The Standard Model of Particle Physics

Andrea Romanino – *Associate Professor, SISSA. The scientific interests include Standard Model and beyond, including supersymmetry, extra dimensions and grand unification; flavour and neutrino physics.*

Abstract: These lectures provide a basic introduction to the Standard Model (SM) of particle physics. While there are several reasons to believe that the Standard Model is just the low energy limit of a more fundamental theory, the SM has been successfully tested at an impressive level of accuracy and provides at present our best fundamental understanding of the phenomenology of particle physics. The perspective I will take will not be historical, I will instead take advantage of our present understanding to find the most direct logical motivations.

The Standard Model of Particle Physics

Andrea Romanino
SISSA/ISAS and INFN, I–34151 Trieste, Italy

1 Introduction

These lectures provide a basic introduction to the Standard Model (SM) of particle physics. While there are several reasons to believe that the Standard Model is just the low energy limit of a more fundamental theory, the SM has been successfully tested at an impressive level of accuracy and provides at present our best fundamental understanding of the phenomenology of particle physics. The perspective I will take will not be historical, I will instead take advantage of our present understanding to find the most direct logical motivations. As the level of the audience is quite diverse, I will summarize (in a concise, qualitative, and pragmatic way) the main theoretical preliminaries needed to make sense of what will follow. The hope is that a part of these lectures could be useful even to undergraduate students, that people who are already familiar with QFT and gauge theories can also benefit from them, and people in between can have a first impact with the tools involved in a non qualitative treatment. I will systematically use natural units, in which $c = \hbar = 1$.

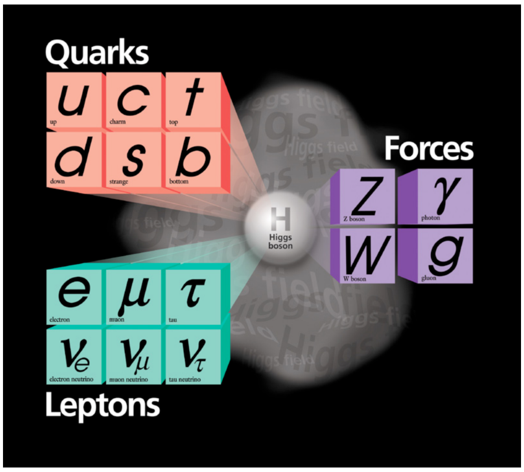


Figure 1: The SM particle content

Let us start with the very basic facts. The main ingredients of the SM are shown in Fig. 1. The particles involved are characterized by their spin, their mass, and the quantum numbers (charges) determining their interactions. The fermion content (spin $= 1/2$) is organized in three families with identical quantum numbers and different masses. The heavier families are unstable and decay into the lightest one, which makes up most of the ordinary matter (us). The four fermions in each family are distinguished by their charges under strong and electromagnetic interactions. Two of them are quarks, which are charged under the strong interactions, and two are leptons, which are not. The two quarks have electromagnetic charges $2/3$ ("up" quarks) and -1/3 ("down" quarks) respectively, and the two leptons have charges -1 (charged, or "down" leptons) and 0 (neutrinos, or "up" leptons), in units in which the electron charge is -1. The neutrinos are peculiar from two, probably theoretically related, points of view: they are neutral under both the strong and the electromagnetic interactions (they feel weak interactions, though) and they are at least six order of magnitudes lighter than all the other SM fermions. The masses of the SM fermions span a range going from the sub-eV neutrino masses to the $1.7 \cdot 10^2$ GeV top mass [1]. They exhibit quite a peculiar structure, with the masses of the different families being hierarchically separated. Each fermion is associated to two so-called chiralities. Chirality is conserved for massless fermions, in which case the chirality coincides with the helicity. That is why the two possible chiralities are called left-handed and right-handed. From a theoretical point of view, chirality, by definition, distinguishes the two irreducible representations of the Lorentz group that can be used to describe spin $1/2$ fermions. Massive charged fermions are necessarily described by two components of different chiralities combined in what is called a Dirac spinor. As for neutrinos, only the left-handed chirality has been observed so far. This can be elegantly understood in terms of the quantum numbers of a possible right-handed component, but the argument goes beyond the scope of these lectures.

The SM interactions are associated to the exchange of four vector bosons (spin = 1). The photon mediates electromagnetic interactions, the gluon strong interactions, the Z and W weak interactions. The photon and the gluons are massless, while the Z and the W are massive, which is the reason why weak interactions are weak at low energy (they are suppressed by powers of $E/M_{Z,W}$, where E is the energy of the process). Despite their weakness, they give rise to distinctive signatures because they violate parity P, charge conjugation C, their combination CP, time-reversal T, and family number, which all are symmetries of the electromagnetic and strong[1] interactions. In particular, the decay of heavier into lighter families is due to weak interactions.

The description outlined above holds at relatively low energy and has long been known before the SM was invented. The SM description becomes necessary when processes involving higher energies are considered. The transition from the low energy "effective" regime and the SM regime takes place around the electroweak scale $v \approx 174\,\text{GeV}$. Above this scale, nature exhibits a higher degree of symmetry. The electromagnetic and weak interactions become indistinguishable and are unified in the "electroweak" interaction. The left-handed chirality components of up and down fermions also become indistinguishable and are unified in electroweak doublets. The electroweak scale is where such an "electroweak" symmetry breaks. Together with the QCD and the Planck scale, it is one of the fundamental scales of nature known at present. The mechanism through which the electroweak symmetry breaks is well established, also experimentally, and is called spontaneous symmetry breaking. It is through such a mechanism that the fermions and the massive gauge bosons acquire

[1]In the limit in which non-perturbative effects associated to the vacuum structure of QCD are neglected.

a mass proportional to the electroweak scale[2]. It is still not know, however, what is the mechanism triggering the spontaneous breaking. The SM encodes the simplest (both from the theoretical and phenomenological consistency point of view) option: the Higgs mechanism [2]. Such a mechanism postulates the existence of a spin $= 0$ field, the Higgs field. Unlike all other fields, that require energy to be switched on, the Higgs field is "on" even in the ground state, where it permeates space-time. It is through their interactions with the Higgs field that the SM massive particles acquire their masses, proportional to the coupling to the Higgs. The identification of the mechanism responsible for the electroweak symmetry breaking and the stability of the weak scale compared to the Planck scale are two central issues in today's particle physics and are two of the most important missions of the LHC.

In order to go beyond the qualitative picture presented so far, I need to introduce some important theoretical tools, at least at a very basic level. The key tools are gauge theories and their spontaneous breaking. In Section 2 I will introduce gauge symmetries, which allow to define the SM gauge sector in Section 3. Spontaneous breaking will be introduced in Section 4, which allows to introduce the SM Higgs and Yukawa sectors in Section 5. In Section 6 we will discuss the phenomenological implications of the SM and mention the open problems.

2 Gauge (and global) transformations

The SM is first of all a quantum field theory (QFT). In QFT, particles are associated to fields $\phi_i(x)$, $i = 1 \dots n$ depending on the space-time coordinates $x = (x^0, x^1, x^2, x^3)$. We consider only fields with spin $s = 0, 1/2, 1$ (no gravity), the only ones needed for the SM, and the only ones for which we know how to write a theoretically consistent QFT. Their dynamics is determined by an action S written in terms of a Lagrangian density $\mathcal{L}(x)$ (which I will simply call "Lagrangian") with dimension 4 in energy. The SM Lagrangian is Lorentz-invariant and local, and it can be written as a sum of monomials in the values of the fields and their derivatives (up to two) in a given space-time point:

$$S = \int d^4x \, \mathcal{L}(x), \quad \mathcal{L}(x) = \sum_k c_k \mathcal{O}_k(x). \tag{1}$$

In particular, the Lagrangian will contain: a term constant in the fields, which is not physical as long as gravity is not taken into account; terms linear in the fields, which can be reabsorbed by a shift redefinition of the fields $\phi(x) \to \phi(x) + c$; terms bilinear in the fields, the "free" Lagrangian $\mathcal{L}_{\text{free}}$, which account for the free propagation of the fields and define their dimensions in energy (1 for the bosons and 3/2 for the fermions); terms with at least three fields, the "interaction" Lagrangian $\mathcal{L}_{\text{int}}$, which

[2]Except possibly the neutrinos, whose masses are likely to be quadratic in the electroweak scale.

account for field interactions. In a perturbative regime, the amplitude of any physical process can be expressed as an expansion in $\mathcal{L}_{\text{int}}$ represented (at a given order) by a set of Feynman diagrams. The simplest example of interacting QFT involves a single real, scalar field $\varphi(x)$ with Lagrangian $\mathcal{L} = (\partial\phi)^2/2 - m^2\phi^2/2 - \lambda\phi^4/4$.

The dimension of the coefficients c_k in eq. (1) is important, as it determines the properties of the corresponding interaction. Terms with negative dimension, $c = 1/\Lambda^D$, $D > 0$, where Λ is an energy scale, decouple at sufficiently low energy E, where their role is suppressed by $(E/\Lambda)^D$. They are called "non-renormalizable". Only "renormalizable" terms with non-negative dimension of the coefficients are therefore relevant at low enough energy. The presence of non-renormalizable terms becomes important at high energy, where they end up making the theory incalculable. As the amplitudes grow with $(E/\Lambda)^D$, in fact, higher order in perturbative expansions will eventually become as important as the lower orders: the theory becomes non perturbative. This is signaled by the fact that the unitarity of the theory appears to be violated at the perturbative level. Therefore, non-renormalizable terms can be tolerated only in the context of an effective theory valid up to a certain energy scale ("cutoff") Λ. At the scale Λ, such terms should be accounted for by renormalizable interactions, if the theory is to remain perturbative and calculable. Such theoretical considerations played an important role in the development of the SM. The latter, as we will see, was born from the need to express the non-renormalizable four-fermion Fermi interaction accounting for weak interactions at low energy in terms of a renormalizable theory.

Let us now come to the role of symmetries. We will be dealing with SU(N) and U(1) symmetry groups only. The generators t of the group parametrize infinitesimal transformations $u \approx \mathbf{1} - i\epsilon t$, form a Lee algebra, $[t_1, t_2] = it_3$, $[t_1, [t_2, t_3]] + [t_2, [t_3, t_1]] + [t_3, [t_1, t_2]] = 0$, and in the case of SU($N$) are Hermitian traceless matrices. A standard choice of SU(N) generators is $t_a = \sigma_a/2$, $a = 1, 2, 3$, for the case $N = 2$, where σ_a are the Pauli matrices; $t_A = \lambda_A/2$, $A = 1 \ldots 8$, for the case $N = 3$, where λ_A are the Gell-Mann matrices; t_i, $i = i \ldots N^2 - 1$, with $\text{tr}(t_i t_j) = \delta_{ij}/2$, in the general case. The structure constants f_{ijk} are defined by $[t_i, t_j] = i f_{ijk} t_k$ and are antisymmetric.

In quantum mechanics a continuous symmetry group is represented by unitary transformations on the states of the system that commute with the Hamiltonian. The generators correspond to conserved quantities. In QFT, we consider symmetries whose action on the states corresponds to a transformation of the fields $\phi_i(x) \to U_{ij}\phi_j(x)$ that is a symmetry of the Lagrangian. If U does not depend on the space-time point x, we call the symmetry "global", or "rigid". The generators of the symmetry, or, more precisely, their actions T on the quantum states, are given by Noether's theorem in terms of conserved currents,

$$T = \int d\boldsymbol{x}\, j^0, \quad j^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi, \quad \partial_\mu j^\mu = 0, \tag{2}$$

where $\delta\phi$ is the infinitesimal change of ϕ associated to the generator t. A global symmetry is a true symmetry, meaning that it relates physically inequivalent states (multiplets) with same mass and spin and their couplings. The classic example of global symmetry in QFT is isospin, which acts on the doublet $Q = (u, d)^T$ made of the up and down quark fields. In the limit $m_u = m_d = 0$ and neglecting electromagnetic interactions, the SU(2) isospin transformation $Q \to UQ$ is a symmetry of the QCD Lagrangian for u, d. The isospin currents are $j^\mu_a = \overline{Q}\gamma^\mu(\sigma_a/2)Q$ (one for each generator). The isospin symmetry relates masses and couplings of inequivalent states. The experimental proof that isospin is indeed an (approximate) symmetry comes from the fact that the light hadrons do organize themselves into isospin multiplets. For example, the proton and the neutron make up a isospin 1/2 doublet with $m \approx 940\,\text{MeV}$, the pions an isospin 1 triplet with $m \approx 140\,\text{MeV}$. The effective pion-nucleon couplings are also related by the isospin symmetry.

Gauge transformations are not symmetries in the sense in which global symmetries are. Their action on the fields does depend on the space-time point and from the physical point of view they do not relate inequivalent physical states. They relate on the contrary equivalent field configurations, whose redundance should be taken into account and factored out.

In order to have a feeling of how gauge symmetries come about, let us consider the prototypical gauge theory, quantum electrodynamics (QED). The gauge symmetry is first of all a property of the classical theory. We know that the physical degrees of freedom of electrodynamics are the electromagnetic field $F_{\mu\nu}$ and the current j^μ, in terms of which Maxwell equations are formulated. On the other hand, $F_{\mu\nu}$ is not a good dynamical variable, in particular it does not allow the Maxwell equations to be obtained from a variational principle. On the other hand, it is possible to express $F_{\mu\nu}$ in terms of the vector potential A_μ as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The equations of motion can then be derived from the Lagrangian $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - j^\mu A_\mu$. On the other hand, A_μ is not a physical variable, as two fields related by the gauge transformation $A(x)_\mu \to A_\mu(x) + \partial_\mu\alpha(x)$ correspond to the same physical observable $F_{\mu\nu}$. Therefore, the gauge transformation just defines the equivalence of configurations corresponding to the same physical observable, it signals a redundancy in the degrees of freedom we are using.

Let us now consider the quantum theory and consider the case in which the electromagnetic current is associated to a single Dirac fermion ψ with charge Qe, $j^\mu = eQ\overline{\psi}\gamma^\mu\psi$, where e is the absolute value of the charge of the electron. The Lagrangian will accordingly be given by

$$\mathcal{L}_{\text{QED}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - j^\mu A_\mu + \overline{\psi}(i\gamma^\mu\partial_\mu - m)\psi, \tag{3}$$

where the free Lagrangian of the Dirac spinor has also been added. We can now check explicitly that: 1) the electromagnetic current j^μ turns out to be the Noether current associated to the U(1) global symmetry under which $\psi(x) \to e^{-ieQ\alpha}\psi(x)$, which therefore can be thought to account for the conservation of the electric charge; and 2) the U(1) symmetry turns out to be also related to gauge transformations. Let us remember in fact that we want the Lagrangian to be invariant with respect to gauge transformations of the vector field, or we would be able to distinguish equivalent configurations. It turns out that the Lagrangian is invariant provided that ψ transforms according to the local version of the global U(1) transformation above:

$$A(x)_\mu \to A_\mu(x) + \partial_\mu \alpha(x), \quad \psi(x) \to e^{-ieQ\alpha(x)}\psi(x). \tag{4}$$

Gauge invariance can therefore be considered as the principle underlying charge conservation. In order to verify the invariance of the Lagrangian, is is convenient to define the "covariant" derivative

$$D_\mu = \partial_\mu + ieQA_\mu, \quad \text{such that} \quad D_\mu(e^{-ieQ\alpha(x)}\psi(x)) = e^{-ieQ\alpha(x)}(D_\mu\psi(x)). \tag{5}$$

The QED Lagrangian can then be written as $\mathcal{L}_{\text{QED}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \overline{\psi}(i\gamma^\mu D_\mu - m)\psi$.

