle. \quad (5.3.10)$$

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, $\langle\bar{\Psi}\Psi\rangle \neq 0$.

5.4 Low Energy Effective Theory and Chiral Perturbation Theory

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

$$\begin{aligned} G &= SU(N_f)_L \otimes SU(N_f)_R \otimes U(1)_B, \\ H &= SU(N_f)_{L=R} \otimes U(1)_B, \end{aligned} \quad (5.4.1)$$

and the corresponding coset space is

$$G/H = SU(N_f). \quad (5.4.2)$$

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

$$U(x)' = LU(x)R^+. \quad (5.4.3)$$

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

$$\partial_\mu U(x)' = L\partial_\mu U(x)R^+, \quad (5.4.4)$$

and we form a chirally invariant Lorentz scalar

$$\mathcal{L}[U] = \frac{f_\pi^2}{4}\mathrm{Tr}(\partial^\mu U(x)^+ \partial_\mu U(x)). \tag{5.4.5}$$

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

$$\begin{aligned}\mathcal{L}[U'] &= \frac{f_\pi^2}{4}\mathrm{Tr}(\partial^\mu U(x)'^+ \partial_\mu U(x)') \\ &= \frac{f_\pi^2}{4}\mathrm{Tr}(R\partial^\mu U(x)^+ L^+ L\partial_\mu U(x) R^+) = \mathcal{L}[U].\end{aligned} \tag{5.4.6}$$

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

$$\mathcal{L}[U] = \frac{f_\pi^2}{4}\mathrm{Tr}(\partial^\mu U(x)^+ \partial_\mu U(x)) + c\mathrm{Tr}(\mathcal{M}(U + U^+)). \tag{5.4.7}$$

Under chiral transformations the additional term transforms into

$$\mathrm{Tr}(\mathcal{M}(U(x)' + U(x)'^+)) = \mathrm{Tr}(\mathcal{M}(LU(x)R^+ + RU(x)^+ L^+)). \tag{5.4.8}$$

If all quark masses are equal, i.e. for $\mathcal{M} = m1$, the Lagrange density is again invariant against rotations $R^+L = 1$ 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

$$\langle 0|\frac{\partial}{\partial m_f}\mathcal{L}|_{\mathcal{M}=0}|0\rangle = \langle 0| - \bar{\Psi}_f \Psi_f|0\rangle = -\frac{1}{N_f}\langle\bar{\Psi}\Psi\rangle. \tag{5.4.9}$$

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

$$U(x) = 1. \tag{5.4.10}$$

Hence, in the effective theory we have

$$\langle 0|\frac{\partial}{\partial m_f}\mathcal{L}|_{\mathcal{M}=0}|0\rangle = c\mathrm{Tr}(\mathrm{diag}(1,0,...,0)(1+1)) = 2c, \tag{5.4.11}$$

and therefore

$$c = -\frac{1}{2N_f}\langle\bar{\Psi}\Psi\rangle. \tag{5.4.12}$$

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, and they contain new 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

$$\mathcal{L}[U] = \frac{f_\pi^2}{4}\mathrm{Tr}(\partial^\mu U(x)^+ \partial_\mu U(x)) - \frac{1}{2N_f}\langle\bar{\Psi}\Psi\rangle\mathrm{Tr}(\mathcal{M}(U+U^+)). \tag{5.4.13}$$

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

$$U(x) = \exp(i\pi^a(x)\eta_a/f_\pi),\ a \in \{1, 2, ..., N_f^2 - 1\}, \tag{5.4.14}$$

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

$$U(x) = 1 + i\pi^a(x)\eta_a/f_\pi,\ \partial_\mu U(x) = i\partial_\mu\pi^a(x)\eta_a/f_\pi, \tag{5.4.15}$$

and hence

$$\begin{aligned}\mathcal{L}[U] &= \frac{1}{4}\mathrm{Tr}(\partial^\mu\pi^a(x)\eta_a\partial_\mu\pi^b(x)\eta_b) - \frac{1}{2N_f}\langle\bar{\Psi}\Psi\rangle\mathrm{Tr}(\mathcal{M}(1+1)) \\ &= \frac{1}{2}\partial^\mu\pi^a(x)\partial_\mu\pi^b(x)\delta_{ab} - \frac{1}{N_f}\langle\bar{\Psi}\Psi\rangle\mathrm{Tr}\mathcal{M}.\end{aligned} \tag{5.4.16}$$

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

$$U(x) = 1 + i\pi^a(x)\eta_a/f_\pi + \frac{1}{2}(i\pi^a(x)\eta_a/f_\pi)^2, \tag{5.4.17}$$

such that for quarks with equal masses, i.e. for $\mathcal{M} = m1$,

$$\mathrm{Tr}(\mathcal{M}(U(x) + U(x)^+)) = m\mathrm{Tr}(U(x) + U(x)^+)$$

$$
\begin{aligned}
&= 2mN_f - m\frac{1}{f_\pi^2}\pi^a(x)\pi^b(x)\mathrm{Tr}(\eta_a\eta_b) \\
&= 2mN_f - 2m\frac{1}{f_\pi^2}\pi^a(x)\pi^a(x). \qquad (5.4.18)
\end{aligned}
$$

Altogether we obtain

$$
\mathcal{L}[U] = \frac{1}{2}\partial^\mu\pi^a(x)\partial_\mu\pi^a(x) - \frac{1}{N_f}\langle\bar{\Psi}\Psi\rangle(mN_f - m\frac{1}{f_\pi^2}\pi^a(x)\pi^a(x)). \quad (5.4.19)
$$

The resulting equation of motion is given by

$$
\partial^\mu\frac{\delta\mathcal{L}}{\delta\partial^\mu\pi^a} - \frac{\delta\mathcal{L}}{\delta\pi^a} = \partial^\mu\partial_\mu\pi^a(x) - \frac{2m\langle\bar{\Psi}\Psi\rangle}{N_f f_\pi^2}\pi^a(x) = 0. \qquad (5.4.20)
$$

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

$$
M_\pi^2 = \frac{2m\langle\bar{\Psi}\Psi\rangle}{N_f f_\pi^2}. \qquad (5.4.21)
$$

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 very instructive 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

$$
\begin{aligned}
M_{K^+}^2 &= M_{K^-}^2 = \frac{(m_u + m_s)\langle\bar{\Psi}\Psi\rangle}{N_f f_\pi^2}, \\
M_{K^0}^2 &= M_{\bar{K}^0}^2 = \frac{(m_d + m_s)\langle\bar{\Psi}\Psi\rangle}{N_f f_\pi^2}, \\
M_\eta^2 = \frac{1}{3}&\frac{(m_u + m_d + 4m_s)\langle\bar{\Psi}\Psi\rangle}{N_f f_\pi^2}. \qquad (5.4.22)
\end{aligned}
$$

This leads to the following mass relation

$$
3M_\eta^2 + M_\pi^2 = 2M_{K^+}^2 + 2M_{K^0}^2, \qquad (5.4.23)
$$

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

$$
m_q = \frac{1}{2}(m_u + m_d) \qquad (5.4.24)
$$

one obtains

$$\frac{M_{K^+}^2 + M_{K^0}^2}{M_\pi^2} = \frac{m_s + m_q}{m_q}, \tag{5.4.25}$$

which yields $m_s/m_q = 24.2$. Similarly

$$\frac{M_\eta^2}{M_\pi^2} = \frac{2m_s + m_q}{3m_q}, \tag{5.4.26}$$

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

$$m_q \approx 0.007\text{GeV}, \tag{5.4.27}$$

and hence

$$m_s \approx 0.16\text{GeV}. \tag{5.4.28}$$

These are so-called current quark masses. They are much smaller than the constituent quark masses. Still, the mass difference is consistent with the value obtained in the constituent quark model.

Chapter 6

Instantons, θ-Vacua and the $U(1)$-Problem

6.1 Why is the η'-Meson no Goldstone Boson?

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 $M_{\eta'} = 0.958\text{GeV}$, 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 people realized that the axial $U(1)$ is not really 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 nontrivial configurations of the gluon field — so-called instantons — give mass to the η'-meson. Instantons are characterized by an integer valued topological invariant — the topological charge. 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. The topological charge has another important effect on the QCD dynamics. The nontrivial topology leads to a periodic vacuum structure that can be characterized by the vacuum angle θ.

6.2 The Anomaly

In the previous chapter we have studied the vector and axial currents

$$V_\mu^a(x) = \bar{\Psi}(x)\gamma_\mu \frac{\sigma^a}{2}\Psi(x),\ A_\mu^a(x) = \bar{\Psi}(x)\gamma_5\gamma_\mu \frac{\sigma^a}{2}\Psi(x), \tag{6.2.1}$$

that transform in the adjoint representation of the flavor symmetry group. For massless quarks these currents are conserved.

$$\partial^\mu V_\mu^a(x) = \partial^\mu A_\mu^a(x) = 0. \tag{6.2.2}$$

Now we also consider the flavor singlet vector and axial currents

$$V_\mu(x) = \bar{\Psi}(x)\gamma_\mu\Psi(x),\ A_\mu(x) = \bar{\Psi}(x)\gamma_5\gamma_\mu\Psi(x). \tag{6.2.3}$$

The flavor singlet vector current is conserved (even for quarks with arbitrary masses)

$$\partial^\mu V_\mu(x) = 0, \tag{6.2.4}$$

and the corresponding charge is the baryon number. The axial charge, however, is not conserved, because the axial current conservation is violated by the Adler-Bell-Jackiw anomaly

$$\partial^\mu A_\mu(x) = 2N_f P(x). \tag{6.2.5}$$

Here N_f is the number of flavors (assuming massless quarks) and

$$P(x) = -\frac{1}{32\pi^2}\varepsilon^{\mu\nu\rho\sigma}\mathrm{Tr}(G_{\mu\nu}(x)G_{\rho\sigma}(x)) \tag{6.2.6}$$

is the so-called Chern-Pontryagin density. The anomaly equation can be derived in perturbation theory. It follows from the Feynman triangle diagram shown in fig.6.1. The Chern-Pontryagin density can be written as a

Figure 6.1: *The triangle graph that yields the anomaly.*

total divergence

$$P(x) = \partial^\mu \Omega^{(0)}_\mu(x), \tag{6.2.7}$$

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

$$\Omega^{(0)}_\mu(x) = -\frac{1}{8\pi^2}\varepsilon^{\mu\nu\rho\sigma}\mathrm{Tr}[G_\nu(x)(\partial_\rho G_\sigma(x) + \frac{2}{3}G_\rho(x)G_\sigma(x))]. \tag{6.2.8}$$

It is a good exercise to convince oneself that this satisfies eq.(6.2.7). We can now construct a conserved current

$$\tilde{A}_\mu(x) = A_\mu(x) - 2N_f\Omega^{(0)}_\mu(x), \tag{6.2.9}$$

because then

$$\partial^\mu \tilde{A}_\mu(x) = \partial^\mu A_\mu(x) - 2N_f P(x) = 0. \tag{6.2.10}$$

One might think that we have found a new axial $U(1)$ symmetry which is free of the anomaly, and that hence the η'-meson should indeed be a Goldstone boson. This is, however, not so, because the current $\tilde{A}_\mu(x)$ contains $\Omega^{(0)}_\mu(x)$ which is not gauge invariant. Although the gauge variant current is formally conserved, this has no gauge invariant physical consequences.

6.3 The Topological Charge

For the rest of this course 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. In the next chapter we will make the transition to Euclidean time in a more systematic manner. Here we just stop distinguishing co- and contravariant indices and we write the topological charge as

$$\begin{aligned} Q &= -\frac{1}{32\pi^2}\int d^4x\ \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}(G_{\mu\nu}(x)G_{\rho\sigma}(x)) = \int d^4x\ P(x) \\ &= \int d^4x\ \partial_\mu \Omega^{(0)}_\mu(x) = \int_{S^3} d^3\sigma_\mu\ \Omega^{(0)}_\mu(x). \end{aligned} \tag{6.3.1}$$

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 gluon 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)

$$G_\mu(x) = U(x)\partial_\mu U(x)^+. \tag{6.3.2}$$

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

$$\begin{aligned} Q &= -\frac{1}{8\pi^2}\int_{S^3} d^3\sigma_\mu\ \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(U(x)\partial_\nu U(x)^+)(\partial_\rho(U(x)\partial_\sigma U(x)^+) \\ &+ \frac{2}{3}(U(x)\partial_\rho U(x)^+)(U(x)\partial_\sigma U(x)^+))] \\ &= -\frac{1}{8\pi^2}\int_{S^3} d^3\sigma_\mu\ \varepsilon_{\mu\nu\rho\sigma} \\ &\times \mathrm{Tr}[-(U(x)\partial_\nu U(x)^+)(U(x)\partial_\rho U(x)^+)(U(x)\partial_\sigma U(x)^+) \\ &+ \frac{2}{3}(U(x)\partial_\nu U(x)^+)(U(x)\partial_\rho U(x)^+)(U(x)\partial_\sigma U(x)^+)] \\ &= \frac{1}{24\pi^2}\int_{S^3} d^3\sigma_\mu\ \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(U(x)\partial_\nu U(x)^+)(U(x)\partial_\rho U(x)^+)(U(x)\partial_\sigma U(x)^+)]. \end{aligned} \tag{6.3.3}$$

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

$$U : S^3 \to SU(3)_c. \tag{6.3.4}$$

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

$$\Pi_3[SU(3)_c] = \mathbf{Z}. \tag{6.3.5}$$

Here the index 3 indicates that we consider mappings of the 3-dimensional sphere S^3. The third homotopy group of $SU(3)$ 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

$$U = VW,\ W = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \tilde{U}_{11} & \tilde{U}_{12} \\ 0 & \tilde{U}_{21} & \tilde{U}_{22} \end{pmatrix}, \tag{6.3.6}$$

where the embedded matrix $\tilde{U}$ is in $SU(2)$. It is indirectly defined by

$$V = \begin{pmatrix} U_{11} & -U_{21}^* & -\frac{U_{31}^*(1+U_{11})}{1+U_{11}^*} \\ U_{21} & \frac{1+U_{11}^*-|U_{21}|^2}{1+U_{11}} & -\frac{U_{31}^*U_{21}}{1+U_{11}^*} \\ U_{31} & -\frac{U_{21}^*U_{31}}{1+U_{11}} & \frac{1+U_{11}^*-|U_{31}|^2}{1+U_{11}^*} \end{pmatrix} \in SU(3). \tag{6.3.7}$$

The matrix V is constructed entirely from the elements U_{11}, U_{21} and U_{31} of the first column of the matrix U. One should convince oneself that V is indeed an $SU(3)$ matrix, and that the resulting matrix $\tilde{U}$ is indeed in $SU(2)$. The idea now is to reduce the expression for the topological charge from $SU(3)$ to $SU(2)$ by using the formula

$$\begin{aligned} & \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(VW)\partial_\nu(VW)^+(VW)\partial_\rho(VW)^+(VW)\partial_\sigma(VW)^+] = \\ & \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(V\partial_\nu V^+)(V\partial_\rho V^+)(V\partial_\sigma V^+) \\ + \ & \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(W\partial_\nu W^+)(W\partial_\rho W^+)(W\partial_\sigma W^+)] \\ & +3\partial_\nu\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[((V\partial_\rho V^+)(W\partial_\sigma W^+)]. \end{aligned} \tag{6.3.8}$$

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

$$\begin{aligned} Q &= \frac{1}{24\pi^2}\int_{S^3} d^3\sigma_\mu \varepsilon_{\mu\nu\rho\sigma} \mathrm{Tr}[(U(x)\partial_\nu U(x)^+)(U(x)\partial_\rho U(x)^+)(U(x)\partial_\sigma U(x)^+)] \\ &= \frac{1}{24\pi^2}\int_{S^3} d^3\sigma_\mu \varepsilon_{\mu\nu\rho\sigma} \mathrm{Tr}[(V(x)\partial_\nu V(x)^+)(V(x)\partial_\rho V(x)^+)(V(x)\partial_\sigma V(x)^+) \\ &+ (W(x)\partial_\nu W(x)^+)(W(x)\partial_\rho W(x)^+)(W(x)\partial_\sigma W(x)^+)]. \end{aligned} \tag{6.3.9}$$

The ∂_ν-term of the formula eq.(6.3.8) 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 U_{11}, U_{21} and U_{31}, it can be viewed as a mapping of S^3 into the sphere S^5

$$V : S^3 \to S^5. \tag{6.3.10}$$

This is because $|U_{11}|^2 + |U_{21}|^2 + |U_{31}|^2 = 1$. Remarkably the corresponding homotopy group is trivial, i.e.

$$\Pi_3[S^5] = \{0\}. \tag{6.3.11}$$

All mappings of S^3 into the higher dimensional sphere S^5 are topologically equivalent (they can be deformed into each other). This can be understood better in a lower dimensional example

$$\Pi_1[S^2] = \{0\}. \tag{6.3.12}$$

Each closed curve on an ordinary sphere can be constricted to the north pole, and hence is topologically trivial. This is illustrated in fig.6.2. Since the mapping V is topologically trivial its contribution to the topological charge vanishes. The remaining W term reduces to the $SU(2)$ contribution

$$Q = \frac{1}{24\pi^2}\int_{S^3} d^3\sigma_\mu \ \varepsilon_{\mu\nu\rho\sigma} \mathrm{Tr}[(\tilde{U}(x)\partial_\nu \tilde{U}(x)^+)(\tilde{U}(x)\partial_\rho \tilde{U}(x)^+)(\tilde{U}(x)\partial_\sigma \tilde{U}(x)^+)]. \tag{6.3.13}$$

The separation of the V contribution works only if the decomposition of U into V and $\tilde{U}$ is non-singular. In fact, the expression for V is singular for $U_{11} = -1$. This corresponds to a 3-dimensional subspace of the 8-dimensional $SU(3)$ group space. The mapping U itself covers a 3-d subspace of $SU(3)$. Hence it is arbitrarily improbable to hit a singularity (it is of zero

Figure 6.2: *All mappings of a circle to a sphere are topologically trivial.*

measure). Since we have now reduced the $SU(3)$ topological charge to the $SU(2)$ case, it remains to be shown that the $SU(2)$ expression is actually an integer. First of all

$$\tilde{U} : S^3 \to SU(2) = S^3, \tag{6.3.14}$$

and indeed

$$\Pi_3[SU(2)] = \Pi_3[S^3] = \mathbf{Z}. \tag{6.3.15}$$

The topological charge specifies how often the $SU(2)$ group space (which is isomorphic to the 3-sphere) is covered by $\tilde{U}$ 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

$$U = \exp(i\varphi) : S^1 \to U(1) = S^1. \tag{6.3.16}$$

The relevant homotopy group is

$$\Pi_1[U(1)] = \Pi_1[S^1] = \mathbf{Z}. \tag{6.3.17}$$

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. An example with $Q = 2$ is shown in figure 6.3. In $U(1)$ the expression for the topological

Figure 6.3: *A mapping from S^1 to $U(1)$ with topological charge $Q = 2$.*

charge is analogous to the one in $SU(N)$

$$\begin{aligned} Q &= -\frac{1}{2\pi}\int_{S^1} d\sigma_\mu \varepsilon_{\mu\nu}(U(x)\partial_\nu U(x)^+) = \frac{1}{2\pi}\int_{S^1} d\sigma_\mu \varepsilon_{\mu\nu}\partial_\nu \varphi(x) \\ &= \frac{1}{2\pi}(\varphi(2\pi) - \varphi(0)). \end{aligned} \tag{6.3.18}$$

If $U(x)$ 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 U 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 $\tilde{U}$ as

$$\begin{aligned} &\tilde{U}(x) = \exp(i\vec{\alpha}(x)\cdot\vec{\sigma}) = \cos\alpha(x) + i\sin\alpha(x)\vec{e}_\alpha(x)\cdot\vec{\sigma}, \\ &\vec{e}_\alpha(x) = (\sin\theta(x)\sin\varphi(x), \sin\theta(x)\cos\varphi(x), \cos\theta(x)). \end{aligned} \tag{6.3.19}$$

It is a good exercise to convince oneself that

$$\begin{aligned} &\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(\tilde{U}(x)\partial_\nu\tilde{U}(x)^+)(\tilde{U}(x)\partial_\rho\tilde{U}(x)^+)(\tilde{U}(x)\partial_\sigma\tilde{U}(x)^+)] \\ &= 12\sin^2\alpha(x)\sin\theta(x)\varepsilon_{\mu\nu\rho\sigma}\partial_\nu\alpha(x)\partial_\rho\theta(x)\partial_\sigma\varphi(x). \end{aligned} \tag{6.3.20}$$

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

$$Q = \frac{1}{2\pi^2}\int_{S^3} d^3\sigma_\mu \sin^2\alpha(x)\sin\theta(x)\varepsilon_{\mu\nu\rho\sigma}\partial_\nu\alpha(x)\partial_\rho\theta(x)\partial_\sigma\varphi(x) = \frac{1}{2\pi^2}\int_{S^3} d\tilde{U}. \tag{6.3.21}$$

The volume of the 3-sphere is given by $2\pi^2$. When the mapping $\tilde{U}$ 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.

6.4 Topology of the Gluon Field on a Compact Manifold

Imagine our universe were 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, the η'-meson mass is non-zero because the topological charge does not vanish. To resolve this apparent contradiction we will now discuss the topology of the gluon 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

$$Q = \int_M d^4x \; P(x). \tag{6.4.1}$$

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

$$P(x) = \partial_\mu \Omega^{(0)}_\mu(x), \tag{6.4.2}$$

and using Gauss' law we obtain

$$Q = \int_M d^4x \; \partial_\mu \Omega^{(0)}_\mu(x) = \int_{\partial M} d^3\sigma_\mu \Omega^{(0)}_\mu(x) = 0. \tag{6.4.3}$$

Here we have used that M has no boundary, i.e. ∂M is an empty set. A gluon 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.(6.4.2) 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 gluon field configuration 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 $c_i \cap c_j = \partial c_i \cap \partial c_j$. This geometry is illustrated in fig.6.4. The next step is to remove

Figure 6.4: *The manifold M with singularities x_i is covered by closed sets c_i.*

the gauge singularities x_i by performing gauge transformations g_i in each local patch

$$G^i_\mu(x) = g_i(x)(G_\mu(x) + \partial_\mu)g_i^+(x). \tag{6.4.4}$$

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

$$\begin{aligned} Q &= \sum_i \int_{c_i} d^4x \; P(x) = \sum_i \int_{\partial c_i} d^3\sigma_\mu \; \Omega^{(0)}_\mu(i) \\ &= \frac{1}{2}\sum_{ij} \int_{c_i \cap c_j} d^3\sigma_\mu \; [\Omega^{(0)}_\mu(i) - \Omega^{(0)}_\mu(j)]. \end{aligned} \tag{6.4.5}$$

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}

$$G^i_\mu(x) = v_{ij}(x)(G^j_\mu(x) + \partial_\mu)v_{ij}(x)^+, \tag{6.4.6}$$

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

$$v_{ij}(x) = g_i(x)g_j(x)^+. \tag{6.4.7}$$

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

$$v_{ik}(x) = v_{ij}(x)v_{jk}(x). \tag{6.4.8}$$

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

$$\Delta\Omega^{(0)}_\mu(i,j) = \Omega^{(0)}_\mu(i) - \Omega^{(0)}_\mu(j). \tag{6.4.9}$$

It is straight forward to show that

$$\begin{aligned}\Delta\Omega^{(0)}_\mu(i,j) = &-\frac{1}{24\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[v_{ij}(x)\partial_\nu v_{ij}(x)^+ v_{ij}(x)\partial_\rho v_{ij}(x)^+ v_{ij}(x)\partial_\sigma v_{ij}(x)^+] \\ &-\frac{1}{8\pi^2}\varepsilon_{\mu\nu\rho\sigma}\partial_\nu\mathrm{Tr}[\partial_\rho v_{ij}(x)^+ v_{ij}(x)G^i_\sigma(x)].\end{aligned} \tag{6.4.10}$$

The above equation for the topological charge then takes the form

$$\begin{aligned}Q &= -\frac{1}{48\pi^2}\sum_{ij}\int_{c_i\cap c_j} d^3\sigma_\mu\ \varepsilon_{\mu\nu\rho\sigma} \\ &\times\ \mathrm{Tr}[v_{ij}(x)\partial_\nu v_{ij}(x)^+ v_{ij}(x)\partial_\rho v_{ij}(x)^+ v_{ij}(x)\partial_\sigma v_{ij}(x)^+] \\ &-\ \frac{1}{16\pi^2}\sum_{ij}\int_{\partial(c_i\cap c_j)} d^2\sigma_{\mu\nu}\ \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[\partial_\rho v_{ij}(x)^+ v_{ij}(x)G^i_\sigma(x)].\end{aligned} \tag{6.4.11}$$

Using the cocycle condition this can be rewritten as

$$\begin{aligned}Q &= -\frac{1}{48\pi^2}\sum_{ij}\int_{c_i\cap c_j} d^3\sigma_\mu\ \varepsilon_{\mu\nu\rho\sigma} \\ &\times\ \mathrm{Tr}[v_{ij}(x)\partial_\nu v_{ij}(x)^+ v_{ij}(x)\partial_\rho v_{ij}(x)^+ v_{ij}(x)\partial_\sigma v_{ij}(x)^+] \\ &-\frac{1}{48\pi^2}\sum_{ijk}\int_{c_i\cap c_j\cap c_k} d^2\sigma_{\mu\nu}\ \varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[v_{ij}(x)\partial_\rho v_{ij}(x)^+ v_{jk}(x)\partial_\rho v_{jk}(x)^+].\end{aligned} \tag{6.4.12}$$

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 (6.3.8) together with eq.(6.4.7) one can show that the topological charge is expressed in terms of the section in the following way

$$\begin{aligned} Q &= \sum_i Q_i \\ &= \frac{1}{24\pi^2} \sum_i \int_{\partial c_i} d^3\sigma_\mu \varepsilon_{\mu\nu\rho\sigma} \mathrm{Tr}[g_i(x)\partial_\nu g_i(x)^+ g_i(x)\partial_\rho g_i(x)^+ g_i(x)\partial_\sigma g_i(x)^+]. \end{aligned} \tag{6.4.13}$$

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

$$g_i : \partial c_i \to SU(3), \tag{6.4.14}$$

and hence

$$Q_i \in \Pi_3[SU(3)] = \mathbf{Z}. \tag{6.4.15}$$

The topological charge Q is a sum of local winding numbers $Q_i \in \mathbf{Z}$, 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

$$G_\mu(x)' = U(x)(G_\mu(x) + \partial_\mu)U(x)^+. \tag{6.4.16}$$

Deriving the gauge transformation properties of the section and using formula (6.3.8) this is again straight forward.

6.5 Cochain Reduction in $SU(2)$

To complete the mathematical investigation of the topological charge let us finally investigate its expression eq.(6.4.12) 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

$$Q = \sum_i \int_{\partial c_i} d^3\sigma_\mu \Omega^{(0)}_\mu(i) = \frac{1}{2!}\sum_{ij}\int_{c_i\cap c_j} d^3\sigma_\mu[\Omega^{(0)}_\mu(i) - \Omega^{(0)}_\mu(j)]. \qquad (6.5.1)$$

The coboundary operator

$$\Delta\Omega^{(0)}_\mu(i,j) = \Omega^{(0)}_\mu(i) - \Omega^{(0)}_\mu(j) = \partial_\nu \Omega^{(1)}_{\mu\nu}(i,j) \qquad (6.5.2)$$

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

$$\begin{aligned}\Omega^{(1)}_{\mu\nu}(i,j) &= -\frac{1}{8\pi^2}(\alpha - \sin\alpha\cos\alpha)\varepsilon_{\mu\nu\rho\sigma}\vec{e}_\alpha\cdot(\partial_\rho\vec{e}_\alpha\times\partial_\sigma\vec{e}_\alpha)\\ &- \frac{1}{8\pi^2}\varepsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[\partial_\rho v_{ij}v^+_{ij}G^i_\sigma]. \qquad (6.5.3)\end{aligned}$$

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

$$v_{ij} = \exp(i\vec{\alpha}\cdot\vec{\sigma}) = \cos\alpha + i\sin\alpha\vec{e}_\alpha\cdot\vec{\sigma},\ \alpha\in[0,\pi]. \qquad (6.5.4)$$

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

$$\Omega^{(1)}_{\mu\nu}(i,j) = -\frac{1}{8\pi}\varepsilon_{\mu\nu\rho\sigma}\vec{e}_\alpha\cdot(\partial_\rho\vec{e}_\alpha\times\partial_\sigma\vec{e}_\alpha). \qquad (6.5.5)$$

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

$$Q = Q^{(1)} + Q^{(1)}_\Sigma, \qquad (6.5.6)$$

where

$$Q^{(1)} = \frac{1}{2!}\sum_{ij}\sum_{x\in c_i\cap c_j}\frac{1}{8\pi}\int_{S^2_\varepsilon(x)} d^2\sigma_{\mu\nu}\ \varepsilon_{\mu\nu\rho\sigma}\vec{e}_\alpha\cdot(\partial_\rho\vec{e}_\alpha\times\partial_\sigma\vec{e}_\alpha). \qquad (6.5.7)$$

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

$$\frac{1}{8\pi}\int_{S^2_\varepsilon(x)} d^2\sigma_{\mu\nu}\ \varepsilon_{\mu\nu\rho\sigma}\vec{e}_\alpha\cdot(\partial_\rho\vec{e}_\alpha\times\partial_\sigma\vec{e}_\alpha) = n^{(1)}(x;i,j), \qquad (6.5.8)$$

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

$$Q^{(1)} = \sum_{\Lambda^{(1)}} n^{(1)}(x;i,j) \tag{6.5.9}$$

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

$$\Lambda^{(1)} = \cup_{ij}\{x \in c_i \cap c_j\}. \tag{6.5.10}$$

The remaining contribution to the topological charge is given by

$$\begin{aligned} Q^{(1)}_\Sigma &= \frac{1}{2!}\sum_{ij}\int_{\partial(c_i\cap c_j)} d^2\sigma_{\mu\nu}\ \Omega^{(1)}_{\mu\nu}(i,j) \\ &= \frac{1}{3!}\sum_{ijk}\int_{c_i\cap c_j\cap c_k} d^2\sigma_{\mu\nu}\ [\Omega^{(1)}_{\mu\nu}(i,j) - \Omega^{(1)}_{\mu\nu}(i,k) + \Omega^{(1)}_{\mu\nu}(j,k)]. \end{aligned} \tag{6.5.11}$$

We identify the coboundary operator

$$\Delta\Omega^{(1)}_{\mu\nu}(i,j,k) = \Omega^{(1)}_{\mu\nu}(i,j) - \Omega^{(1)}_{\mu\nu}(i,k) + \Omega^{(1)}_{\mu\nu}(j,k) = \partial_\rho\Omega^{(2)}_{\mu\nu\rho}(i,j,k), \tag{6.5.12}$$

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

$$\begin{aligned} &\Omega^{(2)}_{\mu\nu\rho}(i,j,k) = \\ &-\frac{1}{8\pi^2}\varepsilon_{\mu\nu\rho\sigma}(1 + 2\cos\alpha\cos\beta\cos\gamma - \cos^2\alpha - \cos^2\beta - \cos^2\gamma)^{-1} \\ &\times\{(\vec{\alpha}+\vec{\beta}-\vec{\gamma})\cdot\sin\alpha\vec{e}_\alpha[\partial_\sigma(\sin\beta\vec{e}_\beta)\cdot\sin\gamma\vec{e}_\gamma - \sin\beta\vec{e}_\beta\cdot\partial_\sigma(\sin\gamma\vec{e}_\gamma)] \\ &+(\vec{\alpha}+\vec{\beta}-\vec{\gamma})\cdot\sin\beta\vec{e}_\beta[\partial_\sigma(\sin\gamma\vec{e}_\gamma)\cdot\sin\alpha\vec{e}_\alpha - \sin\gamma\vec{e}_\gamma\cdot\partial_\sigma(\sin\alpha\vec{e}_\alpha)] \\ &+(\vec{\alpha}+\vec{\beta}-\vec{\gamma})\cdot\sin\gamma\vec{e}_\gamma[\partial_\sigma(\sin\alpha\vec{e}_\alpha)\cdot\sin\beta\vec{e}_\beta - \sin\alpha\vec{e}_\alpha\cdot\partial_\sigma(\sin\beta\vec{e}_\beta)]\}. \end{aligned} \tag{6.5.13}$$

Here we have parametrized the transition functions as

$$v_{ij} = \exp(i\vec{\alpha}\cdot\sigma),\ v_{jk} = \exp(i\vec{\beta}\cdot\sigma),\ v_{ik} = \exp(i\vec{\gamma}\cdot\sigma). \tag{6.5.14}$$

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

$$\begin{aligned} \cos\gamma &= \cos\alpha\cos\beta - \sin\alpha\sin\beta\vec{e}_\alpha\cdot\vec{e}_\beta, \\ \sin\gamma\vec{e}_\gamma &= \sin\alpha\cos\beta\vec{e}_\alpha + \cos\alpha\sin\beta\vec{e}_\beta - \sin\alpha\sin\beta\vec{e}_\alpha\times\vec{e}_\beta. \end{aligned} \tag{6.5.15}$$

Figure 6.5: *Three transition functions define a spherical triangle on S^3.*

These equations define a spherical triangle on the sphere S^3 with side lengths α, β and γ. This is illustrated in fig.6.5. 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

$$(\vec{\alpha}+\vec{\beta}-\vec{\gamma})\cdot\vec{e}_\alpha = (\vec{\alpha}+\vec{\beta}-\vec{\gamma})\cdot\vec{e}_\beta = -(\vec{\alpha}+\vec{\beta}-\vec{\gamma})\cdot\vec{e}_\gamma = 2\pi. \quad (6.5.16)$$

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

$$\Omega^{(2)}_{\mu\nu\rho}(i,j,k) = -\frac{1}{4\pi^2}\varepsilon_{\mu\nu\rho\sigma}(\vec{\alpha}+\vec{\beta}-\vec{\gamma})\cdot(\vec{e}_s\times\partial_\sigma\vec{e}_s). \quad (6.5.17)$$

Here we have defined another unit vector

$$\vec{e}_s = \cos\alpha\sin\beta\sin\gamma\vec{e}_\beta\times\vec{e}_\gamma + \cos\beta\sin\gamma\sin\alpha\vec{e}_\gamma\times\vec{e}_\alpha + \cos\gamma\sin\alpha\sin\beta\vec{e}_\alpha\times\vec{e}_\beta. \quad (6.5.18)$$

Using Gauss' law we now obtain

$$Q^{(1)}_\Sigma = Q^{(2)} + Q^{(2)}_\Sigma, \quad (6.5.19)$$

with

$$Q^{(2)} = \frac{1}{3!}\sum_{ijk}\sum_{x\in c_i\cap c_j\cap c_k}\frac{1}{4\pi^2}\int_{S^1_\varepsilon(x)} d\sigma_{\mu\nu\rho}\varepsilon_{\mu\nu\rho\sigma}(\vec{\alpha}+\vec{\beta}-\vec{\gamma})\cdot(\vec{e}_s\times\partial_\sigma\vec{e}_s). \quad (6.5.20)$$

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

$$\frac{1}{4\pi^2}\int_{S^1_\varepsilon(x)} d\sigma_{\mu\nu\rho}\varepsilon_{\mu\nu\rho\sigma}(\vec{\alpha}+\vec{\beta}-\vec{\gamma})\cdot(\vec{e}_s\times\partial_\sigma\vec{e}_s) = n^{(2)}(x;i,j,k), \quad (6.5.21)$$

where $n^{(2)}(x;i,j,k)$ is a winding number from $\Pi_1[S^1]=\mathbf{Z}$, 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

$$Q^{(2)} = \sum_{\Lambda^{(2)}} n^{(2)}(x;i,j,k), \quad (6.5.22)$$

where the set of singular points is now denoted by

$$\Lambda^{(2)} = \cup_{ijk}\{x\in c_i\cap c_j\cap c_k\}. \quad (6.5.23)$$

The remaining contributions to the topological charge take the form

$$\begin{aligned} Q^{(2)}_\Sigma &= \frac{1}{3!}\sum_{ijk}\int_{\partial(c_i\cap c_j\cap c_k)} d\sigma_{\mu\nu\rho}\Omega^{(2)}_{\mu\nu\rho}(i,j,k) \\ &= \frac{1}{4!}\sum_{ijkl}\int_{c_i\cap c_j\cap c_k\cap c_l} d\sigma_{\mu\nu\rho}[\Omega^{(2)}_{\mu\nu\rho}(i,j,k)-\Omega^{(2)}_{\mu\nu\rho}(i,j,l) \\ &+ \Omega^{(2)}_{\mu\nu\rho}(i,k,l)-\Omega^{(2)}_{\mu\nu\rho}(j,k,l)]. \end{aligned} \quad (6.5.24)$$

It comes as no surprise that the coboundary operator

$$\begin{aligned} \Delta\Omega^{(2)}_{\mu\nu\rho}(i,j,k,l) &= \Omega^{(2)}_{\mu\nu\rho}(i,j,k)-\Omega^{(2)}_{\mu\nu\rho}(i,j,l)+\Omega^{(2)}_{\mu\nu\rho}(i,k,l)-\Omega^{(2)}_{\mu\nu\rho}(j,k,l) \\ &= \partial_\sigma\Omega^{(3)}_{\mu\nu\rho\sigma}(i,j,k,l) \end{aligned} \quad (6.5.25)$$

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. The geometry is illustrated in fig.6.6. It is quite tedious to show that

$$\Omega^{(3)}_{\mu\nu\rho\sigma}(i,j,k,l) = \frac{1}{2\pi^2}\varepsilon_{\mu\nu\rho\sigma}V(i,j,k,l). \quad (6.5.26)$$

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

$$\partial_\sigma V(i,j,k,l) = \frac{1}{2}(\alpha\partial_\sigma A+\beta\partial_\sigma B+\gamma\partial_\sigma\Gamma+\delta\partial_\sigma\Delta+\varepsilon\partial_\sigma E+\zeta\partial_\sigma Z), \quad (6.5.27)$$

Figure 6.6: *Four cocycle conditions define a spherical tetrahedron on the sphere S^3.*

where the angles $A, B, ..., Z$ are defined in fig.6.6. 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

$$Q_\Sigma^{(2)} = Q^{(3)} + Q_\Sigma^{(3)}. \tag{6.5.28}$$

The first contribution has the form

$$Q^{(3)} = \sum_{\Lambda^{(3)}} n^{(3)}(x; i, j, k, l), \tag{6.5.29}$$

where

$$\Lambda^{(3)} = \cup_{ijkl}\{x \in c_i \cap c_j \cap c_k \cap c_l\} \tag{6.5.30}$$

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

$$n^{(3)}(x; i, j, k, l) = \frac{1}{2\pi^2}\Delta V(i, j, k, l) \tag{6.5.31}$$

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

$$Q_\Sigma^{(3)} = \frac{1}{4!} \sum_{\partial(c_i \cap c_j \cap c_k \cap c_l)} \sigma_{\mu\nu\rho\sigma} \Omega^{(3)}_{\mu\nu\rho\sigma}(i, j, k, l)$$

$$
\begin{aligned}
&= \frac{1}{5!}\sum_{c_i\cap c_j\cap c_k\cap c_l\cap c_m}\sigma_{\mu\nu\rho\sigma}[\Omega^{(3)}_{\mu\nu\rho\sigma}(i,j,k,l)-\Omega^{(3)}_{\mu\nu\rho\sigma}(i,j,k,m)\\
&+ \Omega^{(3)}_{\mu\nu\rho\sigma}(i,j,l,m)-\Omega^{(3)}_{\mu\nu\rho\sigma}(i,k,l,m)+\Omega^{(3)}_{\mu\nu\rho\sigma}(j,k,l,m)]\\
&= \frac{1}{5!}\sum_{c_i\cap c_j\cap c_k\cap c_l\cap c_m}\sigma\frac{1}{2\pi^2}[V(i,j,k,l)-V(i,j,k,m)\\
&+ V(i,j,l,m)-V(i,k,l,m)+V(j,k,l,m)]. \qquad (6.5.32)
\end{aligned}
$$

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

$$\Lambda^{(4)}=\{x\in\cup_{ijklm}c_i\cap c_j\cap c_k\cap c_l\cap c_m\}, \qquad (6.5.33)$$

such that

$$Q^{(3)}_\Sigma=Q^{(4)}=\sum_{\Lambda^{(4)}}n^{(4)}(x;i,j,k,l,m) \qquad (6.5.34)$$

with

$$
\begin{aligned}
n^{(4)}(x;i,j,k,l,m) &= \sigma\frac{1}{2\pi^2}[V(i,j,k,l)-V(i,j,k,m)+V(i,j,l,m)\\
&- V(i,k,l,m)+V(j,k,l,m)]. \qquad (6.5.35)
\end{aligned}
$$

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

$$
\begin{aligned}
Q &= \sum_{\Lambda^{(1)}}n^{(1)}(x;i,j)+\sum_{\Lambda^{(2)}}n^{(2)}(x;i,j,k)\\
&+ \sum_{\Lambda^{(3)}}n^{(3)}(x;i,j,k,l)+\sum_{\Lambda^{(4)}}n^{(4)}(x;i,j,k,l,m). \qquad (6.5.36)
\end{aligned}
$$

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.

6.6 The Instanton in $SU(2)$

We have argued mathematically that gluon 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 QCD.

Here we concentrate on $SU(2)$. This is sufficient, because we have seen that the $SU(3)$ 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

$$G_\mu(x) = U(x)\partial_\mu U(x)^+. \tag{6.6.1}$$

Provided the gauge field is otherwise smooth, the topology resides entirely in the mapping U. We want to construct a field configuration with topological charge $Q = 1$, i.e. one in which the mapping U 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

$$U(x) = \frac{x_0 + i\vec{x}\cdot\vec{\sigma}}{|x|},\ |x| = \sqrt{x_0^2 + |\vec{x}|^2}. \tag{6.6.2}$$

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.(6.6.1) because U is singular at $x = 0$. To avoid this singularity we make the ansatz

$$G_\mu(x) = f(|x|)U(x)\partial_\mu U(x)^+, \tag{6.6.3}$$

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

$$S[G_\mu] = \int d^4x\ \frac{1}{2g_s^2}\mathrm{Tr}[G_{\mu\nu}(x)G_{\mu\nu}(x)]. \tag{6.6.4}$$

Let us consider the following integral

$$\int d^4x \ \mathrm{Tr}[(G_{\mu\nu}(x) \pm \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}G_{\rho\sigma}(x))(G_{\mu\nu}(x) \pm \frac{1}{2}\varepsilon_{\mu\nu\kappa\lambda}G_{\kappa\lambda}(x)) =$$
$$\int d^4x \ \mathrm{Tr}[G_{\mu\nu}(x)G_{\mu\nu}(x) \pm \varepsilon_{\mu\nu\rho\sigma}G_{\mu\nu}(x)G_{\rho\sigma}(x) + G_{\mu\nu}(x)G_{\mu\nu}(x)] =$$
$$4g_s^2 S[G_\mu] \pm 32\pi^2 Q[G_\mu]. \tag{6.6.5}$$

We have integrated a square. Hence it is obvious that

$$S[G_\mu] \pm \frac{8\pi^2}{g_s^2}Q[G_\mu] \geq 0 \ \Rightarrow \ S[G_\mu] \geq \frac{8\pi^2}{g_s^2}|Q[G_\mu]|, \tag{6.6.6}$$

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

$$S[G_\mu] = \frac{8\pi^2}{g_s^2}|Q[G_\mu]|. \tag{6.6.7}$$

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

$$G_{\mu\nu}(x) = \pm\frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}G_{\rho\sigma}(x). \tag{6.6.8}$$

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

$$f(|x|) = \frac{|x|^2}{|x|^2 + \rho^2} \tag{6.6.9}$$

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

$$G_\mu(x) = \frac{|x|^2}{|x|^2 + \rho^2}U(x)\partial_\mu U(x)^+. \tag{6.6.10}$$

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 $S[G_\mu] = 8\pi^2/g_s^2$.