One can go one step further and argue that not only gauge invariance is the principle underlying charge conservation but that the QED Lagrangian itself follows from the principle of gauge invariance. The point is that the Lagrangian in eq. (3) can be obtained by forcing the spinor free Lagrangian $\overline{\psi}(i\gamma^\mu\partial_\mu - m)\psi$, invariant under global U(1) transformations $\psi(x) \to e^{-ieQ\alpha}\psi(x)$, to be invariant under local U(1) transformations $\psi(x) \to e^{-ieQ\alpha(x)}\psi(x)$. The argument goes as follows. In order to make the derivative term invariant, one is forced to introduce a vector field A_μ transforming as in eq. (4). A gauge invariant kinetic term for A_μ can then be built in terms of the gauge invariant quantity $F_{\mu\nu}$. Gauge invariance also explains why the photon is massless, as a mass term for A_μ would break gauge invariance.

There is also a second reason to believe that gauge invariance should be regarded as a fundamental principle, related to the fact that the quantization of the photon field is not straightforward. I will not enter the details of this argument, but will just mention that promoting a vector field to a gauge field is the only known consistent way to quantize it, and that its quantization requires adding a "gauge fixing" term, for example $-(\partial_\mu A^\mu)^2/(2\xi)$, to the Lagrangian in order to get rid of the redundant degrees of freedom.

Given the impressive success of QED as a gauge theory (both in confronting the experiment and in addressing the quantization of vector bosons), it is natural to consider the generalization of the gauge principle to a generic (compact, Lie) group. Particularly interesting is the case of non-Abelian groups, which leads to at least two qualitative differences with the simple Abelian case of QED: the vector bosons acquire self-interactions and the coupling (which in QFT depends on the energy scale of the process in which it is measured) may get stronger at lower scales. Non-Abelian gauge theories are therefore a candidate to describe strong interactions. Indeed, it turns out that strong interactions can be described by a gauge theory based on a SU(3) gauge group acting on a "color" degree of freedom, quantum chromodynamics (QCD). QCD will be addressed in the lectures by Michelangelo Mangano.

Let us then consider a general gauge theory. Having motivated the construction in some detail in the simple Abelian case, I will now just provide the result of the generalized construction in the form of a recipe to construct a general renormalizable gauge theory of spin 0, 1/2, 1 fields in four space-time dimensions (up to non-perturbative effects), without any further proof or motivation. The ingredients that need to be specified are the following. 1) A compact gauge group G (which specifies the vector field content). 2) The scalar fields ϕ_i, $i = 1 \ldots n_s$ and fermion fields ψ_i, $i = 1 \ldots n_f$ (collectively denoted in the following as Φ). 3) The transformations (quantum numbers) of the latter under the action of the gauge group. 4) A renormalizable Lagrangian $\mathcal{L}_0(\Phi, \partial\Phi)$ for the scalar and fermion fields and their derivatives symmetric under the global (space-time independent) transformations of the fields specified above. Before seeing how the Lagrangian is obtained in terms of the above ingredients, let us elaborate on the specification of the fermion field content and quantum numbers. The point is that the fermions are usually described in terms of Dirac spinors ψ_i that, as we said in the introduction, combine two independent components with left and right chirality respectively, $\psi_{iL} = (1-\gamma_5)\psi_i/2$ and $\psi_{iR} = (1+\gamma_5)\psi_i/2$ (in a given convention for the gamma matrices), each being an irreducible representation of the Lorentz group. Now, the only requirement on the action of the gauge group on fermions is that it commutes with the Lorentz group. As a consequence, its most general action is not bound to transform ψ_{iL} and ψ_{iR} in the same way: they could have different quantum numbers. This is not the case in QED and QCD, or in any parity conserving theory of charged particles, but turns out to be the case for the SM. A theory like the SM, where the left and right chirality components of fermions transform under inequivalent representations of the gauge group is called "chiral". A chiral theory violates parity. Actually, something even more general could happen. Since the left and right chirality representations of the Lorentz group are conjugated to each other, ψ^*_{iR} is left handed, as ψ_{iL}. Therefore, the most general action of the gauge group on the fermions could mix the ψ_{iL} and ψ^*_{iR} fields. This is not the case in the SM, but does happen in grand-unified theories. In summary, the fermion field content should be specified in terms of the left-handed fermions, by specifying the quantum numbers of both ψ_{iL} and ψ^*_{iR}(or equivalently ψ_{iL} and ψ_{iR}).

Let us now construct the Lagrangian associated to the above ingredients by the gauge principle. The recipe is the following. Choose a basis of generators t_a for the group G as above and associate a real vector field A^a_μ and a field strength $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g f_{abc} A^b_\mu A^c_\nu$ to it. Define the covariant derivative $D_\mu = \partial_\mu + igA^a_\mu T_a$, where T_a is the action of the generator t on the fields Φ (then given a gauge transformation $\Phi(x) \to U(x)\Phi(x)$ we have $F^a_{\mu\nu} T_a \equiv \mathbf{F}_{\mu\nu} \to U \mathbf{F}_{\mu\nu} U^{-1}$ and $D_\mu(U(x)\Phi(x)) =$

$U(x)(D_\mu\Phi(x)))$. The gauge Lagrangian is then

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a + \mathcal{L}_0(\Phi, D\Phi) - \frac{1}{2\xi}(\partial_\mu A^\mu_a)^2 + \text{ghosts}, \tag{6}$$

where the first term is the so-called Yang-Mills Lagrangian for the vector fields, the derivative has been promoted to a covariant derivative, the next to last term fixes the gauge and the last term involves auxiliary, unphysical anticommuting scalars (ghosts) important to preserve unitarity, whose discussion goes well beyond the purpose of these lectures (they will not play a role in what follows). As a simple exercise one can recover the QED Lagrangian in eq. (3) from the free Lagrangian of a Dirac fermion. The gauge interactions between matter and gauge fields come from the terms involving the covariant derivative, i.e. from the kinetic terms in the Lagrangian $\mathcal{L}_0$ before gauging. In the case of a set of fermions $\Psi = (\psi_1 \dots \psi_{n_f})^T$ subject to an action $\Psi \to T_a\Psi$ of the generators, the kinetic term $\overline{\Psi}i\partial^\mu\gamma_\mu\Psi$ gives

$$\overline{\Psi}i\partial_\mu\gamma^\mu\Psi \to \overline{\Psi}iD_\mu\gamma^\mu\Psi = \overline{\Psi}i\partial_\mu\gamma^\mu\Psi - gA^a_\mu\overline{\Psi}\gamma^\mu T_a\Psi. \tag{7}$$

The gauge bosons self-interactions are $gf_{abc}\partial_\mu A^a_\nu A^{b\mu}A^{c\nu} - \frac{1}{4}g^2 f_{abc}f_{ab'c'}A^{b\mu}A^{c\nu}A^{b'}_\mu A^{c'}_\nu$. All gauge interactions are determined in terms of a universal gauge coupling g. In the case in which the gauge group is made of several factors (as in the case of the SM, as we will see), there is one independent gauge coupling for each irreducible factor.

3 The SM gauge interactions

We are now ready (at last) to start presenting the Standard Model Lagrangian. As the SM is a gauge theory, it suffices to specify the ingredients listed above to define it completely. It is however instructive to outline some of the logic that lead to the choice of such a model (and ingredients) to describe electroweak interactions at energies of the order of the electroweak scale or higher. The way the SM emerges from the experimental information that was available when it was discovered can be considered (in quite a reductive way) as one of the nicest, and definitely most useful, model-building exercises ever [3]. Let us then review it, following a shortcut making use of some of our present knowledge.

Let us consider to begin with the fermion content of a single family of fermions, whose electromagnetic, strong, and weak interactions we want to describe in the context of a theoretically consistent theory that can be extrapolated at and beyond what we now know to be the electroweak scale. We then have the fermion fields ν, e, u, d. QED and QCD can be accounted for by a gauge Lagrangian based on gauge groups $\mathrm{U}(1)_{\mathrm{em}}$ and $\mathrm{SU}(3)_c$ respectively ("c" stands for color), where the electric charges of the above fermions where given in the introduction and the quarks transform as triplets under the color SU(3). We are left with weak interactions. While QED and QCD are parity conserving theories of charged particles and as such can be described in terms of Dirac spinors, we know that weak interactions violate parity. In order to describe weak interactions, we therefore expect to have to split the above spinors into left- and right-handed components ν_L, e_L, e_R, u_L, u_R, d_L, d_R (I did not include a right-handed neutrino component, so far unobserved). Indeed, it was known that weak interactions could be described in terms of the effective four-fermion interaction [4]

$$\mathcal{L}_{\text{weak}} = 4\frac{G_F}{\sqrt{2}} j_c^\mu j_{c\mu}^\dagger, \quad \text{where} \quad G_F^{-1/2} \sim 250\,\text{GeV} \quad \text{and} \quad j_c^\mu = \overline{\nu_L}\gamma^\mu e_L + \overline{u_L}\gamma^\mu d_L \tag{8}$$

only involves left-handed fields (thus breaking parity). We have written the charged current directly in terms of quarks, thus taking advantage of our present knowledge. The experimental determination of the Lorentz and chirality structure of the interaction above had been crucial for the discovery of the SM.

The coefficient of the interaction in eq. (8) has negative energy dimension. We are therefore dealing with a non-renormalizable interaction. The effect of such an interaction at energies $E \ll G_F^{-1/2}$ will be suppressed by powers of $E/G_F^{-1/2}$. That's why weak interactions are weak at low energy. At energies comparable to the scale $G_F^{-1/2}$ and above, however, the interaction above becomes stronger and stronger and the theory eventually becomes incalculable. This is signaled by the failure of the perturbative calculations of unitarity. The cross section for processes induced by the interaction in eq. (8) would in fact grow as $G_F E^2$ at high energy, instead of falling with $1/E^2$, as predicted by unitarity. In order to be able to make sense of weak interactions at $E \sim G_F^{-1/2}$ and beyond, we need to replace the effective description in eq. (8) with a description in terms of renormalizable interactions. This is easily done. The effective interaction in eq. (8) can be generated by the exchange of an heavy field with mass M, as shown in Fig. 2 for the case of the leptonic interactions. When the energy E involved in the process is smaller than M, the propagator $(E^2 - M^2)^{-1}$ can be approximated by $-1/M^2$, and one recovers the effective description of weak interactions. In other words, the wavelength of the process is too large to probe the detailed renormalizable structure of the interaction, which appears to be pointlike, as in the left hand side of Fig. 2. In this $E < M$ regime, the amplitude for the process in the Figure grows as $A \sim g^2/M^2$ and the cross section as $\sigma \sim g^4(E^2/M^4)$, where g is the coupling involved in the renormalizable vertexes. On the other hand, when the energy becomes large, $E > M$, the propagator $(E^2 - M^2)^{-1}$ can be approximated by $1/E^2$, the amplitude goes as $A \sim g^2/E^2$ and the cross section as $\sigma \sim g^4/E^2$, in agreement with the unitarity bound.

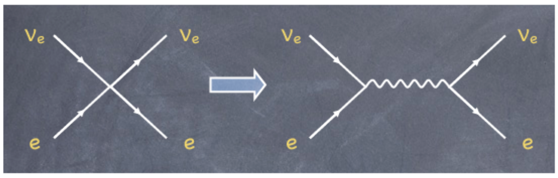


Figure 2: Weak interactions in terms of renormalizable physics.

We now have to identify the degrees of freedom exchanged in Fig. 2 and their renormalizable interactions with the SM fermions. Here is where the determination of the Lorentz and chirality structure of the effective weak interactions turns crucial. The Lorentz structure indicates that the heavy particle exchanged is a (charged) vector, $W^\pm_\mu$, and the chirality structure shows that it only couples to left-handed fermions. We can then write the fermion-vector interaction entering the Feynman diagram in Fig. 2 (and the corresponding quark one) as

$$-\mathcal{L}_W = \frac{g}{\sqrt{2}} W^+_\mu \overline{\nu_L} \gamma^\mu e_L + \frac{g}{\sqrt{2}} W^+_\mu \overline{u_L} \gamma^\mu d_L + \text{h.c.}, \tag{9}$$

where the $\sqrt{2}$ is conventional. We can then express the Fermi constant G_F in terms of the W mass M_W and coupling g as $G_F/\sqrt{2} = g^2/(8M_W^2)$. We have then learned two things. First, weak interactions involve a new vector field $W^\pm_\mu$. As said, vectors should be described by means of a gauge theory. We should then find a gauge group G_{SM} and assign quantum numbers to the fermions in such a way that the interaction in eq. (9) be given by a gauge Lagrangian in the form in eq. (6). Second, we will have to understand how a vector boson, $W^\pm_\mu$, can get massive, as we pointed out before that a mass term for a vector boson is not gauge invariant. This will require the spontaneous symmetry breaking of the gauge symmetry, which we will address in Section 4. In the meanwhile, in this Section, we will ignore the latter issue and determine the gauge structure of the SM.

In order to determine the gauge group G_{SM} and the fermion quantum numbers generating the gauge interactions in eq. (9), we need to compare eq. (9) with the general form of the vector-fermion interactions in eq. (7). Since eq. (7) involves real vectors, we first write $W^\pm_\mu = (W^1_\mu \mp i W^2_\mu)/\sqrt{2}$ in terms of the real vectors $W^{1,2}_\mu$. The interactions in eq. (9) can then be written as $-\mathcal{L}_W = g\overline{L}\gamma^\mu(W^1_\mu T_1 + W^2_\mu T_2)L + g\overline{Q}\gamma^\mu(W^1_\mu T_1 + W^2_\mu T_2)Q$, where $T_{1,2} = \sigma_{1,2}/2$ and we have introduced the left handed lepton and quark doublets

$$L \equiv \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} \qquad Q \equiv \begin{pmatrix} u_L \\ d_L \end{pmatrix}. \tag{10}$$

It is now easy to compare with eq. (7) and identify two of the generators of the gauge theory we want to determine to be T_1 and T_2. Since the generators of a gauge theory form a Lie algebra and $[T_1, T_2] = iT_3$, with $T_3 = \sigma_3/2$, we conclude that T_3 is also a generator, to which a third gauge boson, W^3_μ, must be associated. Note that W^3_μ cannot be identified with a photon or a gluon, the gauge bosons associated to QED and QCD (the couplings to fermions are different). The gauge principle therefore leads to the prediction of a new (neutral) interaction. The three generators above constitute the algebra of generators of $\text{SU}(2)_L$, which must therefore be a factor of the gauge group, $G_{\text{SM}} \supseteq \text{SU}(2)_L$. The index L refers to the fact that $\text{SU}(2)_L$ only acts on the left handed components of the fields. The latter organize themselves in doublets of $\mathrm{SU}(2)_L$, as in eq. (10), whereas the right handed components should be $\mathrm{SU}(2)_R$ singlets, as they do not appear in $\mathcal{L}_W$.