6.7 θ-Vacua

The existence of topologically nontrivial gauge transformations has drastic consequences for QCD. 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 R^3 to S^3. This is just a technical trick which makes live 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 R^3 itself. The classical vacuum solutions of such a theory are the pure gauge fields

$$G_i(x) = U(x)\partial_i U(x)^+. \tag{6.7.1}$$

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

$$n \in \Pi_3[SU(3)] = \mathbf{Z}, \tag{6.7.2}$$

which is given by

$$n = \frac{1}{24\pi^2}\int_{S^3} d^3x\ \varepsilon_{ijk}\mathrm{Tr}[U(x)\partial_i U(x)^+ U(x)\partial_j U(x)^+ U(x)\partial_k U(x)^+]. \tag{6.7.3}$$

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

$$Q = n - m. \tag{6.7.4}$$

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

$$\langle n|U(\infty, -\infty)|m\rangle = \int \mathcal{D}G_\mu^{(n-m)} \exp(-S[G_\mu]). \tag{6.7.5}$$

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.

$$T\Psi[G_i] = \Psi[g(G_i + \partial_i)g^+]. \tag{6.7.6}$$

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

$$T|n\rangle = |n + 1\rangle, \tag{6.7.7}$$

i.e. T acts as a ladder operator. Since the operator T implements a special gauge transformation, it commutes with the QCD Hamiltonian, just because QCD is gauge invariant. This means that the QCD Hamiltonian and T can be diagonalized simultaneously, and each QCD 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 a QCD eigenstate — for example the vacuum — can be written as $|\theta\rangle$ with

$$T|\theta\rangle = \exp(i\theta)|\theta\rangle. \tag{6.7.8}$$

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

$$|\theta\rangle = \sum_n c_n|n\rangle. \tag{6.7.9}$$

Using

$$\begin{aligned} T|\theta\rangle &= \sum_n c_n T|n\rangle = \sum_n c_n |n+1\rangle \\ &= \sum_n c_{n-1}|n\rangle = \exp(i\theta)\sum_n c_n|n\rangle, \end{aligned} \tag{6.7.10}$$

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

$$|\theta\rangle = \sum_n \exp(-in\theta)|n\rangle. \tag{6.7.11}$$

The true vacuum of QCD 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 QCD T induces discrete translations between classical vacua, with analogous mathematical consequences.

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

$$\begin{aligned} &\langle\theta|U(\infty,-\infty)|\theta'\rangle = \sum_{m,n}\exp(in\theta)\exp(-im\theta')\langle n|U(\infty,-\infty)|m\rangle \\ &= \sum_{n,Q=n-m}\exp(in\theta - i(n-Q)\theta')\int \mathcal{D}G_\mu^{(Q)}\exp(-S[G_\mu]) \\ &= \delta(\theta-\theta')\sum_Q\int\mathcal{D}G_\mu^{(Q)}\exp(-S[G_\mu])\exp(i\theta Q[G_\mu]) \\ &= \int \mathcal{D}G_\mu \exp(-S_\theta[G_\mu]). \end{aligned} \tag{6.7.12}$$

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

$$S_\theta[G_\mu] = S[G_\mu] - i\theta Q[G_\mu]. \tag{6.7.13}$$

Finally, let us consider the theory with at least one massless quark. 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

$$\gamma_\mu(G_\mu(x)+\partial_\mu)\Psi(x) = 0. \tag{6.7.14}$$

These eigenvectors have a definite handedness, i.e.

$$\frac{1}{2}(1 \pm \gamma_5)\Psi(x) = \Psi(x), \tag{6.7.15}$$

because

$$\gamma_5\gamma_\mu(G_\mu(x) + \partial_\mu)\Psi(x) = -\gamma_\mu(G_\mu(x) + \partial_\mu)\gamma_5\Psi(x) = 0. \tag{6.7.16}$$

The Atiyah-Singer index theorem states that

$$Q = n_L - n_R, \tag{6.7.17}$$

where n_L and n_R are the numbers of left and right handed zero modes. Hence, a topologically nontrivial gluon field configuration necessarily has at least one zero mode. This zero mode of the Dirac operator eliminates topologically nontrivial field configurations from QCD with a massless quark, 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. My personal opinion is that the solution of the strong CP problem is beyond QCD, and probably beyond the Standard model, and is maybe related to some cosmological effect.

6.8 A Quantitative Solution of the $U(1)$ Problem

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

$$\partial_\mu A_\mu(x) = 2N_f P(x), \tag{6.8.1}$$

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 so solve it analytically. Still, one can classify the subset of Feynman diagrams that contributes 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. This is illustrated in fig.6.7. In this way any large N_c QCD diagram

Figure 6.7: *In the large N_c limit the gluon propagator can be replaced by a quark-antiquark color flow. The gluon self energy diagram is also shown.*

can be represented as a quark diagram. This is illustrated for the gluon self energy diagram also in fig.6.7. The internal loop 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

$$g^2 = g_s^2 N_c, \tag{6.8.2}$$

and we perform the large N_c limit such that g_s goes to zero but g remains finite. Let us now consider the 2-loop diagram of fig.6.8. 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

Figure 6.8: *Planar 2-loop and 4-loop gluon diagrams in the large N_c limit.*

to g^4 and hence it is finite. Let us also consider the 4-loop diagram in fig.6.8. 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 the non-planar diagram of fig.6.9. The

Figure 6.9: *A non-planar 4-loop diagram and a quark diagram that vanish in the large N_c limit.*

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 from fig.6.9. 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 better.

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 antiquark, 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

$$S[G_\mu] \geq \frac{8\pi^2}{g_s^2}|Q[G_\mu]| = \frac{8\pi^2 N_c}{g^2}|Q[G_\mu]|, \tag{6.8.3}$$

which is valid for all $SU(N_c)$. In the large N_c limit the action of an instanton diverges, and topologically nontrivial 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 we should be able to derive a mass formula for the η'-meson just as we did 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 a 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

$$\chi_t = \int d^4x \ {}_{pg}\langle 0|P(0)P(x)|0\rangle_{pg} \frac{Q^2}{V} \tag{6.8.4}$$

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)

$$\int d^4x \ \langle 0|P(0)P(x)|0\rangle = 0, \tag{6.8.5}$$

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

$$\chi_t - \sum_m \frac{\langle 0|P|m\rangle\langle m|P|0\rangle}{M_m^2} = 0, \tag{6.8.6}$$

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

$$\chi_t = \frac{|\langle 0|P|\eta'\rangle|^2}{M_{\eta'}^2}. \tag{6.8.7}$$

Using the anomaly equation one obtains

$$\langle 0|P|\eta'\rangle = \frac{1}{2N_f}\langle 0|\partial_\mu A_\mu|\eta'\rangle = \frac{1}{\sqrt{2N_f}} M_{\eta'}^2 f_{\eta'}. \tag{6.8.8}$$

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

$$\chi_t = \frac{f_\pi^2 M_{\eta'}^2}{2N_f}. \tag{6.8.9}$$

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

$$\chi_t = \frac{1}{6} f_\pi^2 (M_{\eta'}^2 + M_\eta^2 - 2M_K^2) = (0.180\text{GeV})^4. \tag{6.8.10}$$

Lattice calculations are at least roughly consistent with this value, which supports this solution of the $U(1)$-problem.

Chapter 7

Lattice Field Theory

7.1 The Quantum Mechanical Path Integral in Real Time

The quantization of field theories is most conveniently performed using the path integral approach. Here we introduce the path integral in quantum 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

$$i\hbar\partial_t|\Psi(t)\rangle = H|\Psi(t)\rangle. \tag{7.1.1}$$

For a time-independent Hamilton operator the time evolution operator is given by

$$U(t',t) = \exp(-\frac{i}{\hbar}H(t'-t)), \tag{7.1.2}$$

such that

$$|\Psi(t')\rangle = U(t',t)|\Psi(t)\rangle. \tag{7.1.3}$$

Let us consider the transition amplitude $\langle x'|U(t',t)|x\rangle$ for a nonrelativistic particle starting at the time t at the position x for arriving at time t' at the final position x'. Using

$$\langle x|\Psi(t)\rangle = \Psi(x,t) \tag{7.1.4}$$

we obtain

$$\Psi(x',t') = \int dx \langle x'|U(t',t)|x\rangle \Psi(x,t), \tag{7.1.5}$$

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

$$\begin{aligned} \langle x|U(t',t)|x\rangle &= \langle x| \exp(-\frac{i}{\hbar}H(t'-t))|x\rangle \\ &= \sum_n |\langle x|n\rangle|^2 \exp(-\frac{i}{\hbar}E_n(t'-t)), \end{aligned} \tag{7.1.6}$$

i.e. 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

$$\begin{aligned} \langle x'|U(t',t)|x\rangle &= \langle x'| \exp(-\frac{i}{\hbar}H(t'-t_1+t_1-t))|x\rangle \\ &= \int dx_1 \langle x'| \exp(-\frac{i}{\hbar}H(t'-t_1))|x_1\rangle \\ &\times \langle x_1| \exp(-\frac{i}{\hbar}H(t_1-t))|x\rangle \\ &= \int dx_1 \langle x'|U(t',t_1)|x_1\rangle \langle x_1|U(t_1,t)|x\rangle. \end{aligned} \tag{7.1.7}$$

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

$$t'-t = N\varepsilon. \tag{7.1.8}$$

Inserting a complete set of position eigenstates at the intermediate times $t_i, i \in \{1,2,...,N-1\}$ we obtain

$$\begin{aligned} \langle x'|U(t',t)|x\rangle &= \int dx_1 \int dx_2 ... \int dx_{N-1} \langle x'|U(t',t_{N-1})|x_{N-1}\rangle ... \\ &\times \langle x_2|U(t_2,t_1)|x_1\rangle \langle x_1|U(t_1,t)|x\rangle. \end{aligned} \tag{7.1.9}$$

In the next step we concentrate on one of the factors and we consider a single nonrelativistic particle moving in an external potential $V(x)$ such

that

$$H = \frac{p^2}{2m} + V(x). \tag{7.1.10}$$

Using the Baker-Campbell-Haussdorff formula and neglecting terms of order ε^2 we find

$$\begin{aligned}
\langle x_{i+1}|U(t_{i+1},t_i)|x_i\rangle &= \langle x_{i+1}|\exp(-\frac{i\varepsilon p^2}{2m\hbar})\exp(-\frac{i\varepsilon}{\hbar}V(x))|x_i\rangle \\
&= \frac{1}{2\pi}\int dp\langle x_{i+1}|\exp(-\frac{i\varepsilon p^2}{2m\hbar})|p\rangle\langle p|\exp(-\frac{i\varepsilon}{\hbar}V(x))|x_i\rangle \\
&= \frac{1}{2\pi}\int dp\exp(-\frac{i\varepsilon p^2}{2m\hbar})\exp(-\frac{i}{\hbar}p(x_{i+1}-x_i)) \\
&\times \exp(-\frac{i\varepsilon}{\hbar}V(x)).
\end{aligned} \tag{7.1.11}$$

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

$$\langle x_{i+1}|U(t_{i+1},t_i)|x_i\rangle = \sqrt{\frac{m}{2\pi i\hbar\varepsilon}}\exp(\frac{i}{\hbar}\varepsilon[\frac{m}{2}(\frac{x_{i+1}-x_i}{\varepsilon})^2 - V(x_i)]). \tag{7.1.12}$$

Inserting this back into the expression for the propagator we obtain

$$\langle x'|U(t',t)|x\rangle = \int \mathcal{D}x\exp(\frac{i}{\hbar}S[x]). \tag{7.1.13}$$

The classical action has been identified in the time continuum limit as

$$\begin{aligned}
S[x] &= \int dt[\frac{m}{2}(\frac{dx}{dt})^2 - V(x)] \\
&= \lim_{\varepsilon\to 0}\sum_i \varepsilon[\frac{m}{2}(\frac{x_{i+1}-x_i}{\varepsilon})^2 - V(x_i)].
\end{aligned} \tag{7.1.14}$$

The integration measure is defined as

$$\int \mathcal{D}x = \lim_{\varepsilon\to 0}\sqrt{\frac{m}{2\pi i\hbar\varepsilon}}^{N-1}\int dx_1\int dx_2 ...\int dx_{N-1}. \tag{7.1.15}$$

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 $\exp(\frac{i}{\hbar}S[x])$ given by the classical 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. It is instructive to compute the propagator for a free particle by performing the path integral explicitly, and to compare it with the result obtained directly from the Schrödinger equation.

7.2 The Quantum Mechanical 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

$$Z = \mathrm{Tr}\exp(-\beta H), \tag{7.2.1}$$

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

$$\beta = \frac{i}{\hbar}(t' - t). \tag{7.2.2}$$

In this sense the system at finite temperature corresponds to one propagating in purely imaginary (Euclidean) time. By dividing the Euclidean time interval into N time steps, i.e. by writing $\beta = Na/\hbar$, and by again inserting complete sets of position eigenstates we now arrive at the Euclidean time path integral

$$Z = \int \mathcal{D}x \exp(-\frac{1}{\hbar}S_E[x]). \tag{7.2.3}$$

The action now takes the Euclidean form

$$\begin{aligned} S_E[x] &= \int dt[\frac{m}{2}(\frac{dx}{dt})^2 + V(x)] \\ &= \lim_{a\to 0}\sum_i a[\frac{m}{2}(\frac{x_{i+1} - x_i}{a})^2 + V(x_i)], \end{aligned} \tag{7.2.4}$$

which is positive definite if $V(x)$ is bounded from below. In contrast to the real time case the measure now involves N integrations

$$\int \mathcal{D}x = \lim_{a\to 0} \sqrt{\frac{m}{2\pi i\hbar a}}^N \int dx_1 \int dx_2 ... \int dx_N. \tag{7.2.5}$$

The extra integration over $x_N = x'$ is due to the trace in eq.(7.2.1). 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

$$\langle \mathcal{O}(x)\rangle = \frac{1}{Z}\mathrm{Tr}[\mathcal{O}(x)\exp(-\beta H)] = \frac{1}{Z}\int \mathcal{D}x\mathcal{O}(x(0))\exp(-\frac{1}{\hbar}S_E[x]). \tag{7.2.6}$$

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 $\beta \to \infty$ 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

$$\langle 0|\mathcal{O}(x)|0\rangle = \lim_{\beta\to\infty} \frac{1}{Z}\int \mathcal{D}x\mathcal{O}(x(0))\exp(-\frac{1}{\hbar}S_E[x]). \tag{7.2.7}$$

It is also interesting to consider 2-point functions of operators at different moments in Euclidean time

$$\begin{aligned}\langle \mathcal{O}(x(0))\mathcal{O}(x(t))\rangle &= \frac{1}{Z}\mathrm{Tr}[\mathcal{O}(x)\exp(-Ht)\mathcal{O}(x)\exp(Ht)\exp(-\beta H)] \\ &= \frac{1}{Z}\int \mathcal{D}x\mathcal{O}(x(0))\mathcal{O}(x(t))\exp(-\frac{1}{\hbar}S_E[x]).\end{aligned} \tag{7.2.8}$$

Again we consider the limit $\beta \to \infty$, but we also separate the operators in time, i.e. we also let $t \to \infty$. 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

$$\lim_{\beta,t\to\infty} \langle \mathcal{O}(x(0))\mathcal{O}(x(t))\rangle - |\langle \mathcal{O}(x)\rangle|^2 = |\langle 0|\mathcal{O}(x)|1\rangle|^2 \exp(-(E_1 - E_0)t). \tag{7.2.9}$$

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 lattice QCD hadron masses are determined from the exponential decay of connected 2-point correlation functions.

7.3 The Ising Model 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 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 is 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.) The Ising spins live on the sites of a d-dimensional hypercubic 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

$$\mathcal{H}[s] = -J \sum_{\langle xy \rangle} s_x s_y - \mu B \sum_x s_x, \tag{7.3.1}$$

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

$$Z = \sum_{configurations} \exp(-\beta \mathcal{H}[s]) = \prod_x \sum_{s_x = \pm 1} \exp(-\beta \mathcal{H}[s]). \tag{7.3.2}$$

The sum over all spin configuration 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

$$\langle s \rangle = \frac{1}{Z} \prod_x \sum_{s_x = \pm 1} s_x \exp(-\beta \mathcal{H}[s]). \tag{7.3.3}$$

Similarly, the spin correlation function is defined by

$$\langle s_x s_y \rangle = \frac{1}{Z} \prod_x \sum_{s_x = \pm 1} s_x s_y \exp(-\beta \mathcal{H}[s]). \tag{7.3.4}$$

At large distances the connected spin correlation function typically decays exponentially

$$\langle s_x s_y \rangle - \langle s \rangle \sim \exp(-|x - y|/\xi), \tag{7.3.5}$$

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, where 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 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 huge 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.

7.4 Analogies between Quantum Mechanics and Classical 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 the Ising model can be defined on a 1-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 $\beta = 1/T$. Indeed $\hbar$ determines the strength of quantum fluctuations, while the temperature T determines the strength of thermal fluctuations. The kinetic energy $((x_{i+1} - x_i)/a)^2/2$ 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 $\mu B s_x$ to an external magnetic field. The magnetization $\langle s \rangle$ 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 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 artefact 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 7.1.

7.5 The Transfer Matrix

The analogy between quantum mechanics and classical statistical mechanics suggests that there is an analog of the quantum Hamilton operator in the context of classical statistical mechanics. This operator is the so-called

Quantum mechanics	Classical statistical mechanics
Euclidean time lattice	1-dimensional spatial lattice
elementary time step a	crystal lattice spacing
particle position $x(t)$	classical spin s_x
particle path	spin configuration
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 $T = 1/\beta$
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 $V(x)$	external field $\mu B s_x$
weight of a path $\exp(-\frac{1}{\hbar}S_E[x])$	Boltzmann factor $\exp(-\beta\mathcal{H}[s])$
vacuum expectation value $\langle 0\|\mathcal{O}(x)\|0\rangle$	magnetization $\langle s\rangle$
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 7.1: *The dictionary that translates quantum mechanics into the language of classical statistical mechanics.*

transfer matrix. The Hamilton operator induces infinitesimal translations in time. In classical statistical mechanics, on the other hand, the analog of continuous time is a 1-dimensional spatial lattice. Hence, the transfer matrix cannot induce infinitesimal space translations. Instead it induces translations by the smallest possible distance — namely by one lattice spacing. For a quantum mechanical system the transfer matrix transports us by one lattice spacing in Euclidean time, and it is given by

$$T = \exp(-\frac{a}{\hbar}H). \tag{7.5.1}$$

Now we want to construct the transfer matrix for the 1-dimensional Ising model without an external magnetic field. The corresponding partition function is given by

$$Z = \prod_x \sum_{s_x=\pm 1} \exp(\beta J \sum_x s_x s_{x+1}). \tag{7.5.2}$$

The transfer matrix obeys

$$Z = \mathrm{Tr} T^N, \tag{7.5.3}$$

where N is the number of lattice points, and its matrix elements are given by the Boltzmann factor corresponding to a nearest neighbor pair by

$$\langle s_{x+1}|T|s_x\rangle = \exp(\beta J s_x s_{x+1}). \tag{7.5.4}$$

This is a 2×2 matrix. The eigenvalues of the transfer matrix can be written as $\exp(-E_0)$ and $\exp(-E_1)$. The energy gap then determines the inverse correlation length as

$$1/\xi = E_1 - E_0. \tag{7.5.5}$$

It is instructive to compute ξ as a function of β to locate the critical point of the 1-d Ising model.

Here we will do the corresponding calculation for the 1-d xy-model. In the xy-model the spins are unit vectors $(\cos\varphi_x, \sin\varphi_x)$ in the xy-plane that are attached to the points x of a d-dimensional lattice. Here we consider $d = 1$, i.e. we study a chain of xy-spins. The standard Hamilton function of the xy-model is given by

$$\mathcal{H}[\varphi] = J \sum_{\langle xy\rangle} (1 - \cos(\varphi_{x+1} - \varphi_x)). \tag{7.5.6}$$

In complete analogy to the Ising model the transfer matrix is now given by

$$\langle\varphi_{x+1}|T|\varphi_x\rangle = \exp(-\beta J(1-\cos(\varphi_{x+1}-\varphi_x)), \tag{7.5.7}$$

which is a matrix with an uncountable number of rows and columns, because there is a continuum of values for φ_x and φ_{x+1}. Still, we can ask about the eigenvalues of this matrix. For this purpose we consider the Fourier representation

$$\langle\varphi_{x+1}|T|\varphi_x\rangle = \sum_{m\in\mathbf{Z}} \langle\varphi_{x+1}|m\rangle \exp(-\beta J) I_m(\beta J)\langle m|\varphi_x\rangle, \tag{7.5.8}$$

where

$$\langle\varphi_x|m\rangle = \exp(im\varphi_x), \tag{7.5.9}$$

are the eigenvectors of the transfer matrix. The eigenvalues are given in terms of modified Bessel functions

$$\exp(-E_m) = \exp(-\beta J) I_m(\beta J). \tag{7.5.10}$$

The energy gap between the ground state and an excited state is given by

$$E_m - E_0 = \log \frac{I_0(\beta J)}{I_m(\beta J)}, \tag{7.5.11}$$

which is nonzero for finite β. In the zero temperature limit $\beta \to \infty$ we have

$$\frac{I_0(\beta J)}{I_m(\beta J)} \sim 1 + \frac{m^2}{2\beta J}, \tag{7.5.12}$$

such that

$$\xi = 1/(E_1 - E_0) \sim 2\beta J \to \infty. \tag{7.5.13}$$

Hence, there is a critical point at zero temperature. In the language of quantum mechanics this implies the continuum limit of a Euclidean lattice theory corresponding to a quantum mechanical problem. In the continuum limit the energies corresponding to the eigenvalues of the transfer matrix take the form

$$E_m - E_0 \sim \frac{m^2}{2\beta J}. \tag{7.5.14}$$

These energies are in lattice units (the lattice spacing was put to 1). Hence, to extract physics we need to consider energy ratios and we find

$$\frac{E_m - E_0}{E_1 - E_0} \sim m^2. \tag{7.5.15}$$

These are the appropriate energy ratios of a quantum rotor — a particle that moves on a circle. Indeed the xy-spins describe an angle, which can be interpreted as the position of the quantum particle. Also the eigenvectors of the transfer matrix are just the energy eigenfunctions of a quantum rotor. Hence, we just solved the Schrödinger equation with a discrete Euclidean time step using the transfer matrix instead of the Hamilton operator. The fact that energy ratios approach physically meaningful constants in the continuum limit is known as scaling. Of course, the discretization introduces an error as long as we are not in the continuum limit. For example, at finite β the energy ratio is

$$\frac{E_m}{E_1} = \frac{\log(I_0(\beta J)/I_m(\beta J)}{\log(I_0(\beta J)/I_1(\beta J)}, \tag{7.5.16}$$

which is different from the continuum answer m^2. This cut-off effect due to a finite lattice spacing is known as a scaling violation.

There are formulations of the path integral that are free of cut-off effects and hence exhibit perfect scaling. The actions in these path integrals are known as perfect actions. As an example of a perfect action let us consider the 1-d xy-model with the so-called Villain action. Then the transfer matrix is modified to

$$\langle\varphi_{x+1}|T|\varphi_x\rangle = \sum_{n\in\mathbf{Z}} \exp(-\frac{\beta J}{2}(\varphi_{x+1} - \varphi_x + 2\pi n)^2). \tag{7.5.17}$$

By a Fourier transformation we now find

$$\langle\varphi_{x+1}|T|\varphi_x\rangle = \sum_{m\in\mathbf{Z}} \langle\varphi_{x+1}|m\rangle \exp(-\frac{m^2}{2\beta J})\langle m|\varphi_x\rangle, \tag{7.5.18}$$

which leads to the same eigenvectors but to new eigenvalues

$$E_m = \frac{m^2}{2\beta J}, \tag{7.5.19}$$

which lead to the exact continuum ratio $(E_m - E_0)/(E_1 - E_0) = m^2$ even at finite temperature, i.e. even away from the continuum limit. Perfect actions are of great importance, because they yield exact continuum physics although the lattice spacing is still finite. In general, it is difficult to construct perfect actions for interacting field theories. Later we will see how perfect actions can be constructed for simple free field theories.

7.6 Scalar 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 $\varphi(x)$ of a quantum field defined at any space-time point x. Again the weight factor in the path integral is given by the classical action. Let us illustrate this for a free scalar field $\varphi(x) \in R$. Its classical Euclidean action is given by

$$S[\varphi] = \int d^4x [\frac{1}{2}\partial_\mu \varphi(x)\partial_\mu \varphi(x) + \frac{m^2}{2}\varphi(x)^2]. \tag{7.6.1}$$

Interactions can be included e.g. by adding a $\lambda\varphi(x)^4$ term to the action. The Feynman path integral for this system is formally written as

$$Z = \int \mathcal{D}\varphi \exp(-S_{[}\varphi]). \tag{7.6.2}$$

(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. Of course, 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 essential for strongly interacting theories like QCD. The above free scalar field theory, of course, does not require nonperturbative treatment. We use it only to illustrate the lattice quantization method in a simple setting. On the lattice the continuum field $\varphi(x)$ is replaced by a lattice field Φ_x, which is restricted to the points x of a d-dimensional 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

$$S[\Phi] = \sum_{x,\mu} \frac{1}{2}(\Phi_{x+\hat{\mu}} - \Phi_x)^2 + \sum_x \frac{m^2}{2}\Phi_x^2. \tag{7.6.3}$$

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

$$Z = \prod_x \int_\infty^\infty d\Phi_x \exp(-S[\Phi]). \tag{7.6.4}$$

For a free field theory the partition function is just a Gaussian integral. In fact, one can write the lattice action as

$$S[\Phi] = \frac{1}{2} \sum_{x,y} \Phi_x \mathcal{M}_{xy} \Phi_y, \tag{7.6.5}$$

where the matrix $\mathcal{M}$ describes the couplings between lattice points. Diagonalizing this matrix by an orthogonal transformation $\mathcal{O}$ one has

$$\mathcal{M} = \mathcal{O}^T \mathcal{D} \mathcal{O}. \tag{7.6.6}$$

Introducing

$$\Phi'_x = \mathcal{O}_{xy} \Phi_y \tag{7.6.7}$$

one obtains

$$Z = \prod_x \int d\Phi'_x \exp(-\frac{1}{2} \sum_x \Phi'_x \mathcal{D}_{xx} \Phi'_x) = (2\pi)^{N/2} \det \mathcal{D}^{-1/2}, \tag{7.6.8}$$

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

$$\langle \Phi_x \Phi_y \rangle = \frac{1}{Z} \int \mathcal{D}\Phi \ \Phi_x \Phi_y \exp(-S[\Phi]). \tag{7.6.9}$$

This is most conveniently done by introducing a source field in the partition function, such that

$$Z[J] = \int \mathcal{D}\Phi \exp(-S[\Phi] + \sum_x J_x \Phi_x). \tag{7.6.10}$$

Then the connected 2-point function is given by

$$\langle \Phi_x \Phi_y \rangle - \langle \Phi \rangle^2 = \frac{\partial^2 \log Z[J]}{\partial J_x \partial J_y}|_{J=0}. \tag{7.6.11}$$

The Boltzmann factor characterizing the problem with the external sources is given by the exponent

$$\frac{1}{2}\Phi\mathcal{M}\Phi - J\Phi = \frac{1}{2}\Phi'\mathcal{M}\Phi' - \frac{1}{2}J\mathcal{M}^{-1}J. \tag{7.6.12}$$

Here we have introduced

$$\Phi' = \Phi - \mathcal{M}^{-1}J. \tag{7.6.13}$$

Integrating over Φ' in the path integral we obtain

$$Z[J] = (2\pi)^{N/2}\det\mathcal{D}^{-1/2}\exp(\frac{1}{2}J\mathcal{M}^{-1}J), \tag{7.6.14}$$

and hence

$$\langle\Phi_x\Phi_y\rangle = \frac{1}{2}\mathcal{M}^{-1}_{xy}. \tag{7.6.15}$$

It is instructive to invert the matrix $\mathcal{M}$ by going to Fourier space, i.e. by writing

$$\Phi_x = \frac{1}{(2\pi)^d}\int_B d^dp\ \Phi(p)\exp(ipx). \tag{7.6.16}$$

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

$$\langle\Phi(-p)\Phi(p)\rangle = [\sum_\mu(2\sin(p_\mu/2))^2 + m^2]^{-1}. \tag{7.6.17}$$

This is the lattice version of the continuum propagator

$$\langle\varphi(-p)\varphi((p)\rangle = (p^2+m^2)^{-1}. \tag{7.6.18}$$

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

$$\Phi(\vec{p})_t = \sum_x \Phi_{\vec{x},t}\exp(-i\vec{p}\cdot\vec{x}), \tag{7.6.19}$$

and we consider its 2-point function

$$\langle\Phi(-\vec{p})_0\Phi(\vec{p})_t\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} dp_d\langle\Phi(-p)\Phi(p)\rangle\exp(ip_dt). \tag{7.6.20}$$

Inserting the lattice propagator of eq.(7.6.17) one can perform the integral. One encounters a pole in the propagator when $p_d = iE$ with

$$(2\sinh(E/2))^2 = \sum_i (2\sin(p_i/2))^2 + m^2. \tag{7.6.21}$$

The 2-point function then takes the form

$$\langle \Phi(-\vec{p})_0 \Phi(\vec{p})_t \rangle = C \exp(-Et), \tag{7.6.22}$$

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

$$E^2 = \vec{p}^2 + m^2. \tag{7.6.23}$$

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. Again, we observe scaling violations, i.e. deviations from continuum results, as long as we are not in the continuum limit. This shows that the standard lattice action is not perfect.

It is easy to derive a perfect lattice propagator for a free scalar field. For this purpose we construct the lattice field Φ_x by averaging the continuum scalar field $\varphi(y)$ over a hypercube c_x of side length 1 centered at x

$$\Phi_x = \int_{c_x} d^d y \; \varphi(y). \tag{7.6.24}$$

In momentum space this relation takes the form

$$\Phi(p) = \sum_{l \in \mathbf{Z}^d} \varphi(p + 2\pi l) \Pi(p + 2\pi l), \tag{7.6.25}$$

where

$$\Pi(p) = \prod_{\mu=1}^{d} \frac{2\sin(p_\mu/2)}{p_\mu}. \tag{7.6.26}$$

Note that the lattice field $\Phi(p)$ is periodic over the Brillouin zone because it is constructed from the continuum field $\varphi(p + 2\pi l)$ by summing over all

integer vectors $l \in \mathbf{Z}^d$. Now we compute the lattice propagator by using the propagator of the continuum theory

$$\begin{aligned}\langle\Phi(-p)\Phi(p)\rangle &= \sum_{l,m\in\mathbf{Z}^d}\langle\varphi(-p-2\pi m)\varphi(p+2\pi l)\rangle\Pi(-p-2\pi m)\Pi(p+2\pi l)\\ &= \sum_{l\in Z^d}[(p+2\pi l)^2+m^2]^{-1}\Pi(p+2\pi l)^2. \qquad (7.6.27)\end{aligned}$$

This lattice propagator defines the inverse of a new matrix $\mathcal{M}$, which in turn defines the couplings of a lattice action. By performing the appropriate Fourier transforms one finds that this action includes couplings to distant neighbors, in principle even to infinitely distant ones. However, the value of the coupling is suppressed exponentially with the distance, and hence the action is still quite local. We will see that this action is in fact perfect. This follows from

$$\begin{aligned}\langle\Phi(-\vec{p})_0\Phi(\vec{p})_t\rangle &= \frac{1}{2\pi}\int_{-\pi}^{\pi}dp_d\langle\Phi(-p)\Phi(p)\rangle\exp(ip_dt)\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}dp_d\sum_{\vec{l}\in\mathbf{Z}^{d-1}}[p_d^2+(\vec{p}+2\pi\vec{l})^2+m^2]^{-1}\Pi(p+2\pi l)^2.\end{aligned} \qquad (7.6.28)$$

Here we have combined the integration over $]-\pi,\pi]$ with the sum over $l_d \in \mathbf{Z}$ to an integral over R. Performing the integral we again encounter contributions from poles which are now located at $p_d = iE$ with

$$E = (\vec{p}+2\pi\vec{l})^2+m^2. \qquad (7.6.29)$$

These are exactly the energies of the continuum theory. In particular, momenta outside the Brillouin zone can be reached because we sum over $\vec{l} \in \mathbf{Z}^{d-1}$. It is remarkable that this lattice theory has the correct continuum spectrum, although its action is manifestly invariant only under discrete translations and rotations. This is because we have derived the lattice theory from the continuum by averaging the continuum field over hypercubes. The averaging procedure reduces the continuum symmetries to those of the lattice, but the physics is unaffected, because we keep contact with the underlying continuum theory. The averaging procedure can be viewed as a renormalization group transformation that integrates out short distance variables on scales smaller than the lattice spacing. The resulting lattice theory has an effective action which is perfect, because it still contains all informations about the full spectrum of the continuum theory.

7.7 Fermionic Path Integrals and Grassmann Algebras

We have defined the path integral by using the classical action. 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

$$\{\eta_i, \eta_j\} = \eta_i \eta_j + \eta_j \eta_i = 0. \tag{7.7.1}$$

An element of the Grassmann algebra is a polynomial in the generators

$$f(\eta) = f + \sum_i f_i \eta_i + \sum_{ij} f_{ij} \eta_i \eta_j + \sum_{ijk} f_{ijk} \eta_i \eta_j \eta_k + ... \tag{7.7.2}$$

The $f_{ij...l}$ are ordinary complex (or sometimes real) numbers, which are antisymmetric in $i, j, ..., l$. One also defines formal differentiation and integration procedures. The differentiation rules are

$$\frac{\partial}{\partial \eta_i} \eta_i = 1, \ \frac{\partial}{\partial \eta_i} \eta_i \eta_j = \eta_j, \ \frac{\partial}{\partial \eta_i} \eta_j \eta_i = -\eta_j, \tag{7.7.3}$$

and integration is defined by

$$\int d\eta_i = 0, \ \int d\eta_i \ \eta_i = 1, \ \int d\eta_i d\eta_j \ \eta_i \eta_j = -1. \tag{7.7.4}$$

These integrals are formal expressions. One should not ask over which range of η_i 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 or color 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

$$\int d\bar{\Psi} d\Psi \ \exp(-m\bar{\Psi}\Psi) = \int d\bar{\Psi} d\Psi \ (1 - m\bar{\Psi}\Psi) = m. \tag{7.7.5}$$

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

$$\prod_x \int d\bar{\Psi}_x d\Psi_x \ \exp(-\bar{\Psi}_x \mathcal{M}_{xy} \Psi_y) = \int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi \ \exp(-\bar{\Psi}\mathcal{M}\Psi) = \det\mathcal{M}. \tag{7.7.6}$$

In the two variable case we have

$$\int d\bar{\Psi} d\Psi \ \bar{\Psi}\Psi \exp(-m\bar{\Psi}\Psi) = 1, \tag{7.7.7}$$

which generalizes to

$$\int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi \ \bar{\Psi}_x \Psi_y \exp(-\bar{\Psi}\mathcal{M}\Psi_y) = \mathcal{M}^{-1}_{ij} \det\mathcal{M}. \tag{7.7.8}$$

7.8 The Naive Lattice Fermion Action and the Doubling Problem

In the continuum the Euclidean free fermion action is given by

$$S[\bar{\psi}, \psi] = \int d^d x \ \bar{\psi}(x)(\gamma_\mu \partial_\mu + m)\psi(x). \tag{7.8.1}$$

In Euclidean space the Dirac matrices γ_μ are Hermitean. Again the partition function

$$Z = \int \mathcal{D}\bar{\psi}\mathcal{D}\psi \ \exp(-S[\bar{\psi}, \psi]) \tag{7.8.2}$$

is a formal expression that requires regularization. On the lattice the continuum fermion field $\bar{\psi}(x), \psi(x)$ is replaced by variables $\bar{\Psi}_x, \Psi_x$ that live on the lattice points x. The continuum derivative is discretized by finite differences, such that

$$S[\bar{\Psi}_x, \Psi_x] = \sum_{x,\mu} \frac{1}{2}(\bar{\Psi}_x \gamma_\mu \Psi_{x+\hat{\mu}} - \bar{\Psi}_{x+\hat{\mu}} \gamma_\mu \Psi_x) + \sum_x m \bar{\Psi}_x \Psi_x. \tag{7.8.3}$$

It is instructive to bring this expression to momentum space, because it allows us to read off the lattice fermion propagator

$$\langle \bar{\Psi}(-p)\Psi(p) \rangle = [i \sum_\mu \gamma_\mu \sin p_\mu + m]^{-1}. \tag{7.8.4}$$

In complete analogy to the scalar field case we can analyze the fermionic 2-point function for its exponential decay. The energies E of lattice fermions with spatial momentum $\vec{p}$ show up as poles in the propagator. Here the lattice dispersion relation takes the form

$$\sinh^2 E = \sum_i \sin^2 p_i + m^2. \tag{7.8.5}$$

Like in the scalar field case the continuum dispersion relation is recovered in the continuum limit for small E, $\vec{p}$ and m. However, there are other momenta where E becomes small for small m. These are at the corners of the Brillouin zone at which at least one component of the momentum vector takes the value $p_i = \pi$, because also then $\sin p_i$ goes to zero. This is in contrast to the scalar field case, and it is due to the fact that the fermionic action contains only a single derivative. As a consequence, the lattice dispersion relation leads to extra states in the spectrum that are absent in the continuum theory. The extra states do not disappear in the continuum limit, such that the naive lattice discretization of the fermionic action does not lead to the correct continuum theory. The extra states appearing in the lattice dispersion relation show up as extra physical particles. These so-called doubler fermions are a manifestation of a fundamental problem of lattice regularized fermionic theories. This so-called doubling problem leads to a multiplication of fermion species. The above lattice fermion propagator has 2^d poles instead of just 1 as in the continuum. Hence, the naively discretized fermion theory contains $2^d - 1$ extra fermion species. The origin of the doubling problem is deeply connected with chiral symmetry and can even be traced back to the Adler-Bell-Jackiw anomaly. The doubler fermions pose a severe problem in lattice field theory. Without removing them we cannot describe QCD (or the Standard model) which has definitely less than 16 quarks.

7.9 The Nielsen-Ninomiya No-Go Theorem

Before we try to eliminate the doubler fermions let us prove a general theorem due to Nielsen and Ninomiya: a chiral invariant free fermion lattice action, which is translation invariant, Hermitean, and local, necessarily has fermion doubling. The theorem is based on topology. In fact, it holds because the lattice momentum space (the Brillouin zone) is a torus. A general

chiral and translation invariant lattice action leads to a propagator

$$\langle \bar{\Psi}(-p)\Psi(p)\rangle = [i\sum_{\mu}\gamma_\mu\rho_\mu(p)]^{-1}. \tag{7.9.1}$$

Of course, the mass term is now absent because it breaks chiral symmetry. The above action is Hermitean only if $\rho_\mu(p)$ is real. Locality of the action, i.e. exponential suppression of couplings at large distances, implies that $\rho_\mu(p)$ is regular in momentum space, i.e. there are no poles or discontinuities. Most important, translation invariance implies that $\rho_\mu(p)$ is periodic over the Brillouin zone. Poles of the propagator — and hence physical or doubler fermions — correspond to points p with $\rho_\mu(p) = 0$ for all μ. Hence, the theorem states that with the above properties, $\rho_\mu(p) = 0$ at more than just one point. Let us prove this first in $d = 1$. Then there is a single regular periodic function $\rho_1(p)$, which should at least have one zero to describe the physical fermion pole. Then the function is positive on one side of the zero and negative on the other side, such that it must go through zero again in order to satisfy periodicity. A double zero (with the function not changing sign) is not allowed, because this would lead to a wrong dispersion relation for the physical fermion. In higher dimensions the proof is analogous. In $d = 2$ there are two functions $\rho_1(p)$ and $\rho_2(p)$. The zeros of $\rho_1(p)$ lie on a closed curve in the 2-dimensional Brillouin zone. This curve may be closed via the periodic boundary conditions. The zeros of $\rho_2(p)$ lie on another closed curve that intersects the first one in the pole position of the physical fermion. Due to the periodic boundary conditions of the Brillouin zone, the curves than necessarily must also intersect somewhere else. Again, the curves cannot just touch each other, because this would lead to a wrong dispersion relation. In d dimensions the zeros of $\rho_\mu(p)$ (with $\mu = 1, 2, ..., d$) lie on d closed $d-1$-dimensional manifolds. Those cannot intersect in just one point. If they intersect once they necessarily intersect also somewhere else. This proves lattice fermion doubling for a chiral invariant lattice action. The theorem does not specify the number of doubler fermions. It is definitely possible to reduce the number of doublers from $2^d - 1$ to 1, but it is impossible to eliminate doublers completely. Of course, one may try to evade the theorem by violating one of its basic assumptions. In fact, people have worked on random lattices because it has no translation invariance, and hence no periodic momentum space. However, there is not much one can do analytically on a random lattice, even for a free fermion, and it is unclear if the consequences of the no-go theorem are indeed evaded. When

one violates Hermiticity one looses contact with Minkowski space and it is unclear what the Euclidean results mean. People have also worked with nonlocal actions. Then $\rho_\mu(p)$ is not a continuous function (it may have poles or discontinuities), and the theorem which relies on topology obviously does not apply. Still, it has turned out that most nonlocal actions have serious problems. For example, the resulting continuum theory may be nonlocal or not Lorentz invariant. Still, as we will see later, a perfect action exists that is nonlocal, and still has no problems in the continuum limit. This is the only example I know in which the theorem is evaded without affecting the physics and without explicitly breaking chiral symmetry. In his first paper on lattice gauge theory Wilson removed the fermion doublers in a very direct and radical way, simply by breaking chiral symmetry explicitly. Then the theorem is evaded because the propagator then contains terms without γ_μ. Of course, this leads to all kinds of complications. In particular, it is nontrivial that chiral symmetry will be recovered in the continuum limit. For vector-like theories like QCD this is under control. For chiral theories like the Standard model, however, it is still unclear if we can define them nonperturbatively due to the notorious fermion doubling problem.