Since we know that electromagnetic and strong interactions are also described by a gauge theory, the gauge group G_{SM} should also include the corresponding generators. Strong interactions can be accounted for by a $\mathrm{SU}(3)_c$ factor commuting with the rest of the SM group, under which u_L, u_R, d_L, d_R all transform as triplets (they carry a "color" index). We do not need to discuss them here. Let us instead concentrate on the electromagnetic interactions. They are associated to a electric charge generator Q (not to be confused with the quark doublet in eq. (10)), whose value on the fermions gives their electric charges (by convention in units of e). As a linear combination of generators is still a generator, the combination $Y \equiv Q - T_3$ should also be a generator, which is called "hypercharge". The interest of Y is that it turns out to commute with the $\mathrm{SU}(2)_L$ generators. The hypercharges of the fermions can be computed in terms of their electric and T_3 charges. One then finds that the hypercharges of the two lepton doublet components, ν_L and e_L, coincide: $Y(\nu_L) = Y(e_L) = -1/2$. Therefore, the hypercharge and $\mathrm{SU}(2)_L$ generators commute on the lepton doublet L. But then they should also commute on the quark doublet Q. This means that the hypercharges of u_L and d_L should also coincide, if what we are doing makes sense. Let us then cross our fingers and compute those hypercharge. For u_L we have $Y(u_L) = Q(u_L) - T_3(u_L) = 2/3 - 1/2 = 1/6$ and for d_L $Y(d_L) = Q(d_L) - T_3(d_L) = -1/3 - (-1/2) = 1/6$. This means that we are one the right track (unfortunately, about 40 years too late for the Nobel prize). The generator Y is therefore associated to a $\mathrm{U}(1)_Y$ subgroup of G_{SM} that commutes with the rest of the group. All in all, this model building exercise identifies the SM gauge group to be

$$G_{\mathrm{SM}} = \mathrm{SU}(3)_c \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y, \tag{11}$$

where the last two factors constitute the electroweak group. The weak and electromagnetic interactions are described by this subgroup, with the electromagnetic interactions associated with a combination of $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ generators, $Q = Y + T_3$. The quantum numbers of the fermions with respect to the three factors of the group are summarized in Table 1. The first two columns show the transformation properties under $\mathrm{SU}(3)_c$ and $\mathrm{SU}(2)_L$, while the last columns shows the hypercharge of each field. The quantum numbers are the same for the three families, which goes under the name of family replication. We do not have a compelling explanation for such a replication. The values of the hypercharges for the right-handed fermions are given by $Y = Q$, as the $\mathrm{SU}(2)_L$ generators vanish on the right-handed fermions. Table 1 fully specifies the SM gauge interactions of the SM fermions. We will analyze them in detail in Section 6.

	$SU(3)_c$	$SU(2)_L$	$U(1)_Y$
L	1	2	$-1/2$
e_R	1	1	-1
Q	3	2	$1/6$
u_R	3	1	$2/3$
d_R	3	1	$-1/3$

Table 1: Gauge quantum numbers of the SM fermions (one family).

Before closing this Section, let us discuss anomalies. The quantum numbers in Table 1 have quite a peculiar pattern. What is (pleasingly) surprising about it is that they satisfy the gauge anomaly cancellation conditions. Anomalies arise when a symmetry of the Lagrangian is not preserved by quantum corrections so that the corresponding current is conserved at the classical level but not at the quantum level. In some case, anomalous symmetries are not a problem, they are actually welcome. This is the case for example of the (global) chiral symmetry of QCD, or of scale invariance, which is broken at the quantum level by the renormalization scale, leading to the anomalous dimensions of fields. In the case of gauge symmetries, however, anomalies should be avoided in order not to make the theory inconsistent. There is a simple condition that needs to be verified in order for gauge anomalies to vanish. The presence of gauge anomalies depends on how the gauge group generators act on the left-handed fermions (ψ_{iL} and ψ^*_{iR}). Let T^L_a give the action of the generators on the left-handed fermions. Then the gauge symmetry survives quantum corrections iff $0 = T_{abc} \equiv \mathrm{tr}(T^L_a\{T^L_b, T^L_c\})$ for each choice of generators. As a consequence of this condition, a gauge symmetry acting in the same way on left and right chirality fermions (as QED, QCD) turns out to be automatically non-anomalous. There is no guarantee, however, that a chiral theory such as the SM be non-anomalous. In order to check whether this is the case we need to compute T_{abc} for a combination of any three SM generators. It is a useful exercise to verify that T_{abc} indeed always vanishes. This is a highly non-trivial property depending on the fact that i) $SU(3)_c$ acts in the same way on left and right chirality fermions, ii) the traces of SU(2) and SU(3) generators vanish, iii) three non-trivial relations hold among the hypercharge quantum numbers: $2Y_Q - Y_{u_R} - Y_{d_R} = 0$, $Y_L + 3Y_Q = 0$, and $2Y^3_L + 6Y^3_Q - 3Y^3_{u_R} - 3Y^3_{d_R} - Y^3_{e_R} = 0$. Moreover, the relation $2Y_L + 6Y_Q - 3Y_{u_R} - 3Y_{d_R} - Y_{e_R} = 0$ accounts for the vanishing of another type of anomaly, the gravitational one. The fact that the SM is anomaly free is reassuring. Of course it would be nice to understand whether there is a reason, besides the cancellation of anomaly itself, why the values of the hypercharge we measure happen to satisfy the anomaly cancellation condition, or, more generally, if there is a reason for the peculiar pattern of gauge quantum numbers in Table 1. Grand-unification theories may answer such questions.

4 Spontaneous symmetry breaking (SSB)

In order to proceed further we need to address the puzzle of vector boson masses. The $W^\pm_\mu$ vector bosons leading to the effective interaction in eq. (8) have a mass of about 80 GeV. On the other hand, the gauge symmetry prevents a mass term for the gauge vectors. The gauge symmetry must then be somehow broken, in order for the gauge vectors to acquire a mass. This can be done through the mechanism of spontaneous symmetry breaking, which we discuss in this Section. Before illustrating it, let us observe that the vector boson masses are not the only problem. The SM fermions would also forced to be massless in the presence of an exact $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry. One can in fact verify that no gauge invariant fermion mass term can be written, given the quantum numbers in Table 1. The most general mass term couples two left-handed fermions in a combination that, as all the terms in the Lagrangian, must be invariant under the gauge symmetry. There are three possible types of mass terms that can be formed: $\psi_L \psi_L$, $\psi^*_R \psi_L$, $\psi^*_R \psi^*_R$. It is easy to verify that no such combination of left or right fermions in Table 1 can be gauge invariant. This is due to the SM gauge symmetry being chiral even when restricted to an arbitrary subset of fermions[3]. Spontaneous breaking of the gauge symmetry is then needed in order to account for the SM spectrum of both vector bosons and fermions. In the following, I will illustrate the main features of spontaneous symmetry breaking of global and gauge symmetries (Higgs mechanism). In the next Section we will see how this applies to the SM.

Spontaneous symmetry breaking (SSB) is interesting and elegant because no explicit breaking of the symmetry is introduced. The equations of the dynamics are exactly symmetric, but they admit solutions that are not. In particular, one has SSB when the ground state of the system is not symmetric. It is then the system itself that "spontaneously" breaks the symmetry. In the context of QFT, SSB is characterized by i) the Lagrangian being invariant under the symmetry, ii) the currents associated to the symmetry being conserved, also at the quantum level, iii) the vacuum (ground state) of the theory being not invariant under the symmetry, iv) the spectrum being not invariant. Its features are i) it allows a consistent breaking of gauge symmetries, in particular a consistent quantization of massive vectors; ii) it turns out to be realized in nature both in the case of global and gauge symmetries.

[3]From the point of view of the physics at the electroweak scale, the chiral structure of the SM is a puzzle. From the point of view of physics at the Planck scale, however, such structure is welcome because it explains why the SM fermions are so light with respect to the Planck scale: they are prevented from getting a mass term until the electroweak symmetry is broken, at a much lower scale. One can take this argument a step further and argue that this is the reason why the SM fermions are fully chiral: because if they were not there would be no reason why they should not be much heavier. Chiral fermions may just be the only ones surviving at low energy because of their very chiral structure.

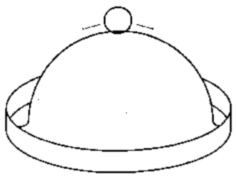


Figure 3: An example of SSB in classical mechanics.

There are several classic examples of SSB. In classical mechanics one can consider a ball forced to live on the symmetric (under rotations around the central axis) surface in Fig. 3 subject to a gravitational field pointing downwards. The equations of the dynamics are symmetric. The central point, where the ball is in the Figure, is also symmetric, but unstable. In order to live in a stable ground state, the system should choose a position a the bottom of the surface, thus spontaneously breaking the rotational symmetry. In quantum mechanics, one can consider a rotationally invariant system of coupled spins (a ferromagnet). The minimum energy state is reached when the spins are aligned. In order to live in the ground state, therefore, the system must choose a common direction for the spins, thus spontaneously breaking rotational invariance. In QFT, a qualitative difference arises, as the number of degrees of freedom is not finite anymore and a quantum superposition of degenerate vacua is not allowed: different ground states are not described within the same Hilbert space.

We now discuss in more detail SSB in QFT. We are interested to the spontaneous breaking of gauge invariance, but we need to discuss the spontaneous breaking of global symmetries first. Let us call Ω the ground state of our QFT. SSB arises iff the vacuum expectation value (vev) of the fields in the theory, $\langle\Phi\rangle \equiv \langle\Omega|\Phi(x)|\Omega\rangle$, is not invariant (therefore not vanishing) under the symmetry ($\langle\Phi\rangle$ is invariant if Ω is). As we do not want to break Poincaré invariance, only scalars can get a non vanishing vev and $\langle\Phi(x)\rangle$ does not depend on the space-time coordinate x. Given a Lagrangian, the value of the vev of the scalar fields can be easily obtained by minimizing the effective scalar potential, i.e. the scalar potential including the so-called one-particle irreducible quantum corrections. It is often possible to neglect quantum corrections and just consider the minimization of the classical potential, which is what we will do in the following. A simple explicit example of SSB in QFT can be obtained in the theory of a complex scalar field ϕ with Lagrangian

$$\mathcal{L} = (\partial_\mu\phi)^\dagger(\partial^\mu\phi) - V(\phi^\dagger\phi), \quad V(\phi^\dagger\phi) = \mu^2\phi^\dagger\phi + \frac{\lambda}{2}(\phi^\dagger\phi)^2. \tag{12}$$

The Lagrangian is symmetric under a global U(1) transformation $\phi(x) \to e^{-i\alpha}\phi(x)$. The parameter λ must be non-negative in order for the potential to be bounded from below. We will take it to be strictly positive. On the other hand, the parameter μ^2 can have both signs (despite it is written as a square to stress that it has the dimension of a squared mass). The shape of the potential V and the structure of the ground state crucially depends on that sign, as shown in Fig. 4. If $\mu^2 < 0$, the minimum of the potential corresponds to $\langle\phi\rangle = ve^{i\theta}$, where

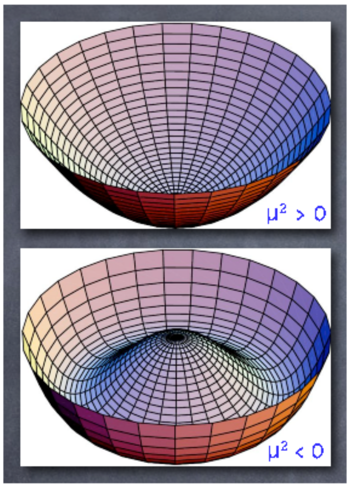


Figure 4: Graphical representation of the potential in (13).

$v^2 = |\mu^2|/\lambda$ and θ parametrizes the position of ϕ in the circle with radius v of degenerate minima. The system chooses an arbitrary value of θ, thus spontaneously breaking the U(1) symmetry. With no loss of generality, we can assume $\theta = 0$.

In view of the shape of the potential, it is convenient to parametrize the field as $\phi(x) = r(x)e^{ig(x)}$, where $g(x)$ parametrizes the "flat directions" along which the potential is constant. As V does not depend on $g(x)$, the corresponding real degree of freedom is massless and only has derivative interactions. Such massless degrees of freedom always arise in the presence of SSB in QFT. They are called "Goldstone" bosons. One can also use a linear parametrization of the Goldstone boson by expanding $\phi(x) = v + \phi'(x) = v + (h(x) + iG(x))/\sqrt{2}$. $G(x)$ can be considered as the linearization of $vg(x)$. In terms of h and G, the potential is

$$V = \frac{\lambda}{8}(h^2 + G^2)^2 + \frac{\lambda}{\sqrt{2}} vh(h^2 + G^2) + |\mu^2|h^2 + \text{const.} \tag{13}$$

We can again verify that G is massless, whereas the physical degree of freedom h acquires a mass proportional to the symmetry breaking scale v and to its self-coupling, $m_h^2 = 2|\mu^2| = 2\lambda v^2$. The two parameters μ^2, λ in the potential V can be traded for v and m_h^2.

The discussion above can be generalized to the case of a generic continuous global symmetry group G and a generic scalar field content. Here are the main features of the generalization. Let us call H the subgroup of G that is not broken by the vev of the scalar fields (under which the vev of the scalar fields is invariant). We can correspondingly divide the generators of G into two sets: the unbroken ones, which annihilate the vacuum, and the broken ones, the orthogonal set. According to the Goldstone theorem [5] each broken generator in G/H is associated to an independent massless scalar (Goldstone boson), carrying the same quantum numbers as the generators.

The prototypical example of spontaneously broken global symmetry in QFT is the chiral symmetry of QCD with two quarks. In the limit in which the up and down quark masses coincide the QCD Lagrangian (not including electromagnetic interactions) for u and d is invariant under independent unitary transformations of the left handed and right handed components of the u and d fields, corresponding to the symmetry group $\mathrm{U}(2)_L \times \mathrm{U}(2)_R = \mathrm{SU}(2)_L \times \mathrm{U}(1)_L \times \mathrm{SU}(2)_R \times \mathrm{U}(1)_R$. Let us call $j_L^{\mu a}$, j_L^{μ}, $j_R^{\mu a}$, j_R^{μ} the currents associated to the four group factors. Then $j_V^{\mu} = j_R^{\mu} + j_L^{\mu}$ is the conserved current associated to Baryon number and $j_A^{\mu a} = j_R^{\mu a} + j_L^{\mu a}$ are the conserved isospin currents. What about the "axial" currents j_A^μ and $j_A^{\mu a}$? The first one, j_A^μ, turns out to be broken by quantum corrections. The corresponding symmetry is anomalous. The second current, $j_A^{\mu a}$, turns out to be conserved, on the other hand. Therefore, it either corresponds to a symmetry or it is spontaneously broken. We do not have evidence for $j_A^{\mu a}$ to correspond to a symmetry. If it did, particles would organize themselves in multiplets with same spin, Baryon number, parity and approximately (in the real world in which $m_u \neq m_d$, electromagnetic interactions exist, and the chiral symmetry $\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$ is only approximate) same mass, which we do not observe. On the other hand, we do have evidence for the symmetry to be spontaneously broken. If that is the case, we should observe three light (not exactly massless in the real world) pseudo- Goldstone bosons with zero Baryon number, spin, negative parity and isospin 1, as the corresponding broken generators. The lightest hadrons, the pions, have indeed all those properties. They are therefore considered to be the pseudo- Goldstone bosons arising because of the spontaneous breaking $\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R \to \mathrm{SU}(2)_V$ of the approximate chiral symmetry of the light quark QCD Lagrangian.

We have discussed so far the spontaneous breaking of a global symmetry. We now discuss the case we are more directly interested in, the spontaneous breaking of gauge invariance. The main feature of such a phenomenon is that the gauge vector associated to each broken generator gets a longitudinal component and a mass. The additional longitudinal degree of freedom is provided by the Goldstone boson associated to the broken generator, which gets "eaten up" by the vector. This is just what we need in order to describe the massive vector boson we observe in nature, the $W^\pm$ and the Z, in a theoretically consistent way.