7.10 Wilson Fermions and Perfect Fermions

In his original work on lattice gauge theory Wilson has avoided the fermion doubling problem by breaking chiral symmetry explicitly. The so-called Wilson term gives the doubler fermions a mass at the order of the cut-off while the physical fermion remains massless. Hence, in the continuum limit chiral symmetry is restored in the physical sector. Wilson's modification of the naive fermion action takes the form of a discretized second derivative

$$\begin{aligned} S[\bar{\Psi},\Psi] &= \sum_{x,\mu}\frac{1}{2}(\bar{\Psi}_x\gamma_\mu\Psi_{x+\hat{\mu}} - \bar{\Psi}_{x+\hat{\mu}}\gamma_\mu\Psi_x) + \sum_x \bar{\Psi}_x\Psi_x \\ &+ \frac{1}{2}\sum_{x,\mu}(2\bar{\Psi}_x\Psi_x - \bar{\Psi}_x\Psi_{x+\hat{\mu}} - \bar{\Psi}_{x+\hat{\mu}}\Psi_x). \end{aligned} \tag{7.10.1}$$

Then the lattice propagator takes the form

$$\langle\bar{\Psi}(-p)\Psi(p)\rangle = [i\sum_\mu \gamma_\mu \sin p_\mu + m + \frac{1}{2}\sum_\mu (2\sin(p_\mu/2))^2]^{-1}. \tag{7.10.2}$$

It is instructive to work out the corresponding dispersion relation, and in particular to determine the fermion mass as a function of the bare mass m. The Wilson term acts as a momentum dependent mass term. For small momenta it vanishes quadratically, and hence it does not effect the dispersion of the physical fermion at least in the continuum limit. For the doubler fermions, on the other hand, the Wilson term is nonzero, and effectively gives them a mass of the order of the cut-off. In the continuum limit the doubler fermions are hence eliminated from the spectrum of the theory.

In complete analogy to the free scalar field case one can also derive a perfect fermion action by blocking from the continuum. For this purpose we average the continuum fermion field ψ over hypercubes

$$\Psi_x = \int_{c_x} d^d y \; \psi(y), \tag{7.10.3}$$

which in momentum space corresponds to

$$\Psi(p) = \sum_{l \in \mathbf{Z}^d} \psi(p + 2\pi l) \Pi(p + 2\pi l). \tag{7.10.4}$$

The lattice fermion propagator is then related to the continuum propagator by

$$\langle \bar{\Psi}(-p) \Psi(p) \rangle = \sum_{l \in \mathbf{Z}^d} [i \gamma_\mu (p_\mu + 2\pi l_\mu) + m]^{-1} \Pi(p + 2\pi l)^2. \tag{7.10.5}$$

This lattice propagator corresponds to a lattice action

$$S[\bar{\Psi}, \Psi] = \sum_{x,y} \bar{\Psi}_x \gamma_\mu \rho_\mu (x - y) \Psi_y, \tag{7.10.6}$$

with couplings $\rho_\mu(x-y)$ calculable by a Fourier transformation. This lattice action is perfect by construction, i.e. its spectrum is identical with the one of the continuum theory, and hence there should be no fermion doubling. On the other hand, the action is manifestly chiral invariant. This seems to contradict the no-go theorem. However, the theorem is evaded because the action turns out to be nonlocal. Its couplings $\rho_\mu(x - y)$ do not decay exponentially at large distances, instead they decay only power-like. As a consequence in momentum space $\rho_\mu(p)$ is not regular (it actually has poles) and therefore the topological arguments behind the Nielsen-Ninomiya

Figure 7.1: *The inverse propagator $\rho_\mu(p)$ (shown over a 2-d Brillouin zone) is not a regular function (it has poles) for the perfect action*

theorem do not apply. The function $\rho_\mu(p)$ is depicted in fig.7.1. This is the only example of a nonlocal action that can solve the fermion doubling problem. Here the nonlocality arises naturally by blocking the theory from the continuum. When one introduces nonlocalities by hand, they usually lead to problems in the continuum limit, e.g. the resulting continuum theory may be nonlocal, or it may not be Lorentz invariant. Still, from a practical point of view the nonlocality of the perfect action is not very convenient. For example, in a numerical simulation it would be very difficult to include couplings to far away neighbors. Also in the framework of perfect actions one may follow Wilson's approach and break chiral symmetry explicitly in order to obtain a local action. This may be done by modifying the above way of blocking from the continuum, which is chirally covariant. If one chooses to break chiral symmetry explicitly in the blocking procedure, the resulting perfect lattice action will not be manifestly chirally invariant, but it will be local. In any case, with a perfect action (local or not) the spectrum of the lattice theory is always identical with the one of the continuum theory.

7.11 Abelian Lattice Gauge Fields

Let us consider an Abelian gauge field a_μ in the continuum. A gauge transformation then amounts to

$$a'_\mu(y) = a_\mu(y) + \partial_\mu \varphi(y). \tag{7.11.1}$$

Since a gauge field is described by a Lorentz vector the corresponding lattice field is naturally associated with the links connecting neighboring lattice points. We denote the lattice gauge potential by $A_{\mu,x}$ where x denotes the center of the corresponding lattice link. The gauge transformation equation for Abelian lattice gauge field then takes the form

$$A'_{\mu,x} = A_{\mu,x} + \Phi_{x-\hat{\mu}/2} - \Phi_{x+\hat{\mu}/2}, \tag{7.11.2}$$

where Φ_x is the lattice gauge transformation associated with the lattice point x. The derivative of the continuum gauge transformation has been replaced by a finite difference, and we have again put the lattice spacing to 1. The field strength of the continuum gauge field is

$$f_{\mu\nu}(y) = \partial_\mu a_\nu(y) - \partial_\nu a_\mu(y), \tag{7.11.3}$$

while the lattice field strength is

$$F_{\mu\nu,x} = A_{\nu,x+\hat{\mu}/2} - A_{\nu,x-\hat{\mu}/2} - A_{\mu,x+\hat{\mu}/2} + A_{\mu,x-\hat{\mu}/2}. \tag{7.11.4}$$

Here x denotes the center of a lattice plaquette (an elementary lattice square) such that the lattice field strength is the oriented sum of gauge links around the plaquette. In complete analogy with the continuum action the standard lattice gauge action takes the form

$$S[A_\mu] = \sum_{\mu\nu,x} \frac{1}{4e^2} F^2_{\mu\nu,x}. \tag{7.11.5}$$

Again one can construct a perfect gauge action by blocking from the continuum. The natural way to block a continuum gauge field is

$$A_{\mu,x} = \int_{c_{x-\hat{\mu}/2}} d^d y \, (1+y_\mu-x_\mu) a_\mu(y) + \int_{c_{x+\hat{\mu}/2}} d^d y \, (1-y_\mu+x_\mu) a_\mu(y). \tag{7.11.6}$$

It is instructive to prove that with this blocking a continuum gauge transformation $\varphi(y)$ induces a lattice gauge transformation

$$\Phi_x = \int_{c_x} d^d y \; \varphi(y). \tag{7.11.7}$$

The derivation of the perfect gauge action is in complete analogy to the scalar and fermionic cases, and will not be presented here. As in the scalar field case, the perfect action turns out to be local.

7.12 The Notion of Lattice Differential Forms

Lattice fields have a natural place to live on the lattice. For example, gauge potentials live on the links, while gauge transformations live on the lattice points and field strength variables live on plaquettes. The notion of lattice differential forms is based on this fact, and it allows to perform complicated manipulations of Abelian lattice fields in a simple and transparent manner.

Let us consider a 4-dimensional hypercubic lattice. It decomposes naturally into k-dimensional oriented elementary cells c_k. These are the lattice points c_0, the links c_1, the plaquettes c_2, the cubes c_3 and the hypercubes c_4. The dual lattice is obtained by shifting the original lattice by half a lattice spacing in each direction. The dual lattice consists of $(4-k)$-dimensional dual cells ${}^*c_k$. The points ${}^*c_4$ of the dual lattice are dual to the hypercubes of the original lattice, the links ${}^*c_3$ are dual to the original cubes, the plaquettes ${}^*c_2$ are dual to the original plaquettes, etc. On the cells c_k one defines fields — the so-called k-forms $\Phi(c_k)$. For example, a gauge transformation is a 0-form because it lives on points, while a gauge potential is a 1-form because it lives on links. To each k-form there is associated a dual $(4-k)$-form simply by

$${}^*\Phi({}^*c_k) = \Phi(c_k). \tag{7.12.1}$$

The exterior differential d increases the rank of a form by one

$$d\Phi(c_{k+1}) = \sum_{c_k \in \partial c_{k+1}} \Phi(c_k), \tag{7.12.2}$$

i.e. $d\Phi$ is a $(k+1)$-form defined on the cells c_{k+1}. It is a sum of the k-forms Φ living on the oriented boundary of c_{k+1}. For example, the field strength

is a 2-form constructed as

$$F = dA, \tag{7.12.3}$$

and a gauge transformation takes the form

$$A' = A + d\Phi. \tag{7.12.4}$$

Applying the exterior differential twice always gives $d^2 = 0$. For example, the Abelian field strength is gauge invariant because

$$F' = dA' = dA + d^2\Phi = dA = F. \tag{7.12.5}$$

The codifferential $\delta =^* d^*$ lowers the rank of a form by one and is given by

$$\delta\Phi(c_{k-1}) = \sum_{^*c_k \in \partial^* c_{k-1}} \Phi(c_k). \tag{7.12.6}$$

For example, the continuum Landau gauge condition $\partial_\mu a_\mu(y) = 0$ takes the form $\delta A = 0$ on the lattice. The Laplacian $\Delta = d\delta + \delta d$ leaves the rank of a form unchanged. Each form can be written as a Hodge decomposition

$$\Phi = d\Delta^{-1}\delta\Phi + \delta\Delta^{-1}d\Phi. \tag{7.12.7}$$

One defines a scalar product of two k-forms Φ and Ψ by

$$(\Phi, \Psi) = \sum_{c_k} \Phi(c_k)\Psi(c_k). \tag{7.12.8}$$

This induces a norm $||\Phi||^2 = (\Phi, \Phi)$. For example, the standard lattice gauge action takes the form

$$S[A] = \frac{1}{2e^2}||F||^2 = \frac{1}{2e^2}(dA, dA). \tag{7.12.9}$$

The rule for partial integration on the lattice simply reads

$$(d\Phi, \Psi) = (\Phi, \delta\Psi), \tag{7.12.10}$$

where Φ is a k-form, and Ψ is now a $(k+1)$-form. For a gauge field in the Landau gauge one can rewrite the gauge action as

$$S[A] = \frac{1}{2e^2}(A, \delta dA) = \frac{1}{2e^2}(A, \delta dA + d\delta A) = \frac{1}{2e^2}(A, \Delta A). \tag{7.12.11}$$

The notion of differential forms is very useful, in particular because it allows us to perform partial integrations on the lattice in a simple manner.

7.13 Wilson loops and the Lattice Coulomb Potential

Let us consider an Abelian lattice gauge theory (a theory of free photons) coupled to external electric currents. The currents may represent infinitely heavy charged particles. In the continuum the interaction of a gauge field with an external current is described by $\int d^d y\, j_\mu(y) a_\mu(y)$. Expressed in form language the corresponding lattice expression reads (J, A). Let us consider the lattice current of a heavy charged particle moving along the links of the lattice. The current is then 1 on the links that form the world line of the particle, and zero on all other links. The Wilson loop observable is associated with the propagation of a charge-anticharge pair that is created and later annihilated. The corresponding world line is a closed loop $\mathcal{C}$ and the so-called Wilson loop is given by

$$W_{\mathcal{C}} = \exp(i(J, A)). \tag{7.13.1}$$

The expectation value of a rectangular Wilson loop of spatial size R and temporal size T is related to the potential $V(R)$ of a static charge-anticharge pair

$$\lim_{T\to\infty} \langle W_{\mathcal{C}} \rangle \sim \exp(-V(R)T). \tag{7.13.2}$$

Let us calculate this quantity for an Abelian gauge theory. The lattice quantized path integral

$$Z = \prod_{c_1} \int_{-\infty}^{\infty} dA(c_1)\ \exp(-\frac{1}{2e^2}||F||^2) \tag{7.13.3}$$

is still undefined because the implicit integration over all gauges results in an infinite factor. The quantization of Abelian (noncompact) gauge fields on the lattice does indeed require gauge fixing. Here we choose to work in the Landau gauge, i.e. we introduce a δ-function constraint in the functional integral

$$Z = \prod_{c_1} \int_{-\infty}^{\infty} dA(c_1)\ \prod_{c_0} \delta(\delta A(c_0)) \exp(-\frac{1}{2e^2}||F||^2). \tag{7.13.4}$$

This expression is now well-defined. In the Landau gauge we can write the Wilson loop expectation value as

$$\langle W_{\mathcal{C}} \rangle \;=\; \frac{1}{Z} \prod_{c_1} \int_{-\infty}^{\infty} dA(c_1)\ \prod_{c_0} \delta(\delta A(c_0)) \exp(-\frac{1}{2e^2}(A, \Delta A) + i(J, A))$$

$$\sim \quad \exp(-\frac{e^2}{2}(J, \Delta^{-1} J)). \tag{7.13.5}$$

Going to momentum space and performing the large T limit one can derive the lattice Coulomb potential from this expression. The lattice Coulomb potential is slightly distorted at short distances due to discretization artifacts, but it agrees with the ordinary Coulomb potential at large separations. The Wilson-loop expectation value is an order parameter of lattice gauge theory. In the case of an Abelian gauge theory it indicates that we are in an ordinary Coulomb phase with the typical interactions between static charged particles.

7.14 Lattice QED

To construct the action of lattice QED we have to add interaction terms to the free fermion (electron) and free gauge field (photon) actions from before. In particular, the free Wilson fermion action

$$\begin{aligned} S[\bar{\Psi}, \Psi] &= \sum_{x,\mu} \frac{1}{2}(\bar{\Psi}_x \gamma_\mu \Psi_{x+\hat{\mu}} - \bar{\Psi}_{x+\hat{\mu}} \gamma_\mu \Psi_x) + \sum_x m \bar{\Psi}_x \Psi_x \\ &+ \frac{1}{2} \sum_{x,\mu} (2\bar{\Psi}_x \Psi_x - \bar{\Psi}_x \Psi_{x+\hat{\mu}} - \bar{\Psi}_{x+\hat{\mu}} \Psi_x) \end{aligned} \tag{7.14.1}$$

is not gauge invariant. Under a $U(1)$ gauge transformation the fermion field transforms as

$$\Psi'_x = \exp(i\Phi_x)\Psi_x, \tag{7.14.2}$$

while the gauge field transforms as

$$A'_{\mu,x} = A_{\mu,x} + \Phi_{x+\hat{\mu}/2} - \Phi_{x-\hat{\mu}/2}. \tag{7.14.3}$$

To make the fermion action gauge invariant we now introduce the parallel transporter field

$$U_{\mu,x} = \exp(-iA_{\mu,x}) \in U(1). \tag{7.14.4}$$

Under a gauge transformation it transforms as

$$U'_{\mu,x} = \exp(i\Phi_{x-\hat{\mu}/2}) U_{\mu,x} \exp(-i\Phi_{x+\hat{\mu}/2}), \tag{7.14.5}$$

such that

$$\begin{aligned} S[\bar{\Psi},\Psi,U_\mu] &= \sum_{x,\mu}\frac{1}{2}(\bar{\Psi}_x\gamma_\mu U_{x+\hat{\mu}/2,\mu}\Psi_{x+\hat{\mu}}-\bar{\Psi}_{x+\hat{\mu}}\gamma_\mu U^+_{x+\hat{\mu}/2,\mu}\Psi_x) \\ &+ \sum_x m\bar{\Psi}_x\Psi_x \\ &+ \frac{1}{2}\sum_{x,\mu}(2\bar{\Psi}_x\Psi_x-\bar{\Psi}_x U_{x+\hat{\mu}/2,\mu}\Psi_{x+\hat{\mu}}-\bar{\Psi}_{x+\hat{\mu}}U^+_{x+\hat{\mu}/2,\mu}\Psi_x) \end{aligned} \tag{7.14.6}$$

is indeed gauge invariant and has the correct naive continuum limit. Note that $^+$ here denotes complex conjugation. The concept of a parallel transporter connecting neighboring lattice sites in a gauge covariant way extends naturally to the continuum. In fact, one can define a parallel transporter along any curve $\mathcal{C}$ connecting the points x_0 and x_1 as

$$U_\mathcal{C} = \exp(-i\int_\mathcal{C} dy_\mu a_\mu(y)). \tag{7.14.7}$$

Under a continuum gauge transformation

$$a'_\mu(y) = a_\mu(y) + \partial_\mu\varphi(y) \tag{7.14.8}$$

the parallel transporter transforms as

$$U'_\mathcal{C} = \exp(i\varphi(x_0))U_\mathcal{C}\exp(-i\varphi(x_1)). \tag{7.14.9}$$

The Wilson loop defined above is nothing but a parallel transport around a closed curve. When one adds the free photon action to the above expression for the electron-photon interaction one obtains the action of lattice QED.

Lattice QED is of theoretical interest because it is unknown how QED behaves for large values of the electric charge. This can only be investigated in a nonperturbative formulation such as lattice gauge theory. Numerical simulations suggest that the renormalized electric charge never becomes large. In fact, some calculations indicate that QED is trivial, i.e. the renormalized electric charge vanishes in the continuum limit. To locate the continuum limit in such calculations it is essential to measure the electron mass, i.e. the inverse correlation length of the electron field. The naive correlation function $\langle\bar{\Psi}_x\Psi_y\rangle$, however, is not gauge invariant, and therefore does not contain information about the electron mass. The reason for the gauge dependence is that the physical electron consists not just of the bare

electric charge carried by the fermion field, but also of the photon cloud surrounding the charge. The charged field of the physical electron can be constructed as follows

$$\Psi_{c,x} = \exp(-i(E_x, A))\Psi_x. \tag{7.14.10}$$

Here E_x is a 1-form describing the electric field of a static charge located at the point x, i.e.

$$\delta E_x = \delta_x, \tag{7.14.11}$$

where δ_x is a 0-form which is 1 at x and zero otherwise. The field E_x is localized in the specific time slice defined by x_d, and it is zero in all other time slices. This is very essential, because otherwise the definition of the electron field would require information about the Euclidean past or future. This would destroy the causal structure of the theory, and would prevent us from constructing a transfer matrix, and hence from rotating our physical results back into Minkowski space-time. Under a gauge transformation one finds

$$\begin{aligned} \Psi'_{c,x} &= \exp(-i(E_x, A'))\Psi'_x = \exp(-i(E_x, A + d\Phi)) \exp(i\Phi_x)\Psi_x \\ &= \exp(-i(\delta E_x, \Phi)) \exp(i\Phi_x)\Psi_{c,x} = \Psi_{c,x}. \end{aligned} \tag{7.14.12}$$

Here $(\delta E_x, \Phi) = (\delta_x, \Phi) = \Phi_x$ has been used. Hence the physical charged field is gauge invariant. Its correlation function $\langle \bar{\Psi}_{c,x} \Psi_{c,y} \rangle$ does indeed decay exponentially with the electron mass.

7.15 Lattice QCD

The concept of the parallel transporter is even more important in non-Abelian gauge theories. Not even the pure gluon part of lattice QCD can be formulated in terms of vector potentials $G_{\mu,x}$. Instead, the whole action is formulated in terms of link parallel transporters

$$U_{x,\mu} = \exp(iG^a_{\mu,x}\lambda_a) = \exp(iG_{\mu,x}) \in SU(3). \tag{7.15.1}$$

Again, the parallel transporter has a meaning already in the continuum. In non-Abelian gauge theory it arises as a path ordered exponential

$$U_{\mathcal{C}} = \mathcal{P} \exp(i \int_{\mathcal{C}} dy_\mu G_\mu(y)) = \lim_{\varepsilon \to 0} \prod_{y \in \mathcal{C}} \exp(i\varepsilon G_\mu(y)). \tag{7.15.2}$$

It is instructive to prove that under a non-Abelian gauge transformation

$$G'_\mu(y) = g(y)(G_\mu(y) + \partial_\mu)g(y)^+, \tag{7.15.3}$$

the parallel transporter transforms as

$$U'_{\mathcal{C}} = g(x_0)U_{\mathcal{C}}g(x_1)^+, \tag{7.15.4}$$

in complete analogy to the Abelian case. Consequently, the quark-gluon interaction part of the lattice QCD action has exactly the same form as the electron-photon interaction part of the lattice QED action of eq.(7.14.6). Of course, now the parallel transporters are $SU(3)$ matrices and the fermion fields have a color index (which is suppressed in our notation).

The pure gluon part of the action, on the other hand, is quite different from the free photon action in QED. This is because the gauge transformation equation (7.15.3) cannot be discretized in a gauge covariant way, when one uses the vector potential G_μ. Instead Wilson has constructed the pure gluon part of the lattice QCD action again using parallel transporters. First, he builds a parallel transporter around an elementary plaquette

$$U_{\mu\nu,x} = U_{\mu,x-\hat{\nu}/2}U_{\nu,x+\hat{\mu}/2}U^+_{\mu,x+\hat{\nu}/2}U^+_{\nu,x-\hat{\mu}/2}. \tag{7.15.5}$$

Then the action is constructed as a sum over all plaquettes

$$S[U_\mu] = \sum_{\mu\nu,x} \frac{1}{2g_s^2}\text{Re Tr}(1 - U_{\mu\nu,x}). \tag{7.15.6}$$

It is a good exercise to prove that this expression reduces to the correct continuum action

$$S[G_\mu] = \int d^d y \ \frac{1}{2g_s^2}\text{Tr}G_{\mu\nu}(y)G_{\mu\nu}(y). \tag{7.15.7}$$

.

To fully define the lattice QCD path integral we also must consider the measure. For the quarks we have the ordinary Grassmann variable integrations. For the gluons we must integrate over the configurations of parallel transporters U_μ. To discuss this, let us concentrate on a pure gluon theory. The lattice partition function of that theory is given by

$$Z = \prod_{x,\mu}\int_{SU(3)} dU_{\mu,x} \ \exp(-S[U_\mu]). \tag{7.15.8}$$

We integrate independently over all link variables using the so-called Haar measure $dU_{\mu,x}$ for each parallel transporter. The Haar measure is a left and right invariant measure, i.e.

$$\int_{SU(3)} dU\ f(g_L U) = \int_{SU(3)} dU\ f(U g_R) = \int_{SU(3)} dU\ f(U), \tag{7.15.9}$$

for any function f and for any $SU(3)$ matrices g_L and g_R. It is convenient to normalize the measure such that

$$\int_{SU(3)} dU\ 1 = 1. \tag{7.15.10}$$

One can show that the Haar measure is unique and that it has the following additional properties

$$\int_{SU(3)} dU\ U_{ij} = 0,\ \int_{SU(3)} dU\ U_{ij} U_{kl} = \frac{1}{3}\delta_{jk}\delta_{il}. \tag{7.15.11}$$

Here ij determines a specific matrix element, i.e. $i, j, k, l \in \{1, 2, 3\}$. For compact groups like $SU(3)$ the integration is over a finite domain. This is in contrast to QED in which we integrated the gauge potential A_μ over all real numbers. This made it necessary to fix the gauge in lattice QED. In lattice QCD, on the other hand, the path integral is finite from the beginning, and gauge fixing is not necessary. This is a big advantage of the formulation using parallel transporters.

The above pure gluon partition function contains a single parameter — the bare gauge coupling g_s. When we want to perform the continuum limit, we must search for values of g_s for which the correlation length of the lattice theory diverges. As always, we are looking for a second order phase transition. To analyze the phase structure of a gauge theory, one needs to study order parameters like the magnetization $\langle s_x \rangle$ in the Ising model. A simple local order parameter like $\langle U_{\mu,x} \rangle$ does not make sense in a gauge theory. This follows from Elitzur's theorem, which states that gauge non-invariant observables vanish in lattice gauge theory (formulated with parallel transporters). This is trivial to prove using eq.(7.15.11). A useful order parameter in a gauge theory must be gauge invariant and, in addition, nonlocal. A good order parameter in the pure gluon theory is the Wilson loop

$$\langle W_{\mathcal{C}} \rangle = \langle \mathrm{Tr} \prod_{\mu,x \in \mathcal{C}} U_{\mu,x} \rangle. \tag{7.15.12}$$

For a rectangular curve with side lengths R and T the Wilson loop describes the instantaneous creation and annihilation of a static quark-antiquark pair at distance R that lives for a time T. As in QED the Wilson loop is related to the static potential

$$\lim_{T\to\infty} \langle W_{\mathcal{C}} \rangle \sim \exp(-V(R)T). \tag{7.15.13}$$

In QCD we expect quarks and antiquarks to be confined to one another by a potential rising linearly at large separations R, i.e.

$$\lim_{R\to\infty} V(R) \sim \sigma R, \tag{7.15.14}$$

where σ is the string tension. In a confinement phase the Wilson loop hence shows an area law

$$\lim_{R,T\to\infty} \langle W_{\mathcal{C}} \rangle \sim \exp(-\sigma RT). \tag{7.15.15}$$

If confinement would be lost (by going through a phase transition) the Wilson loop would no longer show area law behavior. A Coulomb phase, for example, is characterized by a perimeter law.

7.16 Confinement in the Strong Coupling Limit

In lattice gauge theory it is straight forward to prove confinement for large values of the bare gauge coupling g_s. In the strong coupling region, however, the correlation length is small and we cannot take the continuum limit. In fact, due to asymptotic freedom we expect the continuum limit to be at $g_s \to 0$. It is an open question if confinement persists in the continuum limit. However, there is a lot of positive numerical evidence that this is indeed the case. Assuming that there is no phase transition between the strong and weak coupling regions, the derivation of confinement in the strong coupling regime would carry over to the continuum limit. In the strong coupling expansion we expand in powers of $1/g_s$ around $g_s = \infty$. To leading order the pure gluon action is then simply zero. The Wilson loop operator takes the form

$$W_{\mathcal{C}} = U_{1ij} U_{2jk} U_{3kl} ... U_{Nmi}, \tag{7.16.1}$$

where $N = 2(R+T)$ is the number of links along the loop. Using the first formula of eq.(7.15.11) then immediately implies $W_{\mathcal{C}} = 0$. The second

formula of eq.(7.15.11) suggests that to get a nonzero result one needs to have the product of two matrix elements for each link variable on the Wilson loop. This can be achieved by expanding the Boltzmann factor of the action to higher orders in $1/g_s$. The lowest nonzero contribution comes from tiling the Wilson loop with plaquettes that result from expanding

$$\exp(-\frac{1}{2g_s^2}\text{Re Tr}(1-U_{\mu\nu,x})) = \exp(-\frac{3}{2g_s^2})(1+\frac{1}{2g_s^2}\text{Re Tr}U_{\mu\nu,x}) \quad (7.16.2)$$

for the plaquettes in the interior of the Wilson loop. Taking the $U_{\mu\nu,x}$ term for all these plaquettes and using the second formula of eq.(7.15.11) gives

$$\langle W_{\mathcal{C}}\rangle = \frac{1}{(g_s^2)^{RT}}. \quad (7.16.3)$$

Indeed we find an area law and we read off the string tension as

$$\sigma = -\log(\frac{1}{g_s^2}). \quad (7.16.4)$$

Higher order corrections arise from deformations of the simple tiling from above. The leading correction comes from eliminating one of the RT plaquettes of the original tiling, and replacing it with five plaquettes at the surface of an elementary lattice cube, such that the resulting surface bounded by the Wilson loop has no holes. The cube can be attached above and below the plane of the Wilson loop, and it can go out in the two possible orthogonal directions. This results in $4RT$ contributions, which all have four more plaquettes than the leading term, and are hence suppressed by $1/(g_s^2)^4$. Up to that order one finds

$$\langle W_{\mathcal{C}}\rangle = \frac{1}{(g_s^2)^{RT}}(1+4RT\frac{1}{(g_s^2)^4}). \quad (7.16.5)$$

The last two terms are the first term in an exponential, and hence we can read off the corrected string tension as

$$\sigma = -\log(\frac{1}{g_s^2}) + 4\frac{1}{(g_s^2)^4}. \quad (7.16.6)$$

The string tension has been computed to higher orders. Still, one is unable to go to high enough orders to reach the physically interesting scaling region in which we can perform the continuum limit. At present, only numerical

simulation techniques are powerful enough to enter that regime. Consequently, the above result for the string tension cannot yet be compared with experimental results. It is instructive to derive the glueball mass in a similar manner. For this purpose one considers the correlation function of two elementary plaquette operators separated in Euclidean time. Using the strong coupling expansion one can identify the exponential decay, and extract the correlation length whose inverse is the massgap or glueball mass.

7.17 The Confinement Mechanism in Compact Abelian Gauge Theory

The above strong coupling calculations prove that lattice QCD confines at large values of the bare gauge coupling. However, the calculation does not really hint at the dynamical mechanism that confines color. It is a subject of ongoing research to understand the confinement mechanism in non-Abelian gauge theories. In abelian gauge theories, on the other hand, the issue is completely understood, even analytically. Of course, as we saw earlier, the Wilson loop in an Abelian gauge theory yields the Coulomb potential, which does not confine in $d = 4$. It is, however, possible to modify the formulation of Abelian gauge theories, such that they also confine at strong coupling. This is achieved simply by working with parallel transporters in the pure gauge action, as we were forced to do in non-Abelian theories. This leads to the compact formulation of Abelian lattice gauge theory, which is very similar to Wilson's lattice formulation of non-Abelian gauge theory. The use of (compact) parallel transporters instead of (noncompact) vector potentials has drastic consequences in Abelian gauge theory. As we will see, the compact theory has magnetic monopoles as additional degrees of freedom, which turn out to be responsible for confinement. In fact, the monopoles condense in the confinement phase (like Cooper pairs condense in a superconductor), while they are ordinary charged particles in the Coulomb phase.

As we will see, the confinement mechanism in a compact Abelian gauge theory is the dual of the Meissner effect in a superconductor. An external magnetic field is constricted to flux strings within the superconductor. The flux strings cost energy proportional to their length, i.e. they have a

nonzero string tension. If there would be single magnetic monopoles, in a superconductor their magnetic flux would form a string ending in an anti-monopole, and hence monopoles and antimonopoles would be confined to one another. In QCD color electric charges (quarks and gluons) are confined. This could be described by a dual Meissner effect. The Meissner effect in a superconductor arises because electric charges (Cooper pairs) are condensed. A dual superconductor would require the condensation of magnetic charges (monopoles). This is exactly what happens in compact QED. It is still unclear if the same mechanism is responsible for confinement in QCD, although there is some positive numerical evidence.

To understand the role of the monopoles let us now consider 4-d compact $U(1)$ lattice gauge theory. This is lattice QED without electrons, but — as we will see — with magnetic charges. For simplicity we consider the theory in the so-called Villain formulation, for which the partition function takes the form

$$Z = \prod_{c_1} \int_{-\pi}^{\pi} d\theta(c_1) \prod_{c_2} \sum_{n(c_2)\in\mathbf{Z}} \exp(-\frac{1}{2e^2}\|d\theta + 2\pi n\|^2), \tag{7.17.1}$$

where the gauge field θ is a 1-form on the links c_1, n is a 2-form on the plaquettes c_2 and e is the bare electric charge. First we replace the link integration in the path integral by an integration over the plaquette angles $\theta_\Box = d\theta \mathrm{mod} 2\pi$. The Bianchi identity is implemented by a δ-function constraint

$$\begin{aligned} Z &= \prod_{c_2} \int_{-\pi}^{\pi} d\theta_\Box(c_2) \prod_{c_3} \sum_{k(c_3)\in\mathbf{Z}} \delta(\frac{1}{2\pi} d\theta_\Box - k) \\ &\quad \prod_{c_2} \sum_{n(c_2)\in\mathbf{Z}} \exp(-\frac{1}{2e^2}\|\theta_\Box + 2\pi n\|^2) \\ &= \prod_{c_3} \sum_{k(c_3)\in\mathbf{Z}} \prod_{c_2} \int_{-\pi}^{\pi} d\theta_\Box(c_2) \\ &\quad \prod_{c_2} \sum_{n(c_2)\in\mathbf{Z}} \exp(-\frac{1}{2e^2}\|\theta_\Box + 2\pi n\|^2) \exp(i(d\theta_\Box, k)). \end{aligned} \tag{7.17.2}$$

The constraint has been eliminated using the Poisson summation formula. Thereby we have introduced an integer valued 3-form k. Now we perform

the integration over the variable $F = \theta_\Box + 2\pi n$, and we obtain

$$\begin{aligned} Z &= \prod_{c_3} \sum_{k(c_3)\in\mathbf{Z}} \prod_{c_2} \int_{-\infty}^{\infty} dF(c_2) \exp(-\frac{1}{2e^2}\|F\|^2) \exp(i(dF,k)) \\ &= \prod_{c_3} \sum_{k(c_3)\in\mathbf{Z}} \prod_{c_2} \int_{-\infty}^{\infty} dF(c_2) \exp(-\frac{1}{2e^2}\|F\|^2) \exp(i(F,{}^* d^* k)) \\ &= \prod_{{}^*c_3} \sum_{{}^*k({}^*c_3)\in\mathbf{Z}} \exp(-\frac{e^2}{2}\|d^* k\|^2). \end{aligned} \tag{7.17.3}$$

Eq.(7.17.3) describes a $\mathbf{Z}$ gauge theory with dual link variables ${}^*k$ and with a quadratic action.

The dual $\mathbf{Z}$ gauge theory is the $\kappa \to \infty$ limit of scalar QED with gauge group $\mathbf{R}$. The dual noncompact gauge field ${}^*A \in \mathbf{R}$ is a 1-form on the dual links ${}^*c_3$, and the complex scalar field $\Phi_x = \exp(i{}^*\chi_x)$ is defined in terms of the dual 0-form ${}^*\chi \in]-\pi, \pi]$. To make the action periodic in ${}^*\chi$ we also introduce a dual integer valued 1-form ${}^*k \in \mathbf{Z}$ such that the partition function reads

$$\begin{aligned} Z &= \prod_{{}^*c_3} \int_{-\infty}^{\infty} d^* A({}^*c_3) \prod_{{}^*c_4} \int_{-\pi}^{\pi} d^* \chi({}^*c_4) \prod_{{}^*c_3} \sum_{{}^*k({}^*c_3)\in\mathbf{Z}} \\ &\quad \exp(-\frac{1}{2g^2}\|d^* A\|^2 - \frac{\kappa}{2}\|d^*\chi + 2\pi {}^*k - {}^*A\|^2). \end{aligned} \tag{7.17.4}$$

The scalar field Φ_x carries the bare charge g. In the unitary gauge ${}^*\chi = 0$ the partition function is

$$\begin{aligned} Z &= \prod_{{}^*c_3} \int_{-\infty}^{\infty} d^* A({}^*c_3) \prod_{{}^*c_3} \sum_{{}^*k({}^*c_3)\in\mathbf{Z}} \\ &\quad \exp(-\frac{1}{2g^2}\|d^* A\|^2 - \frac{\kappa}{2}\|2\pi {}^*k - {}^*A\|^2). \end{aligned} \tag{7.17.5}$$

Now it is clear that for $\kappa \to \infty$ only configurations with ${}^*A = 2\pi {}^*k$ contribute to the functional integral and that we thus recover the dual $\mathbf{Z}$ gauge theory with

$$Z = \prod_{{}^*c_3} \sum_{{}^*k({}^*c_3)\in\mathbf{Z}} \exp(-\frac{e^2}{2}\|d^* k\|^2) \tag{7.17.6}$$

for $e = 2\pi/g$. This is exactly the Dirac quantization condition for the bare electric and magnetic charges. Since the scalar particle of the dual gauge

theory carries the charge g, it is natural to identify it with the magnetic monopole of the original model. We will see later that such an identification is indeed justified.

Lüscher has derived an inequality between the bare and renormalized charge in noncompact scalar QED in the Coulomb phase

$$g_r \leq g. \tag{7.17.7}$$

Here g_r is the renormalized magnetic charge. The Dirac quantization condition relates the renormalized electric and magnetic charges by

$$e_r = \frac{2\pi}{g_r}. \tag{7.17.8}$$

Duality in the $\kappa \to \infty$ limit together with the Dirac quantization condition thus turns the inequality eq.(7.17.7) into an inequality between the bare and the renormalized electric charge in compact $U(1)$ lattice QED

$$e_r = \frac{2\pi}{g_r} \geq \frac{2\pi}{g} = e. \tag{7.17.9}$$

Consequently, magnetic monopoles lead to antiscreening of electric charge, as opposed to the screening of electric charges in ordinary QED. Now it becomes clear why the presence of monopoles can lead to confinement of electric charges. Because monopoles enhance the renormalized electric charge, they can drive the theory to strong coupling, where Coulomb's law breaks down and confinement sets in. This is in fact what happens in compact lattice QED. At small bare electric charges the theory is in the Coulomb phase. Then the monopoles are stable magnetically charged particles, and external electric charges interact via Coulomb forces. For large bare electric charges the model is in the confined phase, in which the monopoles are condensed and magnetic charge is not a good quantum number. In the dual formulation the confined phase is equivalent to the Higgs phase of scalar QED, in which the scalar charged particle (the monopole) is condensed. The Coulomb phase of compact QED, on the other hand, corresponds to the Coulomb phase of noncompact scalar QED.

In the dual formulation of the theory the monopole is an ordinary charged scalar field. Therefore it is straightforward to construct a creation operator for it. As discussed before the bare charged field Φ_x cannot create the physical charged state, because it is not gauge invariant. Φ_x only

creates the bare monopole, but not the dual photon cloud surrounding it. The gauge dependence of Φ_x is fixed by going to the Coulomb gauge. We perform a gauge transformation ${}^*\varphi$ in the time slice containing the point x

$$\begin{aligned} {}^*A' &= {}^*A + d^*\varphi, \\ {}^*\chi' &= ({}^*\chi + {}^*\varphi) \mathrm{mod} 2\pi. \end{aligned} \tag{7.17.10}$$

Coulomb gauge fixing generates the dual Coulomb field surrounding the monopole. The appropriate gauge transformation at the point x is given by

$${}^*\varphi_x = -({}^*B_x, {}^*A), \tag{7.17.11}$$

with

$$\delta^* B_x = \delta_x, \tag{7.17.12}$$

where δ_x is the Kronecker delta function on the dual lattice, i.e.

$$\delta_x = \begin{cases} 1 & \text{for } y = x \\ 0 & \text{otherwise.} \end{cases} \tag{7.17.13}$$

${}^*B_x$ is the magnetic field of a monopole located at the point x. It is nonzero only on the dual links located in the time slice containing the monopole creation point x. The explicit form of ${}^*B_x$ is given by

$${}^*B_x = d_3 \Delta_3^{-1} \delta_x, \tag{7.17.14}$$

where d_3 is the 3-dimensional exterior differential and Δ_3 is the 3-dimensional Laplacian. In the Coulomb gauge the charged field is

$$\Phi_{c,x} = \exp(i {}^*\chi'_x) = \exp(-i({}^*B_x, {}^*A))\Phi_x. \tag{7.17.15}$$

This has exactly the same form as the charge creation operator discussed earlier, except that we are now using a dual formulation. If we again go to the limit $\kappa \to \infty$ and fix to unitary gauge $\Phi_x = 1$, ${}^*A = 2\pi {}^*k$ we find

$$\Phi_{c,x} = \exp(-2\pi i({}^*B_x, {}^*k)). \tag{7.17.16}$$

This is the monopole field from which one can construct all monopole Green functions. Now we will investigate the vacuum expectation value of the monopole field

$$\langle \Phi_{c,x} \rangle = \frac{1}{Z} \prod_{{}^*c_3} \sum_{{}^*k({}^*c_3) \in \mathbf{Z}} \Phi_{c,x} \exp(-\frac{e^2}{2} \| d^* k \|^2) \tag{7.17.17}$$

after transforming back to the original compact QED. From there one arrives at a Coulomb gas representation of the correlation function, which makes the creation and annihilation of monopoles explicit.

When one performs the duality transformation again one goes back to the original compact lattice QED. The expectation value of the monopole field then takes the form

$$\begin{aligned}\langle \Phi_{c,x} \rangle &= \frac{1}{Z} \prod_{c_1} \int_{-\pi}^{\pi} d\theta(c_1) \prod_{c_2} \sum_{n(c_2) \in \mathbf{Z}} \\ &\exp(-\frac{1}{2e^2} \| d\theta + 2\pi n + 2\pi \delta \Delta^{-1}(B - \omega) \|^2), \quad (7.17.18)\end{aligned}$$

The integer valued 3-form $\omega \in \mathbf{Z}$ is a remnant of the Dirac string which emanates from the monopole position x. It obeys

$$^*(d\omega) = \delta_x \quad (7.17.19)$$

such that $d(B-\omega) = 0$. B and ω give rise to a plaquette shift $2\pi\delta\Delta^{-1}(B-\omega)$ which acts as an external background field.

In the present form it is unclear why eq.(7.17.18) describes the creation of monopoles. Therefore we now rewrite the theory as a Coulomb gas of monopole world lines. First we perform a Hodge decomposition of n

$$n = d\Delta^{-1}\delta n + \delta\Delta^{-1} dn = d\Delta^{-1}\delta n + \delta\Delta^{-1} m, \quad (7.17.20)$$

where the monopoles are described by the 3-form $m = dn$. The dual 1-form $^*m$ describes monopole world lines on the dual links, which form closed loops because of the continuity equation for magnetic charge

$$\delta {}^*m = {}^*dm = {}^*d^2 n = 0. \quad (7.17.21)$$

When n is shifted to $n' = n + dl$, where $l \in \mathbf{Z}$ is a 1-form we still have

$$m' = dn' = dn + d^2 l = dn = m. \quad (7.17.22)$$

One may thus eliminate n from the path integral in favor of m and l

$$\begin{aligned}\langle \Phi_{c,x} \rangle = \frac{1}{Z} \prod_{c_3} \sum_{m(c_3) \in \mathbf{Z}, dm = 0} \prod_{c_1} \sum_{l(c_1) \in \mathbf{Z}} \prod_{c_1} \int_{-\pi}^{\pi} d\varphi(c_1) \\ \exp(-\frac{1}{2e^2} \| d(\varphi + 2\pi l + 2\pi\delta\Delta^{-1} n) + 2\pi\delta\Delta^{-1}(m + B - \omega) \|^2). \\ (7.17.23)\end{aligned}$$

Introducing a noncompact gauge field $A = \varphi + 2\pi l + 2\pi\delta\Delta^{-1}n$ one now writes

$$\begin{aligned}\langle\Phi_{c,x}\rangle &= \frac{1}{Z}\prod_{c_3}\sum_{m(c_3)\in\mathbf{Z},dm=0}\prod_{c_1}\int_{-\infty}^{\infty}dA(c_1)\\ &\exp(-\frac{1}{2e^2}\|dA+2\pi\delta\Delta^{-1}(m+B-\omega)\|^2). \quad (7.17.24)\end{aligned}$$

Using partial integration and $d(m+B-\omega)=0$ one finds

$$\langle\Phi_{c,x}\rangle = \frac{1}{Z}\prod_{c_3}\sum_{m(c_3)\in\mathbf{Z},dm=-dB}\exp(-\frac{2\pi^2}{e^2}(m+B,\Delta^{-1}(m+B))) \quad (7.17.25)$$

with

$$Z = \prod_{c_3}\sum_{m(c_3)\in\mathbf{Z},dm=0}\exp(-\frac{2\pi^2}{e^2}(m,\Delta^{-1}m)). \quad (7.17.26)$$

The expectation value is expressed as a ratio of two partition functions. Z in the denominator describes an ensemble of closed monopole world lines. The monopoles interact with each other via long-range Coulomb forces represented by the inverse Laplacian. In the numerator there is, in addition, one open monopole world line that starts at x. Now it is clear that a monopole has indeed been created at the point x. This shows explicitly that the expectation value, which was originally constructed in the dual scalar QED, does indeed describe the creation of the topological excitations of compact QED. Note that the invisibility of the Dirac string is now obvious, because ω has disappeared from the final expression.