In order to see how this works, let us consider the complex scalar field again and promote the global U(1) symmetry to a gauge symmetry. Our gauge theory machinery gives

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi^\dagger\phi) + \text{gauge fixing}, \quad D_\mu = \partial_\mu + igA_\mu. \tag{14}$$

Let us break U(1) spontaneously by taking $\mu^2 < 0$. The complex field develops a vev $\langle\phi\rangle = v$, as before. Correspondingly, a mass term $M^2 = 2g^2v^2$ is generated for the vector boson, proportional to the symmetry breaking scale and to its (gauge) coupling to ϕ, as can be seen from $(D_\mu\phi)^\dagger(D^\mu\phi) = (\partial_\mu\phi')^\dagger(\partial^\mu\phi') + M^2A^\mu A_\mu/2 + \ldots$. The Goldstone boson gets eaten by the vector boson, of which it becomes the longitudinal component. This can be seen by parametrizing $\phi(x)$ in terms of $r(x)$ and $g(x)$ as above and by noticing that

$$\begin{cases} \phi(x) = r(x)e^{ig(x)} \\ A_\mu(x) \end{cases} \quad \text{is equivalent to} \quad \begin{cases} \phi(x) = r(x) \\ A_\mu(x) - \dfrac{1}{g}\partial_\mu g(x) \end{cases}, \tag{15}$$

as the two configurations are related by a gauge transformation. We can therefore choose a "unitary" gauge in which the field $\phi(x)$ is real, as on the right side of eq. (15), and does not contain $g(x)$, which can be recognized as the longitudinal component of $A_\mu(x)$. This can be generalized to the case of a generic continuous global symmetry group G and a generic scalar field content. In the general case, as mentioned, the gauge vector associated to each broken generator gets a mass by absorbing the corresponding Goldstone boson. This is what goes under the name of "Higgs" mechanism. In the context of gauge theories, unlike the case of global symmetries, the Goldstone bosons do not correspond to physical scalar degrees of freedom. The scalar fields spontaneously breaking the gauge symmetry are called Higgs fields.

Before closing this Section we mention that global and gauge symmetries can be spontaneously broken without the need of scalar fields. The role of the vev of the scalar field is then played by the condensate of a fermion bilinear that arises dynamically as a consequence of new strong interactions. This is how chiral symmetry breaking in QCD is thought to arise. While it is not excluded that such dynamical symmetry breaking mechanism plays a role in the breaking of the electroweak symmetry, such a possibility is at present disfavored. We will not further consider it in these lectures.

5 The SM Higgs and flavour

We are now ready to go back to the SM and discuss its spontaneous breaking. The existence of fermion and vector boson masses is evidence that the SM gauge invariance should be spontaneously broken. We know on the other hand that strong and electromagnetic interactions are not broken as, for example, the electric charge is conserved and the photon is massless. Therefore, the SSB of the SM should preserve $\mathrm{SU}(3)_c\times\mathrm{U}(1)_{\mathrm{em}}$ as an unbroken subgroup.

In order to spontaneously break the SM, we should introduce a scalar field, the Higgs, developing a vev and specify its quantum numbers under the SM gauge group. Such quantum numbers are dictated by the need of the fermions to get a mass term, as we now see.

Let us consider the electron mass term, which has the form $m(\overline{e_R}e_L + \text{h.c.})$. As the left component e_L is contained in the lepton doublet L, that interaction should originate from a SM invariant interaction involving the bilinear $\overline{e_R}L$. Now, such a bilinear is not invariant under G_{SM}. It transforms as a doublet with $Y = 1/2$ under $\mathrm{SU}(2)_L\times\mathrm{U}(1)_Y$ (and is of course invariant under $\mathrm{SU}(3)_c$). That is why the electron mass is not allowed by the gauge symmetry. However, $\overline{e_R}L$ can be part of a gauge invariant renormalizable interaction involving an additional field with appropriate quantum numbers. The only possibility is the Yukawa interaction $\lambda\overline{e_R}LH^*$ with a complex doublet scalar field H (conventionally taken as the conjugated of the field appearing in the interaction), the Higgs field. H is a doublet with $Y = 1/2$ under $\mathrm{SU}(2)_L\times\mathrm{U}(1)_Y$ and is a $\mathrm{SU}(3)_c$ singlet. Its two components contract with the two components in L. A mass term for the electron can now be generated if H gets a vev, once the value of the vev is substituted to the field in the Yukawa interaction above. It turns out that one Higgs doublet H is enough to give rise to the mass of all the SM fermions. Let us in fact write the most general Yukawa Lagrangian involving the Higgs field. Such a Lagrangian is given by

$$-\mathcal{L}_Y = \lambda^E_{ij}\overline{e_{iR}}L_j H^* + \lambda^D_{ij}\overline{d_{iR}}Q_j H^* + \lambda^U_{ij}\overline{u_{iR}}Q_j H + \text{h.c.}, \tag{16}$$

where we have included all the three SM families through the family indices $i, j = 1, 2, 3$. As a consequence of this family structure, each of the three Yukawa couplings in eq. (16) is a generic 3×3 complex matrix. Once the Higgs gets a vev, all the SM fermions get a mass proportional to their Yukawa couplings. SU(2) and SU(3) contractions have been understood in eq. (16). The SU(2) invariant contraction of the doublet indices of Q and H in the up quark Yukawa interaction is obtained by means of the 2×2 antisymmetric tensor ϵ_{ab} as $QH = Q_a\epsilon_{ab}H_b$ ($\epsilon_{12} = 1$). The Lagrangian $\mathcal{L}_Y$ is the origin of the flavour structure of the SM. It is because of that Lagrangian that we can tell for example an electron from a muon. The gauge Lagrangian, in fact, does not make any difference between them.

In order to fully specify the SM Lagrangian we only miss the Higgs Lagrangian. In the same spirit used for the Yukawa Lagrangian, we can simply write the most general renormalizable Lagrangian involving the Higgs field. That turns out to be

$$\mathcal{L}_H = (D_\mu H)^\dagger (D^\mu H) - V, \qquad V = \mu^2 H^\dagger H + \frac{\lambda_H}{2}(H^\dagger H)^2. \tag{17}$$

As in the SSB example with a single complex scalar field, we take $\lambda_H > 0$ and $\mu^2 < 0$ in order to obtain a stable, symmetry breaking potential.

This completes the definition of the SM. We have chosen the gauge group as in eq. (11); we have specified the fermion and scalar content (the three replications of the fermions in Table 1 and the Higgs doublet); their quantum numbers under the SM gauge group (Table 1 and the Higgs assignment above); we have also specified the most general renormalizable globally symmetric Lagrangian for the above fields (the kinetic terms of all the fermions + $\mathcal{L}_Y$ + $\mathcal{L}_H$). The gauge theory machinery then allows to specify the full Lagrangian as in eq. (6). We will analyze the Higgs Lagrangian, together with the rest of the SM Lagrangian, in the next Section.

6 Analysis of the SM Lagrangian

Having defined and motivated the SM, let us know analyze the SM Lagrangian and spell out some phenomenological implications. Let us start from where we stopped, the Higgs sector.

In the minimum of its potential, the Higgs doublet develops a vev. With no loss of generality, such a vev can be written as

$$\langle H\rangle = \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad \text{with} \quad v > 0 \quad \text{and} \quad v^2 = \frac{|\mu^2|}{\lambda_H} \approx (174\,\mathrm{GeV})^2. \tag{18}$$

Any other form of the Higgs vev, in fact, is equivalent to that in eq. (18), up to a $\mathrm{SU}(2)_L\times\mathrm{U}(1)_Y$ gauge transformation. The scale $v \approx 174\,\mathrm{GeV}$ is called the electroweak symmetry breaking scale, or electroweak scale. Let us identify the unbroken part of the SM group, or equivalently the generators that annihilate the Higgs vev. Since the Higgs does not feel strong interactions, the latter are certainly unbroken. Let us therefore concentrate on the electroweak group $\mathrm{SU}(2)_L\times\mathrm{U}(1)_Y$. Its generic generator can be written as $T = aY + b_a T_a$, with a, b_a real. When acting on the Higgs, $T_a = \sigma_a/2$ and $Y = 1/2$. Therefore $T\langle H\rangle = (v/2)(b_1 - ib_2, a - b_3)^T$ and the unbroken generators, for which $T\langle H\rangle = 0$, are those for which $b_1 = b_2 = 0$ and $a = b_3$. There is then only one (up to normalization) electroweak generator unbroken by the Higgs vev, given by $T_3 + Y = Q$. The electric charge is unbroken, as wished. Note that the latter can be considered as a prediction, as the Higgs quantum numbers, determining the unbroken generators, were fixed by independent considerations (obtaining a mass for the electron). Out of the 4 generators of the electroweak group only one is unbroken, which means that 3 are broken. We then expect 3 vector bosons to acquire a mass and 3 Higgs real degree of freedom (the Goldstones) to be eaten up by them. Out of the 4 real (2 complex) Higgs degrees of freedom, only one then correspond to a physical scalar, the Higgs boson. In order to identify the Goldstone (and thus the physical) degrees of freedom we can use a general property of the Goldstone bosons: they correspond to displacements from the vev along the flat directions of the potential. We can move along the flat directions of the potential by performing G_{SM} transformation (which leave the potential invariant) along an arbitrary set of broken generators. From $\delta\langle H\rangle = i\epsilon G_a T_a\langle H\rangle = \epsilon(v/2)(iG_1 + G_2, -iG_3)^T$ we see that we can write the Higgs doublet in terms of the Goldstone components $G^\pm = (G_1 \mp iG_2)/\sqrt{2}$, $G_0 = -G_3$ and the physical component h as

$$H = \begin{pmatrix} iG^+ \\ v + \dfrac{h + iG^0}{\sqrt{2}} \end{pmatrix}. \tag{19}$$

With respect to the CP transformation under which $H \to H^*$, h is even and the Goldstones are odd. We can write the Higgs potential in the unitary gauge in which the Goldstones are removed from the Higgs fields and incorporated in the corresponding vector bosons as follows:

$$V(h) = V(H)_{G=0} = \frac{m_h^2}{2}h^2 + \frac{\lambda_H}{\sqrt{2}}vh^3 + \frac{\lambda_H}{8}h^4 + \text{const}, \tag{20}$$

where the Higgs mass is given by $m_h^2 = 2|\mu|^2 = 2\lambda_H v^2$ and is proportional to the electroweak symmetry breaking scale and to the Higgs self-coupling λ_H.

While the electroweak scale is known from the measurement of the Fermi constant G_F, as we will see, the Higgs mass (or equivalently the Higgs coupling) is at the moment an unknown parameter. We have however three different constraints on it.

The first is a theoretical constraint: in order to avoid a strong coupling regime, the Higgs mass should be lighter than about a TeV. If the Higgs mass was heavier than that, the theory would become strongly interacting before the Higgs could be produced. This is not a priori excluded. However, keeping the theory perturbative allows a quantitative extrapolation to higher energies. Moreover, there are constraints from precision tests on generic effects of strong interactions at a scale as low as a TeV. The onset of a perturbative regime can be seen as follows. One can compute the amplitude $A(W_L W_L \to W_L W_L)$ for the scattering of the longitudinal component of the W boson. The latter are nothing but the Goldstone bosons originally sitting in the Higgs doublet together with the physical Higgs boson whose mass we are trying to constrain. The expansion in partial waves gives $A = \sum_l a_l A_l$, where a_l are partial wave amplitudes. The s-wave amplitude is bound by unitarity to be $|a_0| \leq 1$. If the physical Higgs is not taken into account, a tree level calculation, involving the gauge boson self couplings, gives $a_0 \sim s/(16\pi v^2)$, where s is the center of mass squared energy. The unitarity bound would then be saturated for $s \approx (1.2\,\text{TeV})^2$, unless such a bad behaviour is cancelled by the diagrams involving the exchange of a Higgs lighter than $1.2\,\text{TeV}$. Since we know that unitarity is not violated, the apparent violation must be due to the failing of the tree level approximation, signaling in turn a strongly interacting regime where higher order perturbative corrections are as large as the lower order contributions.

A stronger, second constraint can be obtained by assuming that the SM holds and is stable and perturbative up to a scale Λ. This argument uses the fact that the Higgs coupling λ_H, as all the couplings in the Lagrangian, depends on the energy scale of the process in which it is involved. Under the assumption that the SM holds up to the scale Λ, the value of the Higgs coupling at any scale up to Λ can be calculated as a function of the value at the electroweak scale, i.e. of the Higgs mass. It turns out that a too large value of the Higgs mass would give rise to a steep raise of λ_H with the energy, leading to a Landau-pole, i.e. a non-perturbative regime, before the scale Λ, thus contradicting the initial hypotheses. We obtain this way an upper limit on the Higgs mass as a function of the scale Λ. On the other hand, a too small value of the Higgs mass would make λ_H negative before Λ, thus giving rise to an instability. The Higgs potential would in fact become deeply negative for values of the Higgs field larger than the scale at which λ_H becomes negative. The electroweak scale minimum we need, eq. (18), would therefore be at best metastable (in which case, assuming we happen to live in such a metastable vacuum, its lifetime should be larger than the life of the universe). All this leads to a lower limit on the Higgs mass as a function of the scale Λ. All in all, one gets a window for the Higgs mass m_h as a function of the scale Λ. Such a window is shown in Fig. 5 for $10^3\,\text{GeV} < \Lambda < 10^{19}\,\text{GeV}$ [6].

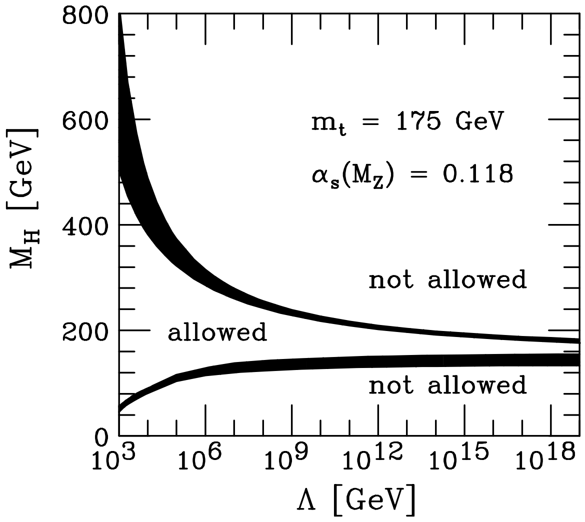


Figure 5: Allowed window for the Higgs mass as a function of the scale Λ up to which the SM is assumed to hold and be stable and perturbative.

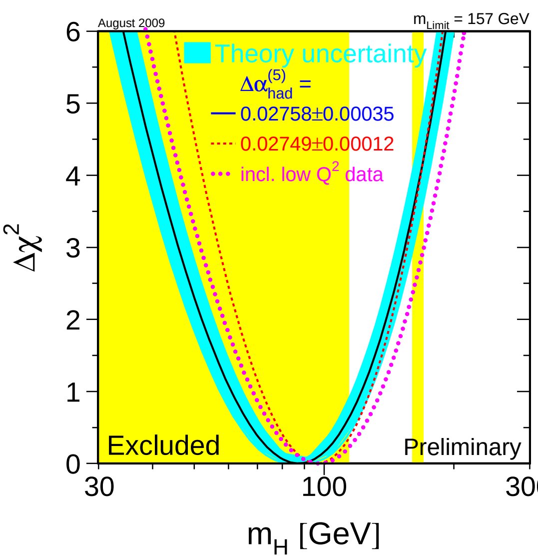


Figure 6: Direct (yellow exclusion regions) and indirect (blue band χ^2) experimental bounds on the Higgs mass [7].

Let us now come to the experimental constraints on the Higgs mass. The direct experimental limit from LEP is $m_h > 114\,\text{GeV}$ at 95% CL. Also, Tevatron has recently excluded the (160-170) GeV window. On top of those, there are indirect experimental bounds. The Higgs mass enters in fact (logarithmically) through loop corrections a number of observables that have been precisely measured at colliders, LEP in particular. A global fit of such precision observables as a function of the Higgs mass actually favors an Higgs mass in the region excluded by LEP, but values above the LEP bound do not give a bad fit. All in all, the fit favours a relatively light Higgs, $m_h < 163\,\text{GeV}$ at 95% CL. The results of the fit are summarized in the "blue band plot" shown in Fig. 6 [7]. The fit could be modified in the presence of new physics.