Because of the Dirac quantization condition strong electric couplings e correspond to weak magnetic couplings g. Hence, the strong coupling limit of the original compact QED corresponds to the weak coupling limit of the dual noncompact QED. We can make use of this to show that monopoles are indeed condensed in the confined phase. For this purpose we investigate the monopole field expectation value $\langle\Phi_{c,x}\rangle$ in the strong coupling $e=\infty$ limit, which corresponds to the weak coupling $g=0$ limit in the dual formulation. In this limit only the configurations with zero dual action contribute to the path integral. Then $k=0$ and consequently

$$\langle\Phi_{c,x}\rangle = 1, \quad (7.17.27)$$

such that monopoles indeed condense in the confined phase.

7.18 The Monte-Carlo Method

The most interesting questions in lattice field theory — especially those that may eventually lead to a solution of QCD — cannot be answered analytically. For example, the strong coupling expansion does not converge well in the weak coupling scaling region, in which we want to take the continuum limit of lattice QCD. Fortunately, the close analogy with classical statistical mechanics allows us to use other techniques developed there. A powerful numerical technique to solve problems in statistical mechanics is the so-called Monte-Carlo method. The idea is to compute the partition function (do the path integral) by generating field configurations numerically. Of course, the path integral is an extremely high dimensional integral, such that doing it with standard numerical integration techniques is completely hopeless. In the Monte-Carlo method predominantly those field configurations are generated that have the largest contribution to the path integral. In fact, the Boltzmann factor $\exp(-S[\Phi])$ is used as the probability to generate the field configuration Φ.

In a Monte-Carlo simulation one generates a sequence of field configurations

$$\Phi^{(1)} \to \Phi^{(2)} \to ... \to \Phi^{(N)}, \tag{7.18.1}$$

which form a so-called Markow chain, by applying an algorithm that turns the configuration $\Phi^{(i)}$ into $\Phi^{(i+1)}$. The initial configuration $\Phi^{(1)}$ is either picked at random or selected otherwise. Ultimately, nothing should depend on this choice. After a (possibly large) number M of Monte-Carlo iterations (applications of the algorithm) an equilibrium is reached, and the system has forgotten about the initial configurations. Only the configurations generated after equilibration are used in the actual calculation. To estimate the expectation value of some observable one averages its values over all configurations of the Monte-Carlo sample

$$\langle \mathcal{O}[\Phi] \rangle \approx \frac{1}{N-M} \sum_{i=M+1}^{N} \mathcal{O}[\Phi^{(i)}]. \tag{7.18.2}$$

In the limit $N \to \infty$ the approximation becomes exact. At finite $N - M$ one makes a calculable numerical error that decreases proportional to $1/\sqrt{N-M}$. Hence, to increase the numerical accuracy by a factor of two one must run the Monte-Carlo algorithm four times as long. The Boltzmann factor $\exp(-S[\Phi])$ is not explicitly included in the above sum. It is

implicitly included, because the configurations in the Markow chain occur with probability $\exp(-S[\Phi])$.

To demonstrate that a particular Monte-Carlo algorithm converges to the correct equilibrium distribution it is sufficient to show that it is ergodic and obeys detailed balance. Ergodicity means that starting from an arbitrary initial configuration the algorithm can in principle reach any other field configuration. This condition is obviously necessary, because the correct value for the path integral can be obtained only if all field configurations (of finite action) are included. Detailed balance means that

$$\exp(-S[\Phi])w[\Phi,\Phi'] = \exp(-S[\Phi'])w[\Phi',\Phi]. \tag{7.18.3}$$

Here $w[\Phi,\Phi']$ is the transition probability for the algorithm to turn the configuration Φ into Φ'. A Monte-Carlo algorithm is completely characterized by its $w[\Phi,\Phi']$. Since the algorithm definitely generates a new configuration the proper normalization is

$$\int \mathcal{D}\Phi' \; w[\Phi,\Phi'] = 1. \tag{7.18.4}$$

When the Monte-Carlo algorithm converges to an equilibrium distribution $p[\Phi]$ of field configurations, this distribution is an eigenvector of $w[\Phi,\Phi']$ with eigenvalue 1

$$\int \mathcal{D}\Phi \; p[\Phi]w[\Phi,\Phi'] = p[\Phi']. \tag{7.18.5}$$

Now we want to show that the canonical Boltzmann distribution

$$p[\Phi] = \exp(-S[\Phi]) \tag{7.18.6}$$

is indeed an eigenvector of $w[\Phi,\Phi']$ if the algorithm obeys detailed balance. We find

$$\begin{aligned} \int \mathcal{D}\Phi \;\; \exp(-S[\Phi])w[\Phi,\Phi'] &= \int \mathcal{D}\Phi \;\; \exp(-S[\Phi'])w[\Phi',\Phi] \\ &= \exp(-S[\Phi']) \int \mathcal{D}\Phi w[\Phi',\Phi] \\ &= \exp(-S[\Phi']). \end{aligned} \tag{7.18.7}$$

Assuming ergodicity one can show that only one eigenvector with eigenvalue 1 exists, and that the equilibrium distribution is therefore unique.

A simple example of an algorithm that is ergodic and obeys detailed balance is the so-called Metropolis algorithm. In this algorithm a new configuration Φ' is randomly chosen in the vicinity of the old configuration Φ. If the action of the new configuration is smaller than the action of the old configuration, the new configuration is accepted, i.e.

$$S[\Phi'] < S[\Phi] \Rightarrow w[\Phi, \Phi'] = 1. \tag{7.18.8}$$

On the other hand, if the new action is larger, the new configuration is accepted only with a certain probability, i.e.

$$S[\Phi'] > S[\Phi] \Rightarrow w[\Phi, \Phi'] = \exp(-S[\Phi'] + S[\Phi]). \tag{7.18.9}$$

Otherwise the old configuration is kept. This algorithm obeys detailed balance. Let us consider two configurations Φ and Φ'. We can assume that $S[\Phi'] < S[\Phi]$ such that $w[\Phi, \Phi'] = 1$. Then of course, $S[\Phi] > S[\Phi']$ such that $w[\Phi', \Phi] = \exp(-S[\Phi] + S[\Phi'])$, and hence

$$\begin{aligned} \exp(-S[\Phi])w[\Phi, \Phi'] &= \exp(-S[\Phi]) = \exp(-S[\Phi'])\exp(-S[\Phi] + S[\Phi']) \\ &= \exp(-S[\Phi'])w[\Phi', \Phi]. \end{aligned} \tag{7.18.10}$$

The Metropolis algorithm is particularly simple, but not very efficient. It turns out that subsequent configurations in the Markow chain are correlated with each other. Hence, to generate a new statistically independent field configuration may require a large number of Monte-Carlo iterations. The autocorrelation time τ of the Metropolis algorithm actually increases when one approaches a second order phase transition (or equivalently the continuum limit). At a second order phase transition the correlation length ξ diverges (or equivalently the particle mass in lattice units vanishes). One finds so-called critical slowing down

$$\tau \propto \xi^z, \tag{7.18.11}$$

where z is a dynamical critical exponent characterizing the efficiency of a Monte-Carlo algorithm. For the Metropolis algorithm one finds $z \approx 2$, which leads to a very bad critical slowing down behavior. In lattice gauge theory the best algorithm that is presently known (the so-called overrelaxation algorithm) has $z \approx 1$. For simpler spin models so-called cluster algorithms exist which have $z \approx 0$, and which hence eliminate critical slowing down.

These algorithms are extremely efficient, and allow one to perform very accurate numerical simulations in these models. Unfortunately, for lattice gauge theory no algorithm with $z \approx 0$ is known.

The inclusion of quarks poses additional very hard problems in numerical simulations of QCD. As a matter of fact, computers cannot deal with Grassmann numbers directly. Therefore, one integrates out the fermions analytically and obtains a fermion determinant, that is a very complicated function of the gluon link variables. One includes the determinant in the effective gluon action, which then becomes highly nonlocal. To evaluate the change in action, which is necessary for the Metropolis step, then is a very time-consuming procedure. Simulations with dynamical quarks are therefore orders of magnitude more complicated than simulations of the pure gluon theory. It is a matter of ongoing research to find better ways to handle quarks numerically. Due to the tremendous difficulties to simulate dynamical quarks, people often use the so-called quenched approximation, in which the fermion determinant is simply put to 1. It is unclear how this very drastic approximation affects the physics. Especially for light quarks one expects that the quenched approximation should be bad. On the other hand, for somewhat heavier quarks the approximation seems to work reasonably well.

Chapter 8

Quarks and Gluons in the Early Universe

8.1 The Standard Big Bang Cosmology

General relativity (without a cosmological constant) predicts that we live in an expanding universe that emerged from a space-time singularity — the big bang — and that 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 $\Lambda_{QCD} \approx 0.2\text{GeV}$. 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. To understand this epoch better let us first discuss 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

$$ds^2 = dt^2 - R(t)^2(\frac{d\rho^2}{1-k\rho^2} + \rho^2(d\theta^2 + \sin^2\theta d\varphi^2)) = g_{\mu\nu}dx^\mu dx^\nu. \quad (8.1.1)$$

Note that we are back 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

$$\Gamma^\rho_{\mu\nu} = \frac{1}{2}g^{\rho\sigma}(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}) \quad (8.1.2)$$

one derives the Riemann tensor

$$R^\sigma_{\mu\nu\rho} = \partial_\nu\Gamma^\sigma_{\mu\rho} - \partial_\rho\Gamma^\sigma_{\mu\nu} + \Gamma^\sigma_{\lambda\nu}\Gamma^\lambda_{\mu\rho} - \Gamma^\sigma_{\lambda\rho}\Gamma^\lambda_{\mu\nu}, \quad (8.1.3)$$

which then leads to the Ricci tensor

$$R_{\mu\sigma} = R^\nu_{\mu\nu\sigma}, \quad (8.1.4)$$

the curvature scalar

$$\mathcal{R} = R^\mu_\mu = g^{\mu\nu}R_{\nu\mu}, \quad (8.1.5)$$

and finally to the Einstein tensor

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R}. \quad (8.1.6)$$

For the above metric one finds

$$\begin{aligned} \mathcal{R} &= -6(\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2}), \\ G_{ij} &= (2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2})g_{ij}, \\ G_{00} &= 3(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}), \\ G_{0i} &= G_{i0} = 0. \end{aligned} \quad (8.1.7)$$

Here i, j denote spatial indices. Einstein's field equation (without a cosmological constant) takes the form

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{8.1.8}$$

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

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu - p g_{\mu\nu}, \tag{8.1.9}$$

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

$$T_{ii} = -p g_{ij},\ T_{00} = \rho. \tag{8.1.10}$$

Using Einstein's field equations we now obtain

$$2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2} = -8\pi G p,\ 3(\frac{\dot{R}^2}{R^2} + \frac{k}{R^2}) = 8\pi G\rho. \tag{8.1.11}$$

From this we infer

$$\begin{aligned} &\frac{d}{dt}(8\pi G\rho R^3) + 8\pi G p \frac{d}{dt}R^3 = \\ &\frac{d}{dt}(3R\dot{R}^2 + 3kR) - (2\frac{\ddot{R}}{R}\frac{\dot{R}^2}{R^2} + \frac{k}{R^2})3R^2\dot{R} = 0. \end{aligned} \tag{8.1.12}$$

We identify this as the first law of thermodynamics

$$\frac{d}{dt}(\rho R^3) = -p\frac{d}{dt}R^3, \tag{8.1.13}$$

or equivalently

$$dE = -pdV, \tag{8.1.14}$$

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

$$p = \frac{1}{3}\rho, \tag{8.1.15}$$

such that

$$\frac{\ddot{R}}{R} = -\frac{8\pi G}{3}\rho < 0. \tag{8.1.16}$$

This shows that the expansion of the universe is decelerating. Going backwards in time this implies the existence of a big bang. Using

$$\begin{aligned} &\frac{d}{dt}(\rho R^3) = -p\frac{d}{dt}R^3 = -\frac{1}{3}\rho 3R^2\dot{R} = -\rho R^2\dot{R} \Rightarrow \\ &\dot{\rho}R^3 + 3\rho R^2\dot{R} + \rho R^2\dot{R} = \dot{\rho}R^3 + 4\rho R^2\dot{R} = 0, \end{aligned} \tag{8.1.17}$$

one finds

$$\rho R^4 = N, \tag{8.1.18}$$

where N is a constant. We then obtain

$$\frac{\ddot{R}}{R} = -\frac{8\pi GN}{3R^4} \Rightarrow \frac{1}{2}\dot{R}^2 = \frac{4\pi GN}{3R^2} + C. \tag{8.1.19}$$

The constant C follows from eq.(8.1.16) to be

$$C = -\frac{k}{2}, \tag{8.1.20}$$

such that

$$\dot{R} = \sqrt{\frac{8\pi GN}{3R^2} - k}. \tag{8.1.21}$$

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

$$t = \int_0^R dRR\sqrt{\frac{3}{8\pi GN}} = \frac{1}{2}R^2\sqrt{\frac{3}{8\pi GN}}, \tag{8.1.22}$$

such that

$$\frac{32\pi GN}{3R^4}t^2 = \frac{32\pi G\rho}{3}t^2 = 1, \tag{8.1.23}$$

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

$$\rho = \frac{\pi^2}{15}T^4, \tag{8.1.24}$$

and for fermions one has

$$\rho = \frac{7\pi^2}{120}T^4. \tag{8.1.25}$$

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

$$m_P = G^{-1/2} = 1.2211 \times 10^{19}\text{GeV}, \tag{8.1.26}$$

or the Planck time

$$t_p = G^{1/2} = 5.3904 \times 10^{-44}\text{sec}. \tag{8.1.27}$$

This allows us to write

$$\frac{32\pi\rho/m_P^4}{3}(t/t_P)^2 = 1. \tag{8.1.28}$$

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.

8.2 The QCD Phase Transition

As we have seen in the previous chapter quantum systems (in particular quantum field theories) at finite temperature characterized by a partition function

$$Z = \text{Tr}\exp(-\beta H), \tag{8.2.1}$$

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 $O(4) \to O(3)$ 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 $SU(3)_L \otimes SU(3)_R \to SU(3)_{L=R}$ 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.

8.3 The High Temperature Phase Transition in the Purge Glue Theory

Let us consider an $SU(3)$ purge gauge theory (without quarks) at finite temperature $T = 1/\beta$, i.e. with periodic boundary conditions in the euclidean time direction

$$G_\mu(\vec{x}, t + \beta) = G_\mu(\vec{x}, t). \tag{8.3.1}$$

Then we can construct the Polyakov loop

$$\Phi(\vec{x}) = \mathrm{Tr}(\mathcal{P} \exp \int_0^\beta dt G_4(\vec{x}, t)) \in C, \tag{8.3.2}$$

which is a complex scalar field in three dimensions. Like the Wilson loop measures the energy of a static quark-antiquark pair, the Polyakov loop measures the free energy F of a single static quark, i.e.

$$\langle \Phi(\vec{x}) \rangle \propto \exp(-\beta F). \tag{8.3.3}$$

In the confined phase, i.e. at low temperatures, quarks are confined and hence a single quark has infinite energy $F = \infty$. Consequently, $\langle \Phi(\vec{x}) \rangle = 0$ in the confined phase. In the deconfined phase, on the other hand, one expects quarks to deconfine such that then $\langle \Phi(\vec{x}) \rangle \neq 0$. This means that the Polyakov loop is an order parameter for the deconfinement phase transition

in the pure gluon theory (with external static quarks). The Polyakov loop is invariant under ordinary periodic gauge transformations, but it transforms into

$$\Phi'(\vec{x}) = \Phi(\vec{x})z^*, \tag{8.3.4}$$

under transformations

$$g(\vec{x}, t+\beta) = g(\vec{x}, t)z, \tag{8.3.5}$$

which are periodic only up to a center element $z \in \mathbf{Z}(3)$. Hence, $\langle\Phi(\vec{x})\rangle \neq 0$ signals the spontaneous breakdown of the $\mathbf{Z}(3)$ center symmetry. It is interesting that the symmetry is then broken spontaneously at high temperatures, in contrast to typical condensed matter examples. The spontaneous breakdown of $\mathbf{Z}(3)$ has been observed in lattice simulations of the pure $SU(3)$ gauge theory, and it has been clearly established that the deconfinement phase transition is first order. For two colors (i.e. for $SU(2)$ gauge theory) again the $\mathbf{Z}(2)$ center symmetry gets spontaneously broken, but then the phase transition turns out to be second order.

8.4 Domain Walls and Gluonic Interfaces

When a discrete symmetry is spontaneously broken different phases exist that are distinguished by some order parameter. Different regions of space may be in different phases and the order parameter changes as one goes from one region to another. Since the change in the order parameter costs free energy, the different regions are separated by domain walls with a free energy proportional to their surface area. Domain walls arise, for example, in the Ising model, where a $\mathbf{Z}(2)$ symmetry gets spontaneously broken. In the Standard model of particle physics no discrete symmetry is spontaneously broken and domain walls do not arise. However, when one neglects quarks — as it is done in the quenched approximation to QCD — the $\mathbf{Z}(3)$ center symmetry of the $SU(3)$ gauge group plays a crucial role. At high temperatures it is spontaneously broken and there are in fact three distinct deconfined phases that are characterized by the values of the Polyakov loop. When different regions of space are in different phases they are separated by deconfined-deconfined domain walls. When the temperature decreases the gluonic system undergoes a phase transition to the confined glueball phase.

The phase transition is of first order, such that the confined and the three deconfined phases coexist at the critical temperature. Hence, at T_c there exist in addition confined-deconfined interfaces.

As we will see, at T_c the deconfined-deconfined interface tension is two times the confined-deconfined interface tension. This leads to the phenomenon of complete wetting: a deconfined-deconfined domain wall splits into two confined-deconfined interfaces with a layer of confined phase in between. Complete wetting is a critical phenomenon of interfaces, which arises although the bulk phase transition is first order. The critical behavior is characterized by the divergence of the thickness of the confined complete wetting layer, which is governed by a certain critical exponent. We will derive the values of the critical exponents of complete wetting from an effective 3-dimensional $\mathbf{Z}(3)$ symmetric Φ^4 model for the Polyakov loop, which is in the same universality class as the pure $SU(3)$ gauge theory. The shape of the interfaces (and also the values of the critical exponents) follow from the classical equations of motion of the effective theory. Complete wetting is a peculiarity of the pure gauge theory. When quarks are included they break the $\mathbf{Z}(3)$ symmetry explicitly, and two of the three deconfined phases become unstable (or at least metastable). Then stable deconfined-deconfined interfaces no longer exist at the phase transition and wetting does not arise. In fact, a region of space that was originally in a metastable $\mathbf{Z}(3)$ phase of the early universe converts into the stable phase about 10^{-14} seconds after the big bang, while the deconfinement phase transition occurred at about 10^{-5} seconds after the bang.

In full QCD the dynamics of a first order phase transition is governed by bubble nucleation: the early high temperature phase supercools until the latent heat is large enough to provide the necessary surface free energy to nucleate bubbles of the hadronic phase. The bubbles expand until the whole universe is converted into the low temperature phase. An important parameter for the dynamics of the phase transition is the confined-deconfined interface tension. It sets the scale for the spatial size of a nucleating bubble, and hence it determines the size of spatial inhomogeneities produced during the phase transition. The inhomogeneities in turn may influence the primordial nucleosynthesis of light elements that takes place after the phase transition, because they induce fluctuations in the baryon density . In the pure gauge theory the dynamics of the phase transition is very different due to complete wetting. Then, above the transition temperature, a network

of deconfined-deconfined domain walls is present. As one approaches the phase transition the domain walls split into pairs of confined-deconfined interfaces with a complete wetting layer of confined phase in between. In this scenario the creation of the confined phase does not require supercooling, because the free energy of two confined-deconfined interfaces is the same as for one deconfined-deconfined domain wall. In fact, the confined phase appears already slightly above T_c. Furthermore, the spatial structure of the confined phase is determined by the domain wall network that emerged from the dynamics of the gluon plasma above T_c, and not by the value of the interface tension. In addition, also the usual bubble nucleation may occur, but it is certainly not the only mechanism by which the phases transform into each other when complete wetting occurs.

Since the confined-deconfined interface tension is an important parameter of the deconfinement phase transition one would like to compute its value. Unlike the critical exponents of complete wetting the value of the interface tension is not universal. It does not follow from the effective Φ^4 model, and hence it must be calculated from the full $SU(3)$ gauge theory. This is possible in numerical lattice simulations of the theory. The interfaces cause characteristic effects in the spectrum of the transfer matrix, which can be used to extract the value of the interface tension.