The existence of a scalar (Higgs) mass parameter gives rise to the so-called naturalness problem of the SM, whose solution has represented in the last decades one of the main guidelines for the theoretical quest for new physics beyond the SM. This issue will be addressed in the lectures by Alexei Gladyshev.

Let us now come to the analysis of the gauge sector of the SM. If the Higgs sector of the SM is the least known part of the SM (we do not even know if the Higgs really exists), the gauge sector is on the contrary the best known sector.

The gauge interactions of SM fermions have been tested with an accuracy up to the ‰ level, which is enough to probe the gauge sector at the loop level (perturbative loop corrections are typically of order $1/(4\pi)^2 \sim \mathcal{O}(\%)$). The fermion gauge interactions can be obtained from the covariant derivative. Let us consider electroweak interactions only, as strong interactions will be discussed elsewhere. The covariant derivative then assumes the form $D_\mu = \partial_\mu + igW^a_\mu T_a + ig' B_\mu Y$, where B_μ is the gauge vector associated to the hypercharge generator. The covariant derivative involves two independent gauge couplings, g and g', corresponding to the two irreducible factors of the electroweak group, $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$. The explicit form of the generators T_a and Y depends on the quantum numbers of the field on which they act. For example, when acting on the Higgs field, $D_\mu = \partial_\mu + igW^a_\mu(\sigma_a/2) + ig' B_\mu(1/2)$. The vector boson masses arise from the $(D_\mu\langle H\rangle)^*(D^\mu\langle H\rangle)$ term in the Higgs Lagrangian. To compute the vector boson masses, we observe that

$$D_\mu\langle H\rangle = \frac{iv}{2}\begin{pmatrix} g(W^1_\mu - iW^2_\mu) \\ gW^3_\mu - g'B_\mu \end{pmatrix} = \frac{iv}{2}\begin{pmatrix} \sqrt{2}gW^+_\mu \\ \sqrt{g^2+g'^2}Z_\mu \end{pmatrix}, \tag{21}$$

where we have defined

$$W^\pm_\mu \equiv \frac{W^1_\mu \mp iW^2_\mu}{\sqrt{2}}, \quad Z_\mu \equiv c_W W^3_\mu - s_W B_\mu, \tag{22}$$

in terms of the "Weinberg angle" θ_W defined by $\tan\theta_W = g'/g$, $0 \leq \theta_W \leq \pi/2$. The charged and neutral vector bosons $W^\pm_\mu$ and Z_μ turn then out to have definite mass. We have in fact $(D_\mu\langle H\rangle)^*(D^\mu\langle H\rangle) = M^2_W W^+_\mu W^{\mu-} + M^2_Z Z_\mu Z^\mu/2$, with the vector boson masses given by

$$M^2_W = \frac{g^2}{2}v^2, \quad M^2_Z = \frac{g^2+g'^2}{2}v^2 \quad \text{and} \quad \frac{G_F}{\sqrt{2}} = \frac{g^2}{8M^2_W} = \frac{1}{4v^2}, \tag{23}$$

determining the electroweak scale in terms of G_F. The fourth vector boson, $A_\mu = s_W W^3_\mu + c_W B_\mu$, the photon, does not get a mass term, as the corresponding generator is not broken. In order to be able to write the gauge interactions in terms of the vector bosons with definite masses, $W^\pm_\mu$, Z_μ, A_μ, it suffices to write the covariant derivative in terms of the latter:

$$D_\mu = \partial_\mu + i\frac{g}{\sqrt{2}}W^+_\mu T^+ + i\frac{g}{\sqrt{2}}W^-_\mu T^- + ieQA_\mu + i\frac{g}{c_W}(T_3 - s^2_W Q)Z_\mu, \tag{24}$$

where e is given by $e = gs_W = g'c_W = gg'/\sqrt{g^2+g'^2}$ and we have defined $T^\pm = T_1 \pm iT_2$. From the above expression of the covariant derivative we recover the electromagnetic interactions of the photon, proportional to the electric charge Q, and the charged current interactions of the $W^\pm$'s in eq. (9). New, "neutral current" interactions involving the Z boson are also predicted proportional to the coupling $(g/c_W)(T_3 - s_W^2 Q)$.

We note the relation

$$\rho \equiv \frac{M_W^2}{M_Z^2 \cos^2\theta_W} = 1 \quad \text{(tree level)}. \tag{25}$$

The above relation is not a general result. It depends on the fact that we have broken the electroweak symmetry by means of a $\mathrm{SU}(2)_L$ doublet and not, for example, a triplet. Indeed, the experimental verification of the above relation (where the Weinberg angle is independently measured) rules out significant contributions to spontaneous breaking by additional scalar triplets. The relation in eq. (25) receives small perturbative corrections and is related to a $\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$ "custodial" symmetry of the Higgs potential.

Let us spell out in greater detail the form of the gauge interactions of the fermions, the vector bosons, and the Higgs. The fermion gauge interactions come from the term

$$\overline{\Psi} i D_\mu \gamma^\mu \Psi = \overline{\Psi} i \partial_\mu \gamma^\mu \Psi - \left(\frac{g}{\sqrt{2}} j_c^\mu W_\mu^+ + \text{h.c.}\right) - \frac{g}{c_W} j_n^\mu Z_\mu - e j_{\text{em}}^\mu A_\mu, \tag{26}$$

where the charged and neutral currents j_c^μ and j_n^μ are given by

$$j_c^\mu = \overline{\nu_{iL}} \gamma^\mu e_{iL} + \overline{u_{iL}} \gamma^\mu d_{iL}, \quad j_n^\mu = \sum_f \overline{f_X} \gamma^\mu (T^3 - s_W^2 Q) f_X \quad (f = \nu_i, e_i, u_i, d_i,\ X = L, R). \tag{27}$$

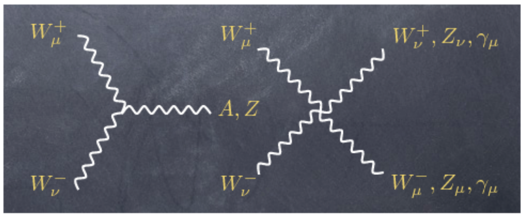


Figure 7: Vector boson gauge self-interactions.

The vector boson gauge self-interactions come from the "Yang-Mills" term:

$$-W_{\mu\nu}^a W^{\mu\nu a}/4,$$

where

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\epsilon_{abc} W_\mu^b W_\nu^c.$$

When expressed in terms of the vector mass eigenstates $W^\pm$, Z, A, the Yang-Mills term gives trilinear and quartic interactions, as shown in Fig. 7. The existence of the $\gamma W^+ W^-$ and $Z W^+ W^-$ vertexes has been experimentally established at LEP. Fig. 8 shows the measurement of the cross Section of $e^+e^- \to W^+W^-$ and the corresponding theoretical predictions obtained taking into account the exchange of a neutrino only (a), adding the exchange of a photon through the $\gamma W^+ W^-$ interactions (a+b), and further adding the exchange of a Z through the $Z W^+ W^-$ vertex (a+b+c) [7].

Finally, the Higgs gauge interactions are given in unitary gauge by

$$\left(\sqrt{2}\frac{h}{v} + \frac{h^2}{2v^2}\right)\left(M_W^2 W_\mu^+ W^{\mu-} + \frac{1}{2} M_Z^2 Z_\mu Z^\mu\right). \tag{28}$$

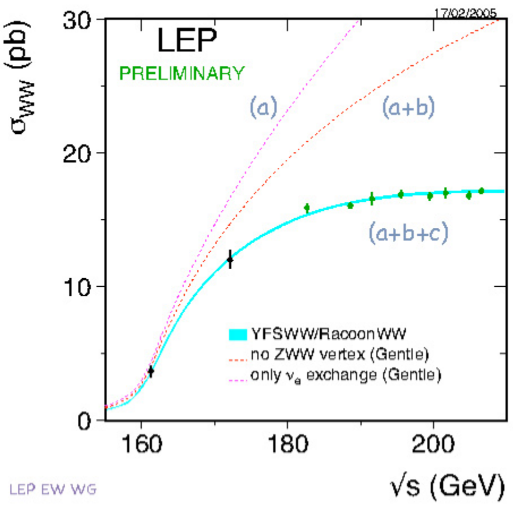


Figure 8: Measurement of the $e^+e^- \to W^+W^-$ cross section as a function of the center of mass energy and comparison with different theoretical predictions.

Having discussed to some extent the gauge and Higgs sectors of the SM, we are only left with the Yukawa sector. The latter will be covered by the lectures by Yossi Nir.

We close these lectures with a summary of the reasons why many of us believe that the SM is not the ultimate theory of everything. Although the SM has been tested with great success, especially in its gauge sector, but also in its flavour sector, there exist a few experimental evidences that are not accounted for by the SM: the existence of dark matter, the baryon asymmetry of the universe and, last but not least, gravity itself. There are then experimental evidences that, although strictly speaking do not contradict the SM, represent strong hint for physics beyond the SM: the peculiar structure of the SM gauge quantum numbers in Table 1, and neutrino masses. The SM quantum numbers can be nicely understood in terms of grand-unified theories, which also lead to the successful, precise prediction of the strong coupling within supersymmetric models. Neutrino masses can be incorporated in the SM by means of an effective interaction of two lepton doublets and two Higgses. Still, the existence of an effective, non-renormalizable interaction represents a strong hint for new physics arising at a higher scale. There are then a number of theoretical puzzles that do not represent a clear indication for new physics but we would be very happy to understand in terms of physics beyond the SM: the smallness of the electroweak scale compared to the Planck scale, family replication, the existence of small Yukawa couplings and the peculiar pattern of fermion masses and mixings. Finally, the SM has a number of theoretical problems. The naturalness/unitarity problem, related to the stability of the Higgs mass with respect to radiative corrections in the presence of a new high scale (of which we have at least an incontrovertible example: the Planck scale); the similar (from a QFT point of view) problem of the smallness of the cosmological constant. The strong CP-problem.

For all the above reasons we believe that the SM is not the end of the story. I wish the younger generations attending this school to have the opportunity to witness and hopefully play an important role in the many developments to come.

Acknowledgments

I would like to thank all the organizers, and Dima in particular, for introducing me to such a special school; the students, the pancake makers, and all the people and fowls involved for the magic atmosphere they contributed to create; and... a hero guy authority in SM.

Bibliography

[1] C. Amsler *et al.* [Particle Data Group], Phys. Lett. B **667** (2008) 1.

[2] F. Englert and R. Brout, Phys. Rev. Lett. **13** (1964) 321. P. W. Higgs, Phys. Rev. Lett. **13** (1964) 508. G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, Phys. Rev. Lett. **13** (1964) 585.

[3] S. L. Glashow, Nucl. Phys. **22** (1961) 579. S. Weinberg, Phys. Rev. Lett. **19** (1967) 1264. A. Salam, *Originally printed in "Svartholm: Elementary Particle Theory, Proceedings of the Nobel Symposium Held 1968 at Lerum, Sweden", Stockholm 1968, 367-377*

[4] E. Fermi, Z. Phys. **88** (1934) 161. G. Gamow and E. Teller, Phys. Rev. **49** (1936) 895. E. C. G. Sudarshan and R. e. Marshak, Phys. Rev. **109** (1958) 1860.

[5] Y. Nambu, Phys. Rev. **117** (1960) 648. J. Goldstone, Nuovo Cim. **19** (1961) 154. J. Goldstone, A. Salam and S. Weinberg, Phys. Rev. **127** (1962) 965.

[6] T. Hambye and K. Riesselmann, Phys. Rev. D **55** (1997) 7255 [arXiv:hep-ph/9610272].

[7] http://lepewwg.web.cern.ch/

The Standard Model of Particle Physics

Uwe-Jens Wiese
Institute for Theoretical Physics
University of Bern

February 19, 2018

Contents

Chapter 1

Introduction

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

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

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

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

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

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

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

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

Part I

FUNDAMENTAL CONCEPTS

Chapter 2

Concepts of Quantum Field Theory and the Standard Model

These lectures are an introduction to the Standard Model of elementary particle physics — the *relativistic quantum field theory* that summarizes all we know today about the fundamental structure of matter, forces, and symmetries. The Standard Model is a *gauge theory* that describes the *strong, weak, and electromagnetic* interactions of *Higgs particles, leptons, and quarks* mediated by *gluons, W- and Z-bosons, and photons.* In addition, it describes the direct (not gauge-boson-mediated) self-couplings of the Higgs field as well as the Yukawa couplings of the Higgs field to leptons and quarks. In this Chapter we discuss fundamental concepts and basic principles of field theory in order to pave the way for a systematic exposition of the subject in the rest of the lectures. In particular, we emphasize the fundamental roles of *locality*, *symmetries*, and *hierarchies of energy scales.* We also provide an overview of the historical development of particle physics and quantum field theory.

2.1 Point Particles versus Fields at the Classical Level

Theoretical physics in the modern sense was initiated by Sir Isaac Newton who published his "Philiosophiae Naturalis Principia Mathematica" in 1687. This spectacular eruption of genius provides us with the description of *classical point particle mechanics*, in terms of ordinary differential equations for the position vectors $\vec{x}_a(t)$ of the individual particles ($a \in \{1, 2, \dots, N\}$) as functions of time t. Classical mechanics is *local in time*, because Newton's equation contains infinitesimal time-derivatives $d\vec{x}_a(t)/dt^2$, but no finite time-differences $t - t'$. On the other hand, Newtonian mechanics is *non-local in space*, because the finite distances $|\vec{x}_a - \vec{x}_b|$ between different particles determine instantaneous forces, including Newtonian gravity. Hence, in classical mechanics there are fundamental differences between space and time. In point particle theories the fundamental degrees of freedom, which are the particle positions $\vec{x}_a(t)$, are mobile: they move around in space. As a consequence, at almost all points space is empty, *i.e.* nothing is happening there, except if a point particle occupies that position.

The fundamental degrees of freedom of a field theory, namely the field values $\Phi(\vec{x}, t)$ are immobile, because they are attached to a given space point $\vec{x}$ at all times t. In this case, it is the field value Φ — and not the position $\vec{x}$ — which changes as a function of time. In a field theory, space plays a very different role than in point particle mechanics. In particular, it is not empty, because field degrees of freedom exist at all points $\vec{x}$ at all times t. Fluid dynamics is an example of a nonrelativistic classical field theory in which the mass density enters as a scalar field $\Phi(\vec{x}, t)$. The classical field equations are partial differential equations (involving both space- and time-derivatives of $\Phi(\vec{x}, t)$) which determine the evolution of the fields. Hence, in contrast to point particle theories, field theories are *local in both space and time.*

The most fundamental classical field theory is James Clark Maxwell's electrodynamics of electric and magnetic fields $\vec{E}(\vec{x}, t)$ and $\vec{B}(\vec{x}, t)$, which was published in 1864. In fact, this theory (in quantized form) is an integral part of the Standard Model. Although this was not known at the time, Maxwell's electrodynamics is a relativistic classical field theory, which is invariant against space-time translations and rotations forming the *Poincaré symmetry group*. Newton's point particle mechanics, on the other hand, is invariant under Galileian instead of Lorentz boosts. Thus, it is nonrelativistic and hence inconsistent with the relativistic space-time underlying Maxwell's electrodynamics.

Albert Einstein's *special theory of relativity* from 1905 modified Newton's point particle mechanics in such a way that it becomes Poincaré invariant. Indeed, in the framework of Einstein's special relativity, charged point particles can interact with classical electromagnetic fields in a Poincaré invariant manner. On the other hand, relativistic point particles cannot interact directly with each other, and thus necessarily remain free in the absence of a mediating electromagnetic field. This follows from Heinrich Leutwyler's *non-interaction theorem* for relativistic systems of N point particles [**?**], which extended an earlier study of the $N = 2$ case [**?**]. Indeed, in the relativistic Standard Model quantum field theory the point particle concept is completely abandoned and all "particles" are in fact just field excitations, which Frank Wilczek sometimes calls "wavicles". This is a very useful distinction which allows us to avoid confusions that might otherwise arise quite easily. In particular, while a Newtonian point particle has a completely well-defined position $\vec{x}_a$, a wavicle does not.

2.2 Particles versus Waves in Quantum Theory

Quantum mechanics (as formulated by Werner Heisenberg in 192? and by Erwin Schrödinger in 192?) applies the basic principles of quantum theory — namely unitarity which implies the conservation of probability — to Newton's point particles. As a consequence, the particle positions $\vec{x}_a$ (which are still conceptually completely well-defined) are then affected by quantum uncertainty. This is described in terms of a wave function $\Psi(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N, t)$, which obeys the nonrelativistic Schrödinger equation — a partial differential equation containing derivatives with respect to time as well as with respect to the N particle positions $\vec{x}_a$. It is important to note that (unlike $\Phi(\vec{x}, t)$) $\Psi(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N, t)$ is not a field in space-time, but just a time-dependent complex function over the N-particle configuration space $(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N)$.

A time-dependent state in a quantum field theory, on the other hand, can be described by a complex-valued wave functional $\Psi[\Phi(\vec{x}), t]$, which depends on the field configuration $\Phi(\vec{x})$, and obeys a functional Schrödinger equation.

When one discusses quantum mechanical *double-slit experiments*, one says that the resulting interference pattern is a manifestation of the wave properties of quantum particles. This does not mean that such a particle is a quantized wave excitation of a field. It is just a point particle with a conceptually completely well defined position, which is, however, affected by quantum uncertainty. In particular, as long as the position of the particle is not measured, it can go through both slits simultaneously, until it hits the detection screen which registers its (unambiguously defined) position. Only after repeating this single-particle experiment a large number of times, the detected positions of the individual particles give rise to an emerging interference pattern. In the context of quantum mechanics, particle-wave duality just means that point particles are described by a quantum mechanical wave function $\Psi(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N, t)$.

When a classical electromagnetic wave is diffracted at a double slit, it shows an interference pattern for very different reasons. As a field excitation, the wave exists simultaneously at all points in a region of space. In fact, unlike a point particle, it does not even have a well-defined position. In contrast to the experiment with quantum mechanical point particles, the interference pattern arises immediately as soon as the classical wave reaches the detection screen. When one repeats this experiment at the quantum level with individual photons, the interference pattern again emerges only after the experiment has been repeated a large number of times. The "particle" character of photons is usually emphasized in the context of the Compton effect. However, while we may be used to thinking of an electron as a point particle with position $\vec{x}_a$ (perhaps affected by quantum uncertainty), we should definitely not think about a photon in a similar way. As a quantized wave excitation of the electromagnetic field, a photon does not even have a well-defined position in space. What do we then mean when we talk about the photon as a "particle"? Unfortunately, in our casual language the term "particle" is associated with the idea of a point-like object, which is not what a photon is like. Frank Wilczek's term "wavicle" serves its purpose when it prevents us from thinking of a photon as a tiny billiard ball. At the end, only mathematics provides an appropriate and accurate description of "particles" like the photon. In the mathematics of quantum field theory, particle-wave duality reduces to the fact that "particles" actually are "wavicles", *i.e.* quantized wave excitations of fields.

When Paul Adrien Maurice Dirac discovered his relativistic equation for the electron in 1928, the 4-component Dirac spinor was initially interpreted as the wave function of an electron or positron with spin up or down. However, due to electron-positron pair creation, it turned out that the Dirac equation does not have a consistent single-particle interpretation. In fact, the Dirac spinor is not a wave function at all, but a fermionic field whose quantized wave excitations manifest themselves as electrons and positrons. In other words, not only photons but all elementary "particles" are, in fact, wavicles. When the Dirac field is coupled to the electromagnetic field one arrives at Quantum Electrodynamics (QED), whose construction was pioneered by Freeman Dyson, Richard Feynman, Julian Schwinger, and Sin-Itiro Tomonaga. QED is an integral part of the Standard Model in which all elementary "particles", including quarks, leptons, and Higgs particles, are quantized wave excitations of the corresponding quark, lepton, and Higgs fields. Unlike point particles, quark, lepton, and Higgs fields can interact directly in a relativistic manner, even without the mediation by gauge fields.

Although in the Standard Model all "particles" are, in fact, wavicles, one often reads that quarks or electrons are "point-like" objects. What can this possibly mean for a wavicle that does not even have a well-defined position in space? Again, this is a deficiency of our casual language, which is properly resolved by the unambiguous mathematics of quantum field theory. What the above statement actually means is that even the highest energy experiments have, at least until now, not revealed any substructure of quarks or electrons, *i.e.* they seem truly elementary. The same is not true for protons or neutrons, which actually consist of quarks and gluons. Interestingly, while being "point-like" in the above sense, an electron is at the same time infinitely extended. This is because electrons are charged "particles" which are surrounded by a Coulomb field that extends to infinity. In reality, this field is usually screened by other positive charges in the vicinity of the electron.

This discussion should have convinced the reader that particle physics is not at all concerned with point particles. Perhaps it should better be called "wavicle physics". However, as long as we are aware that our casual language is not sufficiently precise in this respect, the nomenclature is secondary. In the mathematics of quantum field theory, all "particles" are indeed quantized waves.

2.3 Classical and Quantum Gauge Fields

Although it also contains non-gauge-field-mediated couplings between quark, lepton, and Higgs fields, in the Standard Model gauge fields play a very important role, because they mediate the fundamental strong, weak, and electromagnetic interactions. While the classical Maxwell equations can be expressed entirely in terms of the electromagnetic field strengths $\vec{E}$ and $\vec{B}$, which form the field strength tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, in relativistic quantum field theory gauge fields are described by the vector potential A_μ. Even in the nonrelativistic quantum mechanics of a charged point particle, an external magnetic field $\vec{B} = \vec{\nabla} \times \vec{A}$ enters the Schrödinger equation via the vector potential $\vec{A}$, which forms a *covariant derivative* together with the momentum operator. In particular, the Aharonov-Bohm effect is naturally expressed through a line integral of the vector potential. While the field strength $F_{\mu\nu}$ is gauge invariant and thus physical, the vector potential can be gauge transformed to $A'_\mu = A_\mu + \partial_\mu \varphi$, where $\varphi(\vec{x}, t)$ is an arbitrary local gauge transformation function of space and time. When we work with vector potentials, we use redundant gauge variant variables to describe gauge invariant physical observables. While this is a matter of choice in classical theories, it seems unavoidable in quantum theories. In particular, in the quantum mechanics of a charged point particle, the complex phase ambiguity of the wave function turns into a local gauge freedom. Similarly, in quantum field theory the complex phase of a Dirac spinor is gauge variant, but it can be combined with the gauge variant vector potential to form the gauge invariant QED Lagrangian. Gauge invariance is a local symmetry which must be maintained exactly in order to guarantee that no unphysical effects can arise due to the redundant gauge variant variables.

Since a local gauge symmetry just reflects a redundancy in our theoretical description of the gauge invariant physics, it has different physical consequences than a global symmetry. Both for gauge and for global symmetries, the Hamiltonian of the theory is invariant under symmetry operations. In case of a global symmetry (at least in the absence of spontaneous symmetry breaking), this implies that physical states belong to (in general nontrivial) irreducible representations of the symmetry group. As a consequence, there are degeneracies in the spectrum whenever an irreducible representation is more than 1-dimensional. In case of a gauge symmetry, on the other hand, all physical states are gauge invariant, *i.e.* they belong to a trivial 1-dimensional representation of the local gauge group. Hence, gauge symmetries do not give rise to degeneracies in the spectrum of physical states. Indeed the gauge variant eigenstates of the gauge invariant Hamiltonian are exiled from the physical Hilbert space, by imposing the Gauss law as a constraint on physical states.

2.4 Ultraviolet Divergences, Regularization, and Renormalization

Field theories have a fixed number of fundamental field degrees of freedom attached to each point in space. In continuous space, the total number of field degrees of freedom is thus uncountably large. While this is no problem in classical field theory, where the solutions of the field equations are smooth functions of space and time, quantum fields undergo violent fluctuations which give rise to ultraviolet divergences. In order to obtain meaningful finite answers for physical quantities, quantum fields must be regularized by introducing an ultraviolet cut-off. This is necessary because, most likely, quantum fields in continuous space are ultimately not the correct degrees of freedom that Nature is built from at ultra-short distances of the order of the Planck length $l_{\text{Planck}} \approx 10^{-35}$ m. The corresponding energy scale is the Planck scale $M_{\text{Planck}} \approx 10^{16}$ TeV, at which gravity, which is extremely weak at low energies, becomes strongly coupled. Although string theory provides a promising framework for its formulation, an established non-perturbative theory of quantum gravity, which is valid all the way up to the Planck scale, currently does not exist. Fortunately, we need not know the ultimate high-energy *theory of everything* before we can address the physics in the TeV energy regime that is accessible to present day experiments, in which the Standard Model has been tested with great scrutiny. Whether there are strings, some tiny wheels turning around at the Planck scale, or some other truly fundamental degrees of freedom, the currently accessible low-energy physics is insensitive to those details.

In order to mimic the effects of the unknown ultimate ultra-short distance degrees of freedom, one can introduce an ultraviolet cut-off in many different ways. It is, however, important that the cut-off procedure does not violate any gauge symmetries, because otherwise unphysical redundant variables would contaminate physical results. In perturbation theory, the most efficient way to introduce a gauge invariant cut-off is *dimensional regularization*, *i.e.* analytic continuation in the space-time dimension away from 4. Beyond perturbation theory, the *lattice regularization*, in which space-time is replaced by a 4-dimensional hyper-cubic grid of discrete lattice points, provides a natural cut-off that again allows us to maintain gauge invariance. In this case, the lattice spacing a, *i.e.* the distance between nearest-neighbor lattice points, acts as an ultraviolet cut-off. Unlike in continuous space-time, in lattice field theory the number of field degrees of freedom becomes countable, which removes the divergences in physical observables. Still, in order to obtain meaningful physical results, one must take the continuum limit $a \to 0$. This is achieved by tuning the coupling constants in the Lagrangian in such a way that the long-distance continuum physics becomes insensitive to the lattice spacing. This process is known as renormalization.

2.5 The Standard Model: Renormalizability, Triviality, and Incorporation of Gravity

The gauge, Higgs, lepton, and quark fields of the Standard Model all have specific gauge transformation properties. They also transform appropriately under the space-time transformations of the Poincaré group. The Lagrangian of the Standard Model comprises all terms that are gauge as well as Poincaré invariant combinations of fields. It is important to note that the Standard Model is renormalizable, *i.e.* a finite number of terms in the Lagrangian is sufficient to remove the ultraviolet divergences. In particular, terms with coupling constants of negative mass dimension are irrelevant and need not be included in the Standard Model. Its renormalizability implies that the Standard Model could, at least in principle, be valid up to arbitrarily high energy scales. However, there is a caveat: the issue of "triviality". There is overwhelming evidence, but no rigorous proof, that the Higgs sector of the Standard Model becomes non-interacting (and thus trivial) when one removes the cut-off all the way to infinity.

While renormalizability implies that the Standard Model physics is insensitive to the ultraviolet cut-off, it is not necessarily physically meaningful to send the cut-off to infinity. In particular, one would expect that, at some energy scale, either near or high above the TeV scale accessible to current experiments, new physics beyond the Standard Model could be discovered. In that case, the scale Λ at which new physics arises would provide a physical cut-off for the Standard Model, which would no longer provide an accurate description of the physics above that energy scale. The Standard Model would then still remain a consistent low-energy effective field theory. However, as one reaches higher and higher energies approaching Λ, more and more non-renormalizable terms with negative mass dimension (suppressed by inverse powers of Λ) would have to be added to the effective Lagrangian.

Even in the absence of new physics close to currently accessible energy scales, the triviality of the Standard Model is a rather academic issue, because the Planck scale already provides a finite (yet extremely high) energy scale at which the Standard Model must necessarily be replaced by a more complete theory that should include non-perturbative quantum gravity. While gravity is usually not considered as belonging to the Standard Model, it can be incorporated perturbatively as a low-energy effective theory, provided that Poincaré invariance is maintained as an exact symmetry. This is necessary because in Einstein's theory of gravity, *i.e.* in general relativity, global Poincaré invariance is promoted to a (necessarily exact) gauge symmetry. In contrast to some claims in the literature, it is not true that gravity resists quantization in the context of quantum field theory. While Einstein gravity is not renormalizable, *i.e.* at higher and higher energies more and more terms enter the Lagrangian, it can be consistently quantized as a low-energy effective field theory. This theory is expected to break down at the Planck scale, where gravity becomes strongly coupled.

2.6 Fundamental Standard Model Parameters

The Standard Model Lagrangian contains a large number of free parameters, whose values can only be determined by comparison with experiments. Remarkably, in the minimal version of the Standard Model there is only one dimensionful parameter, which determines the vacuum value $v = 246$ GeV of the Higgs field as well as the Higgs boson mass. The masses of the heavy $W^{\pm}$ and Z^0 gauge bosons, which mediate the weak interaction, are given by $M_W = gv$ and $M_Z = \sqrt{g^2 + g'^2}v$, where g and g' are the dimensionless gauge coupling constants associated with the Standard Model gauge groups $SU(2)_L$ and $U(1)_Y$, respectively. The strong interactions between quarks and gluons are described by Quantum Chromodynamics (QCD) — an $SU(3)_c$ color gauge theory — which is another integral part of the Standard Model. Since scale invariance is broken by quantum effects, by dimensional transmutation the dimensionless $SU(3)_c$ gauge coupling g_s is traded for the dimensionful QCD scale $\Lambda_{\mathrm{QCD}} = 0.260(40)$ GeV. Strongly interacting particles, including protons, neutrons, and other hadrons, receive the dominant portion of their masses from the strong interaction energy of quarks and gluons, which is proportional to Λ_{QCD}, and only a small fraction of their masses is due to the non-zero quark masses. The masses of quarks, $m_q = y_q v$, and of leptons, $m_l = y_l v$, are products of v with the dimensionless Yukawa couplings y_q and y_l. Quarks and leptons arise in three *generations* with the same quantum numbers, but with different masses. The mixing angles between the quark or lepton fields of the different generations are additional fundamental Standard Model parameters, whose values can only be determined experimentally.

In the original minimal version of the Standard Model the neutrinos were massless particles, because only left-handed neutrino fields were considered. Since the discovery of neutrino oscillations, it is clear that (at least some) neutrinos must have a non-zero mass. This naturally suggests to extend the minimal Standard Model by adding right-handed neutrino fields. In this way further dimensionful parameters, the Majorana masses M_{ν_R} of the right-handed neutrinos, enter the Standard Model Lagrangian. A mass mixing mechanism — also known as the see-saw mechanism — leads to small neutrino masses, provided that $M_{\nu_R} \gg v$. The parameters M_{ν_R} set the scale

Λ at which new physics beyond the minimal Standard Model arises. The low-energy effects of this new physics — in particular, the non-zero neutrino masses — can also be described correctly by adding non-renormalizable terms to the minimal Standard Model Lagrangian, which are suppressed by the inverse of the scale $\Lambda \approx M_{\nu_R}$.