8.5 Complete versus Incomplete Wetting

Above the deconfinement phase transition of the $SU(3)$ pure gauge theory a $\mathbf{Z}(3)$ symmetry is spontaneously broken, and three deconfined phases coexist with each other. Spatial regions of different phases are separated by domain walls with a free energy

$$F = \sigma_{dd}(T) A T, \tag{8.5.1}$$

where $\sigma_{dd}(T)$ is the so-called reduced deconfined-deconfined interface tension, T is the temperature and A is the interface area. At very high temperatures the interface tension has been computed perturbatively

$$\sigma_{dd}(T_c) = \frac{8\pi^2 T^2}{9g}, \tag{8.5.2}$$

where g is the gauge coupling renormalized at the scale T. At the critical temperature T_c of the deconfinement phase transition in addition the confined phase arises, which is separated from the deconfined phases by confined-deconfined interfaces with a reduced interface tension σ_{cd}. Thermodynamical stability requires

$$\sigma_{dd}(T_c) \leq 2\sigma_{cd}. \tag{8.5.3}$$

For $\sigma_{dd}(T_c) > 2\sigma_{cd}$ a deconfined-deconfined domain wall could lower its free energy by splitting into two confined-deconfined interfaces, which would lead to

$$\sigma_{dd}(T_c) = 2\sigma_{cd}. \tag{8.5.4}$$

In this case one speaks of complete wetting, because a deconfined-deconfined interface is then always completely wet by the confined layer between the two confined-deconfined interfaces. In general, equilibrium of forces at a vertex where two deconfined phases are in contact with the confined phase implies

$$\sigma_{dd}(T_c) = 2\sigma_{cd} \cos\frac{\theta}{2}, \tag{8.5.5}$$

where θ is the opening angle of a lens of confined phase that forms at the deconfined-deconfined interface. The geometrical situation is illustrated in fig.8.1. When $\sigma_{dd}(T_c) < 2\sigma_{cd}$ the angle θ is different from zero (fig.8.1a). Then one speaks of incomplete or partial wetting, because the deconfined-deconfined interface is only partially wet by the confined phase. For complete wetting ($\sigma_{dd}(T_c) = 2\sigma_{cd}$) the angle θ vanishes and the lens degenerates to a confined layer that completely wets the deconfined-deconfined domain wall (fig.8.1b). Complete wetting is a critical phenomenon of interfaces that arises although the phase transition in the bulk is of first order and not critical. In particular, the thickness z_0 of the complete wetting layer grows to macroscopic size as one approaches the phase transition from above

$$z_0 \propto (T - T_c)^{-\psi}. \tag{8.5.6}$$

The divergence of z_0 is characterized by the critical exponent ψ. At the same time the order parameter (in our case the expectation value of the Polyakov loop) at the center of the interface

$$\Phi_1(0) \propto (T - T_c)^{\beta} \tag{8.5.7}$$

Figure 8.1: The shape of a confined region that wets a deconfined-deconfined domain wall both for (a) incomplete and for (b) complete wetting.

goes to zero with another critical exponent β. In the next section we will compute the critical exponents analytically and we will see that the values are

$$\psi = 0,\ \beta = \frac{1}{2}. \tag{8.5.8}$$

In particular, the thickness of the confined complete wetting layer z_0 diverges only logarithmically as one approaches the phase transition. The critical exponents will be computed using an effective 3-dimensional $\mathbf{Z}(3)$ symmetric Φ^4 theory for the Polyakov loop, which is in the same universality class (for critical interface effects) as the pure gauge theory.

8.6 An Effective 3-d $\mathbf{Z}(3)$ Symmetric Φ^4 Model

Let us again consider the Polyakov loop

$$\Phi(\vec{x}) = \mathrm{Tr}(\mathcal{P} \exp \int_0^\beta dt G_4(\vec{x}, t)) \in C. \tag{8.6.1}$$

It is invariant under ordinary periodic gauge transformations, but it transforms into

$$\Phi'(\vec{x}) = \Phi(\vec{x})z^*, \tag{8.6.2}$$

under transformations which are periodic only up to a center element $z \in \mathbf{Z}(3)$. Another important symmetry of the system is charge conjugation, which transforms the Polyakov loop into its complex conjugate

$$\Phi'(\vec{x}) = \Phi^*(\vec{x}). \tag{8.6.3}$$

The Polyakov loop is a 3-dimensional complex scalar field, whose dynamics is governed by the effective action

$$\exp(-S_{eff}[\Phi]) = \int \mathcal{D}G_\mu \; \delta[\Phi(\vec{x}) - \mathrm{Tr}(\mathcal{P}\exp\int_0^\beta dt G_4(\vec{x},t))] \exp(-S[G_\mu]). \tag{8.6.4}$$

The δ-functional ensures that Φ obeys eq.(8.3.2). The full effective action is certainly an unpleasant nonlocal functional, which is not well suited for further analytic investigations. When one is interested in universal critical phenomena — in this case in complete wetting — one may, however, take any action from the same universality class as $S_{eff}[\Phi]$. A pleasant local action is

$$S[\Phi] = \int d^3x [\frac{1}{2}\partial_i\Phi^*\partial_i\Phi + V(\Phi)]. \tag{8.6.5}$$

In order to belong to the same universality class as the full effective action the action S should at least have the same symmetries, i.e. it should be invariant under charge conjugation and against $\mathbf{Z}(3)$ transformations. The kinetic term is invariant under both. The potential term, however, must obey

$$V(\Phi^*) = V(\Phi), \;\; V(\Phi z^*) = V(\Phi), \tag{8.6.6}$$

where $z \in \mathbf{Z}(3)$. For simplicity let us assume that $V(\Phi)$ is a polynomial in Φ of order at most Φ^4. Decomposing the Polyakov loop into real and imaginary parts

$$\Phi = \Phi_1 + i\Phi_2, \tag{8.6.7}$$

and using the symmetry requirements of eq.(8.6.6) one obtains

$$V(\Phi) = a|\Phi|^2 + b\Phi_1(\Phi_1^2 - 3\Phi_2^2) + c|\Phi|^4. \tag{8.6.8}$$

Stability of the problem requires $c > 0$. For $a > 0$ the potential has a minimum at $\Phi_1 = \Phi_2 = 0$, which corresponds to the confined phase and which

Figure 8.2: The equipotential lines of $V(\Phi)$ in the complex Φ plane at the critical temperature. The parameters are $a = 0.25$, $b = -1$, $c = 1$.

we denote by $\Phi^{(4)} = (0, 0)$. In addition, there may be minima corresponding to the three deconfined phases. They occur for $0 < a < 9b^2/32c$, $b < 0$ and they are located at

$$\begin{aligned}
\Phi^{(1)} &= (\Phi_0, 0), \\
\Phi^{(2)} &= (-\frac{1}{2}\Phi_0, \frac{\sqrt{3}}{2}\Phi_0), \\
\Phi^{(3)} &= (-\frac{1}{2}\Phi_0, -\frac{\sqrt{3}}{2}\Phi_0),
\end{aligned} \tag{8.6.9}$$

where

$$\Phi_0 = \frac{-3b + \sqrt{9b^2 - 32ac}}{8c}. \tag{8.6.10}$$

The phase transition occurs when all four minima are degenerate, i.e. when

$$V(\Phi^{(1)}) = V(\Phi^{(2)}) = V(\Phi^{(3)}) = V(\Phi^{(4)}). \tag{8.6.11}$$

In our parametrization this corresponds to $b^2 = 4ac$. The equipotential lines of $V(\Phi)$ for this situation are shown in fig.8.2.

8.7 Interfaces and Critical Exponents

Let us consider the theory in a spatial volume with periodic boundary conditions in the x- and y-directions with lengths L_x and L_y, and with an infinite z-direction. Then between different bulk phases interfaces arise, which are closed due to periodic boundary conditions. A planar interface perpendicular to the z-direction has the area $A = L_x L_y$ and it is described by an order parameter profile

$$\Phi(\vec{x}) = \Phi(z). \tag{8.7.1}$$

The form of $\Phi(z)$ follows from the classical equations of motion corresponding to the action $S[\Phi]$, which are given by

$$\begin{aligned} \Phi_1'' &= \frac{\partial V}{\partial \Phi_1} = \Phi_1(2a + 4c|\Phi|^2) + 3b(\Phi_1^2 - \Phi_2^2), \\ \Phi_2'' &= \frac{\partial V}{\partial \Phi_2} = \Phi_2(2a + 4c|\Phi|^2) - 6b\Phi_1\Phi_2. \end{aligned} \tag{8.7.2}$$

Here a prime denotes a derivative with respect to z. The interface action is related to the reduced interface tension by

$$S = L_x L_y \int dz(T + V) = 2A \int dz V = \sigma A. \tag{8.7.3}$$

Note that interfaces are topological excitations, like kinks or instantons, that interpolate between two different vacuum states (phases) of the model.

The simplest interface is the one between the confined and the deconfined phase, which exists only at the phase transition temperature corresponding to $b^2 = 4ac$. The interface interpolates between the confined phase $\Phi^{(4)}$ at large negative z and the deconfined phase $\Phi^{(1)}$ at large positive z. At the critical point the solution of the classical equations of motion gives

$$\Phi_1(z) = \frac{\Phi_0}{2}[1 + \tanh\frac{\mu}{2}(z - z_0)],\ \Phi_2(z) = 0, \tag{8.7.4}$$

where $\mu = -b/\sqrt{2c}$ and $b < 0$. The corresponding action yields a reduced interface tension

$$\sigma_{cd} = S/A = \frac{1}{6}\mu\Phi_0^2. \tag{8.7.5}$$

Now we increase the temperature slightly above the phase transition, such that we are in the deconfined phase. Then also deconfined-deconfined interfaces arise. Such an interface separating the phases $\Phi^{(2)}$ and $\Phi^{(3)}$ can be

constructed by combining two confined-deconfined interfaces, one centered around $z = z_0$ and the other one centered around $z = -z_0$

$$\Phi_1(z) = -\frac{1}{4}\Phi_0[2 + \tanh\frac{\mu}{2}(z - z_0) - \tanh\frac{\mu}{2}(z + z_0)],$$
$$\Phi_2(z) = \frac{\sqrt{3}}{4}\Phi_0[\tanh\frac{\mu}{2}(z - z_0) + \tanh\frac{\mu}{2}(z + z_0)]. \tag{8.7.6}$$

Minimizing the action with respect to the free parameter z_0 — the thickness of the confined layer between the two confined-deconfined interfaces — one obtains

$$z_0 = -\frac{1}{2\mu}\ln(\frac{1}{2} - \frac{2ac}{b^2}). \tag{8.7.7}$$

The critical temperature corresponds to $b^2 = 4ac$. Then z_0 diverges logarithmically. This shows that at the critical point complete wetting occurs. Because of the logarithmic divergence the critical exponent is

$$\psi = 0. \tag{8.7.8}$$

For the expectation value of the Polyakov loop at the center of the interface one finds

$$\Phi_1(0) = -\Phi_0\sqrt{\frac{1}{2} - \frac{2ac}{b^2}}, \tag{8.7.9}$$

which goes to zero as one approaches the phase transition. Because of the square root, the corresponding critical exponent is

$$\beta = \frac{1}{2}. \tag{8.7.10}$$

The same critical exponents occur for condensed matter systems with short range forces between different interfaces. For the gluon system one would also expect short range forces, because there are no massless excitations even in the plasma phase. The reduced deconfined-deconfined interface tension of the above solution is given by

$$\sigma_{dd}(T) = S/A = \mu\Phi_0^2[\frac{1}{3} - 2(\frac{1}{2} - \frac{2ac}{b^2})]. \tag{8.7.11}$$

Therefore

$$2\sigma_{cd} - \sigma_{dd}(T) = 2\mu\Phi_0^2(\frac{1}{2} - \frac{2ac}{b^2}) \propto (T - T_c)^{1-\psi}, \tag{8.7.12}$$

in agreement with general scaling arguments. The analytic instanton solution from above assumes a dilute gas of confined-deconfined interfaces and treats interactions between interfaces only to lowest order. In general, the equations of motion can only be solved numerically. For example, one may consider the deconfined-deconfined interface between the phases $\Phi^{(2)}$ and $\Phi^{(3)}$ for any temperature above T_c. A typical interface shape is shown in fig.8.3 deep in the deconfined phase (a), in a situation where wetting sets in (b) and very close to the phase transition (c). Fig. 8.3c indicates that the deconfined-deconfined interface splits into two deconfined-confined interfaces when one approaches the phase transition. This is again the phenomenon of complete wetting.

We have seen that wetting is always complete in the effective Φ^4 model. Of course, this does not mean that complete wetting also occurs in the pure $SU(3)$ gauge theory. Such a question can only be decided in the gauge theory itself, for example by performing numerical simulations. Indeed, numerical simulations show that the pure $SU(3)$ gauge theory shows complete wetting.

8.8 The Interface Tension and the Spectrum of the Transfer Matrix

When one wants to determine the value of the interface tension for the gluon system one cannot do this in the effective theory, because σ_{cd} is not a universal quantity. The determination of the confined-deconfined interface tension is a complicated nonperturbative problem, which one can attack by numerical lattice simulations. Then the question arises what strategy is best suited to extract the interface tension from numerical data. It turns out that the presence of interfaces influences the spectrum of the transfer matrix in volumes with a cylindrical geometry, i.e. with two short x- and y-directions and one much longer z-direction. Then several bulk phases, aligned along the z-direction, coexist with each other. They are separated by interfaces, which are closed via the short periodic x- and y-directions. The interfaces describe transitions from one phase to another, and may hence be interpreted as tunneling events between different vacuum states. The associated tunneling energy is exponentially suppressed by the interface

Figure 8.3: The shape of the interface between two different deconfined phases (for $b = -1$, $c = 1$ and $a = a_c - \Delta a = b^2/4c - \Delta a$) deep in the deconfined phase (a) (at $\Delta a = 10^{-1}$), in the region where wetting sets in (b) (at $\Delta a = 10^{-3}$) and very close to the phase transition (c) (at $\Delta a = 10^{-8}$). The solid line is the real part Φ_1 and the dashed line represents the imaginary part Φ_2.

free energy, i.e. by the interface tension times the interface area $A = L_x L_y$. The tunneling energy — as the splitting between the lowest states of the transfer matrix — can be computed in numerical simulations, and the interface tension can be extracted from its finite volume effects. This approach makes use of the necessarily present finite size effects, instead of trying to avoid them by going to large lattices. In fact, one uses the cylindrical spatial geometry to allow the different phases to coexist in the finite volume. Furthermore, the analytic understanding of the finite volume effects allows to extrapolate reliably to the infinite volume limit.

The transfer matrix $\mathcal{T}(z)$ describes the evolution of the gluonic system over a distance z in the spatial z-direction. Its elements are transition amplitudes, which can be represented as path integrals

$$\mathcal{T}_{ij}(z) = \int \mathcal{D}\Phi \exp(-S[\Phi]). \tag{8.8.1}$$

Here the integration is restricted to fields in the interval $[0, z]$ and the boundary conditions $\Phi(x, y, 0) = \Phi^{(i)}(x, y)$ and $\Phi(x, y, z) = \Phi^{(j)}(x, y)$ are understood, i.e. the initial and final field configurations are fixed. The transfer matrix is a Euclidean evolution operator that can be represented as

$$\mathcal{T}(z) = \exp(-Hz), \tag{8.8.2}$$

where H is a Hamilton operator. The spectrum of H shows characteristic finite size effects, which can be used to determine the value of the confined-deconfined interface tension.

First we consider a system with C-periodic (charge conjugate) boundary conditions, i.e.

$$\Phi(\vec{x} + L\vec{e}_x) = \Phi(\vec{x} + L\vec{e}_y) = \Phi^*(\vec{x}), \tag{8.8.3}$$

in the short x- and y-directions. Then the boundary conditions eliminate two of the three deconfined phases, and only the phase $\Phi^{(1)}$ coexists with the confined phase $\Phi^{(4)}$. We concentrate on the subspace of these two phases, and we can express the transfer matrix in this subspace as

$$t(z) = \begin{pmatrix} t_{dd}(z) & t_{cd}(z) \\ t_{cd}(z) & t_{cc}(z) \end{pmatrix}. \tag{8.8.4}$$

The quantities $t_{dd}(z)$, $t_{cd}(z)$ and $t_{cc}(z)$ are transition amplitudes between the two phases, which determine the eigenvalues of the transfer matrix

$$t_0(z) = \frac{1}{2}[t_{dd}(z) + t_{cc}(z) + \sqrt{(t_{dd}(z) - t_{cc}(z))^2 + 4t_{cd}(z)^2}],$$

$$t_1(z) = \frac{1}{2}[t_{dd}(z) + t_{cc}(z) - \sqrt{(t_{dd}(z) - t_{cc}(z))^2 + 4t_{cd}(z)^2}]. \qquad (8.8.5)$$

The transition amplitudes can be computed using a dilute interface approximation. One integrates over configurations with several bulk phases, aligned along the z-direction, which are separated by confined-deconfined interfaces spanned in the short x- and y-directions. The interface area is $A = L_x L_y$. Because in practice the phase transition temperature is not known with high enough precision, we must work in a temperature region around the critical point, and we must take into account the small free energy difference between the confined and the deconfined phase slightly off T_c. A layer of confined phase of thickness z_0 costs a free energy $F = f_c A z_0$, where f_c is the free energy density of the confined phase. The corresponding Boltzmann weight is

$$\exp(-F/T) = \exp(-f_c A z_0/T). \qquad (8.8.6)$$

A layer of deconfined phase gets a factor $\exp(-F/T) = \exp(-f_d A z_0/T)$. Introducing the variable

$$x = \frac{1}{2}(f_c - f_d)A/T, \qquad (8.8.7)$$

which measures the free energy difference between the confined and the deconfined phase, one may write a factor $\exp(-xz_0)$ for the confined phase and a factor $\exp(xz_0)$ for the deconfined phase. Of course, also the interfaces that separate layers of bulk phases cost free energy and are suppressed by the corresponding Boltzmann factor. The free energy of a confined-deconfined interface is $F = \sigma_{cd} A T$, and the Boltzmann factor is

$$\exp(-F/T) = \delta \exp(-\sigma_{cd} A). \qquad (8.8.8)$$

The pre-exponential factor δ is due to so-called capillary wave fluctuations of the interface. In three dimensions this factor is (to leading order) independent of the interface area.

The leading contributions to the amplitude $t_{dd}(z)$, which describes the transition from the deconfined back to the deconfined phase, are given by

$$t_{dd}(z) = \exp(xz) + \int_0^z dz_0 \int_{z_0}^z dz_0' \\ \exp(xz_0)\delta \exp(-\sigma_{cd}A) \exp(-x(z_0' - z_0))\delta \exp(-\sigma_{cd}A) \exp(x(z - z_0')). \qquad (8.8.9)$$

The first term represents a layer of deconfined phase of thickness z that fills the whole volume. The next term is the two-interface contribution. The integration is over the interface locations z_0 and z_0'. Note that the extension of an interface is neglected in the dilute interface approximation. The first factor in the integrand represents a deconfined layer of thickness z_0. The next factor describes the interface at z_0 followed by a confined layer of thickness $z_0' - z_0$. The last factors represent the interface at z_0' and a deconfined layer of thickness $z - z_0'$. Similarly, for the transition amplitude between the confined and the deconfined phase one obtains

$$t_{cd}(z) = \int_0^z dz_0 \exp(xz_0)\delta \exp(-\sigma_{cd}A) \exp(-x(z - z_0)). \tag{8.8.10}$$

This is the one-interface contribution. In complete analogy one finds the expression for $t_{cc}(z)$. It is straightforward, although somewhat tedious, to work out the multi-interface contributions. In the final result all contributions are summed up, and one finds

$$E_1 - E_0 = 2\sqrt{x^2 + \delta^2 \exp(-2\sigma_{cd}A)}, \tag{8.8.11}$$

where $t_{0,1}(z) = \exp(-E_{0,1}z)$. The energy splitting assumes its minimum at the finite volume critical point at which the free energies of the two phases are equal. Then $x = 0$ and

$$E_1 - E_0 = 2\delta \exp(-\sigma_{cd}A) \quad (\text{for } x = 0). \tag{8.8.12}$$

The energy splitting vanishes exponentially with the interface area times the reduced confined-deconfined interface tension σ_{cd}. This behavior has been used in numerical simulations to determine the value of σ_{cd}.

Fig.8.4 shows a typical configuration of a Monte-Carlo ensemble for an $8 \times 8 \times 128 \times 2$ lattice with C-periodic boundary conditions. One can see that the confined phase coexists with one deconfined phase, and that the two phases are separated by confined-deconfined interfaces. Fig.8.5 shows the energy E_1 on a $4 \times 6 \times 64 \times 2$ lattice with C-periodic boundary conditions in the x- and y-directions. The minimum of the energy splitting indeed occurs at the critical coupling. From the finite size effect of the value at the minimum one determines the reduced confined-deconfined interface tension using eq.(8.8.12). A fit to the numerical data is shown in fig.8.6. It gives

$$\sigma_{cd} = 0.035(1) = 0.140(4)T_c^2. \tag{8.8.13}$$

Figure 8.4: *A typical configuration on the* $8 \times 8 \times 128 \times 2$ *lattice with* C*-periodic boundary conditions.*

Figure 8.5: *The energy* E_1 *on a* $4 \times 6 \times 96 \times 2$ *lattice with* C*-periodic boundary conditions.*

Figure 8.6: *A fit to the finite size dependence of the energy splittings.*

This value of the interface tension is rather small. In full QCD the interface tension is most likely even smaller than in the pure gauge theory (it may even be zero if the phase transition is second order). This means that the spatial fluctuations produced at the QCD phase transition in the early universe were most likely rather small. Then the primordial nucleosynthesis has most likely taken place under spatially homogeneous conditions.