In view of the large number (of about 25) free parameters, one may expect that there could be an even more fundamental structure beyond the Standard Model that would allow us to understand the origin of its free parameters. Ultimately, the Standard Model will definitely break down at the Planck scale, when non-perturbative quantum gravity comes into play. The minimal Standard Model has already been extended by new physics associated with the Majorana neutrino mass scale M_{ν_R}, and there is no reason to believe that no further extensions will be necessary before we reach M_{Planck}. The extensions might include techni-color theories, supersymmetric theories, grand unified theories (GUT), or other structures that have been a subject of intense theoretical investigation. At the time of the writing of these notes, there is no conclusive experimental evidence for physics beyond the Standard Model. There is evidence for *dark matter*, which might be of supersymmetric origin, but could also simply be related to right-handed Majorana neutrinos. The idea of *cosmic inflation* suggests that there could be an *inflaton field*. Then there is evidence for *dark energy — i.e. vacuum energy* — which might arise as dynamical *quintessence* or as a static *cosmological constant* Λ_c. The latter is just a free low-energy parameter of Einstein gravity, another being Newton's constant G, which determines the Planck scale $M_{\text{Planck}} = \sqrt{\hbar c/G}$. When we include perturbative quantum gravity as well as right-handed neutrino fields in the Standard Model, we can currently not exclude that it might be valid all the way up to the Planck scale.

2.7 Hierarchies of Scales and Approximate Global Symmetries

In the minimal Standard Model extended by perturbative quantum gravity we encounter four dimensionful parameters: the Planck scale, $M_{\text{Planck}} \approx 10^{19}$ GeV, derived from Newton's constant, which determines the strength of gravity, the vacuum expectation value $v \approx 10^{-17} M_{\text{Planck}}$ of the Higgs field, the QCD scale $\Lambda_{\text{QCD}} \approx 10^{-20} M_{\text{Planck}}$, and the cosmological constant $\Lambda_{\text{c}}^{1/4} \approx 10^{-30} M_{\text{Planck}}$. Why are these scales so vastly different, or, in other words, what is the origin of these hierarchies of energy scales? Since, according to our present understanding, these scales are free parameters, answering these questions requires to go beyond the Standard Model or perturbative quantum gravity. Staying within the framework of these theories, one can at least ask whether the hierarchies may arise naturally. At first glance, it may seem unnatural that the QCD scale is so much smaller than the Planck scale. However, QCD's property of asymptotic freedom provides an explanation for this hierarchy, because, without unnatural fine-tuning of the strong coupling constant g_s, Λ_{QCD} is exponentially suppressed with respect to the ultraviolet cut-off, which we may identify with M_{Planck}.

The same is not true for the hierarchy between the electroweak scale and the Planck scale. The puzzle to understand why $v \ll M_{\text{Planck}}$ is known as the *hierarchy problem*, which has no natural solution within the Standard Model because the self-coupling of the Higgs field is not asymptotically free. Potential solutions of the hierarchy problem may be associated with new physics beyond the Standard Model, such as, for example, supersymmetric or techni-color models. Despite intensive investigations, at the time of the writing of these notes there is no experimental evidence for these ideas. In a non-perturbative context supersymmetry may, in fact, be unnatural, because the construction of the symmetry itself may require fine-tuning.

The fact that $\Lambda_{\text{c}} \ll M_{\text{Planck}}$ confronts us with the cosmological constant problem — the most severe hierarchy problem in all of physics. If the correct theory of non-perturbative quantum gravity would have a property like asymptotic freedom, one may speculate that the cosmological constant problem might find a natural explanation. Alternatively, one may invoke the *anthropic principle*. One then relates the value of Λ_c to the fact of our own existence. Alternative Universes with a larger negative or positive cosmological constant would either collapse or expand very quickly. In these cases, it seems unlikely that intelligent life could evolve. The idea of eternal cosmic inflation actually provides us with an incredible number of different Universes, forming a very large Multiverse. If the Multiverse indeed exists, which is a matter of speculation, we can only evolve in a pocket Universe with hospitable hierarchies of energy scales. The anthropic principle should, however, be invoked only as a last resort, when all alternative explanations fail. In particular, the somewhat cheap anthropic-principle-based explanations of various hierarchies should not prevent us from thinking hard about everything that can possibly be understood without invoking this principle.

The Standard Model provides us with even more hierarchy puzzles. While the dimensionless Yukawa coupling y_t of the heavy top quark is of order 1, such that the mass of the top quark $m_t = y_t v = 174$ GeV is near the electroweak scale v, the Yukawa couplings of the light up and down quarks are much smaller, $y_u, y_d \approx 10^{-5}$, such that $m_u, m_d \ll \Lambda_{\mathrm{QCD}}$. The hierarchy between the masses of the light quarks and the QCD scale gives rise to an approximate global $SU(2)_L \times SU(2)_R$ chiral symmetry. Its $SU(2)_{L=R}$ isospin subgroup manifests itself in the hadron spectrum and "explains" why proton and neutron have almost the same mass. However, this is a proper explanation only if we take the hierarchy $m_u, m_d \ll \Lambda_{\mathrm{QCD}}$ for granted. However, since we don't understand the origin of the experimental values of the quark masses, we should admit that the approximate isospin symmetry and thus the almost degenerate proton and neutron masses appear just as an "accident". As intelligent beings, we recognize the symmetry (although we may not understand its origin) and utilize it to simplify our theoretical investigations.

2.8 Local and Global Symmetries

As we have discussed in Section 2.3, local symmetries — *i.e.* gauge symmetries — must be exact in order to prevent unphysical effects of the redundant gauge variables. This includes Poincaré symmetry, which is promoted to a gauge symmetry in the context of general relativity. Gauge invariance is very restrictive and, in combination with renormalizability, implies large predictive power, with only one free parameter — the gauge coupling constant associated with the corresponding gauge group. Other non-gauge-mediated interactions, as, for example, the Yukawa couplings between Higgs and quark or lepton fields, give rise to a much larger number of free parameters and thus restrict the predictive power of the theory.

In contrast to gauge symmetries, global symmetries such as isospin are in general only approximate and result from an (often not understood) hierarchy of energy scales. For example, the discrete symmetries of charge conjugation C and parity P are broken by the weak but not by the electromagnetic and strong interactions. Due to the hierarchy $\Lambda_{\mathrm{QCD}} \ll v$, which "explains" the weakness of the W- and Z-boson-mediated interactions (but whose origin is again not understood), C- or P-violating processes are relatively rare. In the Standard Model the origin of C- and P-violation is the chiral nature of the theory — the fact that left- and right-handed quark or lepton fields transform differently under $SU(2)_L \times U(1)_Y$ gauge transformations. This is characteristic of a *chiral gauge theory*. While interactions between gauge and matter fields may break C and P, they leave the combined symmetry CP intact.

In the Standard Model, CP-violating processes arise only due to mixing between the three generations of quarks and leptons, and they are hence even rarer. It is an open question whether these sources of CP violation are sufficiently strong to explain the observed *baryon asymmetry* between matter and anti-matter in the Universe. It is still a puzzle — known as the strong CP problem — why the self-interactions of the gluons respect CP symmetry, *i.e.* why the experimental value of the QCD vacuum angle θ is compatible with zero. A potential explanation beyond the Standard Model (which still awaits experimental confirmation) is related to an approximate $U(1)_{\mathrm{PQ}}$ Peccei-Quinn symmetry, which would be associated with a new light particle — the axion.

Remarkably, as a result of the *CPT theorem*, the combination CPT of CP with time-reversal T is an exact symmetry of any relativistic field theory. In fact, the CPT symmetry is indirectly protected by the necessarily exact Poincaré symmetry, which acts as the gauge symmetry of general relativity.

Exact global symmetries other than CPT are, however, suspicious. In fact, they should either be gauged or explicitly broken. In the minimal version of the Standard Model without right-handed neutrino fields, the difference between baryon and lepton number $B - L$ is an exact global symmetry. In the $SO(10)$ GUT extension of the Standard Model, $U(1)_{B-L}$ is indeed gauged and appears as a subgroup of the $SO(10)$ gauge group. In the extended Standard Model with additional right-handed neutrino fields only, on the other hand, the global $U(1)_{B-L}$ symmetry is explicitly broken by Majorana mass terms. Fermion number conservation modulo 2 then still remains as an exact global symmetry. However, just as CPT, this symmetry automatically follows from Poincaré invariance.

2.9 Explicit versus Spontaneous Symmetry Breaking

As we just discussed, gauge symmetries must be exact, while global symmetries are in general only approximate. A simple source of explicit global symmetry breaking are non-invariant terms in the Lagrangian. A typical example is the $SU(2)_L \times SU(2)_R$ chiral symmetry of QCD, which is explicitly broken by the non-zero Yukawa couplings between the light up and down quarks and the Higgs field.

QCD with massless up and down quarks, on the other hand, has an exact chiral symmetry. Interestingly, this symmetry does not manifest itself directly in the QCD spectrum, because it is *spontaneously broken.* This means that, despite the fact that the Hamiltonian of massless QCD is invariant against $SU(2)_L \times SU(2)_R$ chiral symmetry transformations, its ground state is not. In fact, there is a continuous family of degenerate vacuum states of massless QCD, which are related to each other by chiral transformations. In the process of spontaneous symmetry breaking, one of these ground states is selected spontaneously. This state is still invariant against transformations in the unbroken $SU(2)_{L=R}$ isospin subgroup of $SU(2)_L \times SU(2)_R$. Small fluctuations around the spontaneously chosen vacuum state cost energy in proportion to the magnitude of their momentum, and thus manifest themselves as massless particles — known as *Goldstone bosons.* As a consequence of spontaneous chiral symmetry breaking, which reduces $SU(2)_L \times SU(2)_R$ to its unbroken $SU(2)_{L=R}$ isospin subgroup, there are three massless Goldstone bosons — the charged and neutral pions π^+, π^0, and π^-. In the real world with non-zero up and down quark masses, chiral symmetry is not only spontaneously but, in addition, also explicitly broken. As a result, the pions turn into light (but no longer massless) pseudo-Goldstone bosons, whose squared masses are proportional to the product of the quark masses and the chiral order parameter, which is proportional to Λ_{QCD}^3.

The Higgs sector of the Standard Model also has an $SU(2)_L \times SU(2)_R$ symmetry. However, unlike in QCD, its $SU(2)_L \times U(1)_Y$ subgroup is gauged and must hence be an exact symmetry. Since a gauge symmetry just reflects a redundancy in our theoretical description, it cannot break spontaneously in the same way as a global symmetry. When one gauges a spontaneously broken global symmetry, one induces the *Higgs mechanism* in which the gauge field picks up a mass. The previously massless Goldstone degrees of freedom are then incorporated as longitudinal degrees of freedom of the massive gauge bosons. One says that the gauge bosons "eat" the Goldstone bosons and become massive. This is indeed how the electroweak gauge bosons $W^{\pm}$ and Z^0 pick up their masses. Techni-color extensions of the Standard Model mimic QCD at the electroweak scale. The $W^{\pm}$ and Z^0 then become massive because they "eat" the massless techni-pions $\Pi^{\pm}$ and Π^0. This would indeed solve the hierarchy problem $v \ll M_{\text{Planck}}$, because the techni-chiral order parameter, which replaces v, is proportional to Λ_{TC} (the techni-color analog of Λ_{QCD}) which is naturally much smaller than M_{Planck} due to asymptotic freedom.

2.10 Anomalies in Local and Global Symmetries

A more subtle form of explicit symmetry breaking does not manifest itself in the Lagrangian, because it affects only the quantum but not the classical theory. Whenever quantum effects destroy a symmetry that is exact at the classical level, one speaks of an *anomaly*. Theories affected by gauge anomalies are mathematically and physically inconsistent, because unphysical redundant gauge variables then contaminate physical observables via quantum effects. Gauge anomalies must therefore be canceled. Gauge *anomaly cancellation* imposes severe constraints on chiral gauge theories including the Standard Model. For example, as a consequence of the cancellation of Witten's so-called *global anomaly*, which would destroy the $SU(2)_L$ gauge symmetry of the Standard model, the number of quark colors N_c (which is 3 in the real world) must be an odd number. In addition, anomaly cancellation has important consequences for *electric charge quantization*.

In contrast to gauge anomalies, anomalies in global symmetries need not be canceled but lead to observable consequences. An important example is scale invariance. In the absence of quark mass terms, the QCD Lagrangian has only one parameter — the dimensionless gauge coupling g_s. Hence, the Lagrangian of massless QCD is exactly scale invariant. This global symmetry is affected by an anomaly, because the quantization of the theory requires the introduction of a dimensionful cut-off — for example, the lattice spacing a in the lattice regularization. Remarkably, even when the cut-off is removed in the continuum limit $a \to 0$, the dimensionful scale Λ_{QCD} emerges via the process of dimensional transmutation.

Another anomaly affects the flavor-singlet axial $U(1)_A$ symmetry of massless QCD. This quantum effect gives a large mass to the η' meson. Only in the large N_c limit, in which the $U(1)_A$ anomaly is suppressed, the η' meson would become a massless Goldstone boson. Yet another anomaly affects the discrete global G-parity symmetry of QCD, which conserves the number of pions modulo 2. Remarkably, via electromagnetic interactions a single neutral pion can decay into two photons. This quantum effect changes the number of pions from one to zero and thus breaks G-parity anomalously. In contrast to many textbooks, we will point out that in a gauge-anomaly-free Standard Model with N_c quark colors, the width of the neutral pion, associated with the decay into two photons, is not proportional to N_c^2 but actually N_c-independent.

2.11 Euclidean Quantum Field Theory versus Classical Statistical Mechanics

The quantization of field theories, in particular, gauge theories, is a subtle mathematical problem. The functional integral approach (*i.e.* Feynman's path integral applied to quantum field theory) offers a very attractive alternative to canonical quantization. When real Minkowski time is analytically continued to purely imaginary Euclidean time, the functional integral becomes mathematically particularly well-behaved. As an extra bonus, Euclidean quantum field theory, in particular, when it is regularized on a 4-dimensional Euclidean space-time lattice, is analogous to a system of classical statistical mechanics. The Euclidean fields then correspond to generalized spin variables and the classical Hamilton function is analogous to the Euclidean lattice action of the quantum field theory. The temperature T, which controls the thermal fluctuations in classical statistical mechanics, is analogous to $\hbar$, which controls the strength of quantum fluctuations. A spin correlation function is analogous to a Euclidean 2-point function, whose decay determines a correlation length $\xi = 1/M$, where M is a particle mass.

The analogy between classical statistical mechanics and Euclidean field theory has far-reaching consequences, because the theory of critical phenomena can then be applied to field theory. In particular, a critical point, where a correlation length diverges in units of the lattice spacing, *i.e.* $\xi/a \to 0$, corresponds to the continuum limit of a Euclidean lattice field theory in which $Ma \to 0$. The insensitivity of the low-energy physics to the details of the regularization of quantum field theory follows from universality. Relevant, marginal, and irrelevant couplings follow from considerations of the *renormalization group*. Furthermore, Monte Carlo methods, which were originally developed for classical statistical mechanics, can be applied to lattice QCD in order to quantitatively address non-perturbative problems, which arise due to the strong interaction between quarks and gluons.

Part II

CONSTRUCTION OF THE STANDARD MODEL

Chapter 3

From Cooper Pairs to Higgs Bosons

In this chapter we introduce the scalar sector of the Standard Model. Even without gauge fields or fermions (leptons and quarks), there is interesting physics of Higgs bosons alone. In the Higgs sector of the Standard Model a global $SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}} \simeq O(4)$ symmetry *breaks spontaneously* down to the subgroup $SU(2)_{\mathrm{L=R}} \simeq O(3)$.[1] According to the Goldstone theorem, this gives rise to three massless *Nambu-Goldstone bosons*. Once *electroweak gauge fields* are included (which will be done in the next chapter), the gauge bosons W and Z become massive due to the *Higgs mechanism*. The photon, on the other hand, remains massless, as a consequence of the *unbroken* $U(1)_{\mathrm{em}}$ gauge symmetry of electromagnetism.

The analogies between the Higgs mechanism and the physics of superconductors have been pointed out, *e.g.*, by Philip Anderson. In particular, the scalar field describing the Cooper pairs of electrons in a superconductor is a condensed matter analogue of the Higgs field in particle physics. When Cooper pairs condense, even the $U(1)_{\mathrm{em}}$ symmetry breaks spontaneously and, consequently, the photon then also becomes massive. However, in this chapter we do not yet include gauge fields. Instead we concentrate on the dynamics of the scalar fields alone.

[1]The symbol $\simeq$ denotes a local isomorphism between two manifolds, which may still differ in their global topology.

Experiments before the LHC had excluded a Higgs particle lighter than 110 GeV. Precision tests of the Standard Model had favored a Higgs mass slightly above this limit, and indeed LHC experiments have found the Higgs particle at a mass of 126 GeV. Theoretically, obtaining an elementary Higgs particle with a renormalized mass much smaller than the Planck scale requires an unnatural fine-tuning of the bare mass. This is known as the *gauge hierarchy problem.* A similar problem does not arise for the composite Cooper pairs in condensed matter physics. Hence, one may wonder whether the physical Higgs particle is also composite. This scenario has been suggested by theoretical approaches beyond the Standard Model, in particular by technicolor models.

Another feature of the scalar sector of the Standard Model is its *triviality*. If one insists on removing the cut-off in the scalar quantum field theory describing the Higgs sector of the Standard Model, the theory becomes non-interacting. As a result, the Standard Model can only be a *low-energy effective field theory*, which is expected to break down at sufficiently high energy scales.

3.1 A Charged Scalar Field for Cooper Pairs

The prototype of a gauge theory is Quantum Electrodynamics (QED), the theory of the electromagnetic interaction between charged particles (*e.g.* electrons and positrons) via photon exchange. Here we consider electrically charged scalar particles (without spin). For example, we can think of the *Cooper pairs* in a superconductor. In ordinary superconductors, at temperatures of a few Kelvin (K), the Coulomb repulsion between electrons is overcome by an attractive interaction mediated by phonon-exchange (*i.e.* by couplings to the vibrations of the crystal lattice of ions). The resulting Cooper pairs form in the s-wave channel and have spin 0.[2] Hence, at energy scales well below the binding energy of a Cooper pair (*i.e.* below the energy gap of the superconductor), they can effectively be described by a scalar field.

---

[2]The mechanism for binding Cooper pairs in high-temperature superconductors, which also have spin 0 but form in the d-wave channel, is not yet understood.

Since we do not yet couple the scalar particle to a gauge field, strictly speaking, the analogous condensed matter systems are *superfluids* (with a spontaneously broken global $U(1)$ symmetry) rather than superconductors (with a spontaneously broken $U(1)_{\rm em}$ gauge symmetry of electromagnetism). In superfluid ^{4}He, the relevant atomic objects are bosons, consisting of an α-particle (two protons and two neutrons forming the atomic nucleus) as well as two electrons. When these bosons condense at temperatures around 2 K, the global $U(1)$ symmetry that describes the conserved boson number breaks spontaneously. Superfluid ^{3}He, on the other hand, consists of fermions: a helium nucleus with two protons but only one neutron, and two electrons. Before the fermionic ^{3}He atoms can Bose condense, they must form bosonic bound states which we also denote as Cooper pairs. This happens at very low temperatures around 2 mK. The Cooper pairs of superfluid ^{3}He form a spin triplet in the p-wave channel. Unlike in a superconductor, the Cooper pairs in superfluid ^{3}He are electrically neutral. Hence their Bose condensation leads to the spontaneous breaking of the *global* $U(1)$ symmetry corresponding to particle number conservation, but not of the local $U(1)_{\rm em}$ symmetry of electromagnetism.

A charged scalar particle is described by a complex field $\Phi(x) \in \mathbb{C}$. In fact, it takes two real degrees of freedom to describe both, a scalar and an anti-scalar. As we have discussed before, a quantum field theory can be defined by a Euclidean path integral over all field configurations

$$Z = \int \mathcal{D}\Phi \ \exp\left(-S[\Phi]\right) \ . \tag{3.1.1}$$

Here

$$S[\Phi] = \int d^4x \ \mathcal{L}(\Phi, \partial_\mu \Phi) \tag{3.1.2}$$

is the Euclidean action of the field $\Phi(x) = \phi_1(x) + \mathrm{i}\phi_2(x)$ ($\phi_i(x) \in \mathbb{R}$), with the Lagrangian

$$\mathcal{L}(\Phi, \partial_\mu \Phi) = \frac{1}{2}\partial_\mu \Phi^* \partial_\mu \Phi + V(\Phi) = \frac{1}{2}\partial_\mu \phi_1 \partial_\mu \phi_1 + \frac{1}{2}\partial_\mu \phi_2 \partial_\mu \phi_2 + V(\phi_1, \phi_2) \ . \tag{3.1.3}$$

A simple form for the potential is that of the $\lambda\Phi^4$-model,

$$V(\Phi) = \frac{m^2}{2}|\Phi|^2 + \frac{\lambda}{4!}|\Phi|^4 \ , \quad |\Phi|^2 = \Phi^*\Phi = \phi_1^2 + \phi_2^2 \ . \tag{3.1.4}$$

In the free case, $\lambda = 0$, the classical Euclidean field equations take the form

$$\partial_\mu \frac{\delta \mathcal{L}}{\delta \partial_\mu \Phi} - \frac{\delta \mathcal{L}}{\delta \Phi} = (\partial_\mu \partial_\mu - m^2)\Phi = 0 \ . \tag{3.1.5}$$

This is the 2-component Euclidean Klein-Gordon equation for a free charged scalar field. The Lagrangian of eqs. (3.1.3) and (3.1.4) has a global symmetry: it is invariant under $U(1)$ transformations

$$\Phi'(x) = \exp(\mathrm{i}e\varphi)\Phi(x) \qquad \Leftrightarrow \qquad \Phi'(x)^* = \exp(-\mathrm{i}e\varphi)\Phi(x)^* \ , \tag{3.1.6}$$

where $\varphi \in \mathbb{R}$ is a phase. Once the $U(1)$ symmetry is gauged, the parameter e will be identified as the electric charge of the field Φ.

We assume the coupling constant λ to be strictly positive to make sure that the potential is bounded from below. One can, however, choose $m^2 < 0$. The following discussion of spontaneous symmetry breaking is essentially classical and does not necessarily reveal the true nature of the quantum ground state. We distinguish two cases:

• For $m^2 \geq 0$ the potential has a single minimum at $\Phi = 0$. The classical solution of lowest energy (the classical vacuum) is simply the constant field $\Phi(x) = 0$. This vacuum configuration is invariant against the $U(1)$ transformations of eq. (3.1.6). Hence, in this case, the $U(1)$ symmetry is unbroken.

• For $m^2 < 0$, the trivial configuration $\Phi(x) = 0$ is unstable because it corresponds to a (local) maximum of the potential. The condition for a minimum now reads

$$\frac{\partial V}{\partial \Phi} = m^2 \Phi + \frac{\lambda}{3!}|\Phi|^2 \Phi = 0 \qquad \Rightarrow \qquad |\Phi|^2 = -\frac{6m^2}{\lambda} \ . \tag{3.1.7}$$

In this case the vacuum is no longer unique. Instead, there is a whole class of degenerate vacua

$$\Phi(x) = v \exp(\mathrm{i}\alpha) \ , \quad v = \sqrt{-\frac{6m^2}{\lambda}} \ , \tag{3.1.8}$$

parametrized by an angle $\alpha \in [0, 2\pi)$. The quantity v is the vacuum expectation value of the field Φ. Let us choose the vacuum state with $\alpha = 0$. Of course, such a choice breaks the $U(1)$ symmetry. Hence, in this case the global symmetry is spontaneously broken. In particular, there is not even a non-trivial subgroup of the $U(1)$ symmetry that leaves the vacuum configuration invariant. Hence, the $U(1)$ symmetry is spontaneously broken down to the trivial subgroup $\{1\}$. Expanding around the spontaneously selected minimum of the potential, one obtains

$$\begin{aligned}
&\Phi(x) = v + \sigma(x) + \mathrm{i}\pi(x) \quad \Rightarrow \quad \Phi(x)^* = v + \sigma(x) - \mathrm{i}\pi(x) \ , \\
&|\Phi(x)|^2 = [v + \sigma(x)]^2 + \pi(x)^2 \ , \\
&\partial_\mu \Phi(x) = \partial_\mu \sigma(x) + \mathrm{i}\partial_\mu \pi(x) \ , \quad \partial_\mu \Phi(x)^* = \partial_\mu \sigma(x) - \mathrm{i}\partial_\mu \pi(x) \ . \qquad (3.1.9)
\end{aligned}$$

We now want to express the Lagrangian in terms of the new, real-valued fields σ and π, which describe fluctuations around the vacuum configuration $\Phi(x) = v$ that we selected. We capture the low-energy physics — *i.e.* the dominant contributions to the path integral — by expanding up to second order in $\sigma(x)$ and $\pi(x)$,

$$\begin{aligned}
\frac{1}{2}\partial_\mu \Phi^* \partial_\mu \Phi &= \frac{1}{2}\partial_\mu \sigma \partial_\mu \sigma + \frac{1}{2}\partial_\mu \pi \partial_\mu \pi \\
V(\Phi) &= \frac{m^2}{2}(v+\sigma)^2 + \frac{m^2}{2}\pi^2 + \frac{\lambda}{4!}\left[(v+\sigma)^2 + \pi^2\right]^2 \\
&\approx \frac{m^2}{2}v^2 + m^2 v\sigma + \frac{m^2}{2}\sigma^2 + \frac{m^2}{2}\pi^2 + \frac{\lambda}{4!}\left(v^4 + 4v^3\sigma + 6v^2\sigma^2 + 2v^2\pi^2\right) \\
&= \frac{1}{2}\left(m^2 + \frac{\lambda}{2}v^2\right)\sigma^2 + c \ . \qquad (3.1.10)
\end{aligned}$$

Here c is an irrelevant additive constant. We interpret the term proportional to σ^2 as a mass term for the σ-field. The corresponding σ-particle has a mass squared

$$m_\sigma^2 = m^2 + \frac{\lambda}{2}v^2 = \frac{\lambda}{3}v^2 = -2m^2 > 0 \ . \qquad (3.1.11)$$

Since there is no term proportional to π^2, the corresponding π-particle is *massless* (*i.e.* $m_\pi = 0$). This massless particle is a *Nambu-Goldstone boson*. Its presence is characteristic for the spontaneous breaking of a global, continuous symmetry. The Goldstone theorem, which determines the number of Nambu-Goldstone bosons (one in case of a spontaneously broken $U(1)$ symmetry), will be discussed in Section 3.3. Once the $U(1)$ symmetry is gauged, which will be done in Chapter 4, the Nambu-Goldstone boson turns into the longitudinal polarization state of the gauge boson, which then becomes massive. For example, in a superconductor the spontaneously broken symmetry is the $U(1)_{\text{em}}$ gauge symmetry of electromagnetism. In that case, the photon becomes massive.

3.2 The Higgs Doublet

We now leave the condensed matter analogue behind and proceed to the Higgs sector of the Standard Model. Again a scalar field Φ plays a central rôle. However, the field Φ now has two complex components, it is a *complex doublet*. We deal with the *Higgs field* [3]

$$\Phi(x) = \begin{pmatrix} \Phi^+(x) \\ \Phi^0(x) \end{pmatrix} , \quad \Phi^+(x),\, \Phi^0(x) \in \mathbb{C} . \tag{3.2.1}$$

We follow the structure of the previous section and first discuss a model with only a *global* symmetry,

$$\begin{aligned} \mathcal{L}(\Phi, \partial_\mu \Phi) &= \frac{1}{2}\partial_\mu \Phi^\dagger \partial_\mu \Phi + V(\Phi) , \\ V(\Phi) &= \frac{m^2}{2}|\Phi|^2 + \frac{\lambda}{4!}|\Phi|^4 , \quad |\Phi|^2 = \Phi^\dagger \Phi = \Phi^{+*}\Phi^+ + \Phi^{0*}\Phi^0 . \end{aligned} \tag{3.2.2}$$

This Lagrangian is invariant under a class of $SU(2)$ transformations, which we denote as $SU(2)_{\text{L}}$,

$$\Phi'(x) = L\Phi(x) , \quad L \in SU(2)_{\text{L}} . \tag{3.2.3}$$

We recall that $SU(2)$ is the group of unitary 2×2 matrices with determinant 1,

$$L^\dagger = L^{T\,*} = L^{-1} , \quad \det L = 1 . \tag{3.2.4}$$

A general $SU(2)$ matrix L can be written in terms of complex numbers z_1 and z_2 with $\det L = |z_1|^2 + |z_2|^2 = 1$,

$$L = \begin{pmatrix} z_1 & -z_2^* \\ z_2 & z_1^* \end{pmatrix} \quad \Rightarrow \quad L^\dagger = \begin{pmatrix} z_1^* & z_2^* \\ -z_2 & z_1 \end{pmatrix} , \quad L^\dagger L = \mathbb{1} . \tag{3.2.5}$$

[3]The subscripts + and 0 will later turn out to correspond to electric charges.

This representation shows that the space of $SU(2)$ matrices is isomorphic to the 3-dimensional sphere S^3. The global $SU(2)_\mathrm{L}$ invariance of the Lagrangian follows from

$$\begin{aligned} |\Phi'(x)|^2 &= \Phi'(x)^\dagger \Phi'(x) = [L\Phi(x)]^\dagger L\Phi(x) = \Phi(x)^\dagger L^\dagger L \Phi(x) = |\Phi(x)|^2 \ , \\ \partial_\mu \Phi'(x)^\dagger \partial_\mu \Phi'(x) &= \partial_\mu \Phi(x)^\dagger L^\dagger L \partial_\mu \Phi(x) = \partial_\mu \Phi(x)^\dagger \partial_\mu \Phi(x) \ . \end{aligned} \tag{3.2.6}$$

In addition to the $SU(2)_\mathrm{L}$ symmetry, there is the so-called $U(1)_Y$ hypercharge symmetry which acts as

$$\Phi'(x) = \exp\left(-\mathrm{i}\frac{g'}{2}\varphi\right)\Phi(x) \ . \tag{3.2.7}$$

Just like the $SU(2)_\mathrm{L}$ symmetry, for the moment, we treat the $U(1)_Y$ hypercharge as a global symmetry. Once these symmetries will be gauged, *i.e.* made local, the constant g' will be identified as the gauge coupling strength of the $U(1)_Y$ gauge field. Interestingly, the global symmetry is actually even larger than the group $SU(2)_\mathrm{L} \times U(1)_Y$ identified so far: the action is indeed invariant under an extended group $SU(2)_\mathrm{L} \times SU(2)_\mathrm{R}$, with $U(1)_Y$ being an Abelian subgroup of $SU(2)_\mathrm{R}$. In order to make the additional $SU(2)_R$ symmetry manifest, we introduce another notation for the same Higgs field by re-writing it as a matrix,

$$\mathbf{\Phi}(x) = \left(\begin{array}{cc} \Phi^0(x)^* & \Phi^+(x) \\ -\Phi^+(x)^* & \Phi^0(x) \end{array}\right) \ . \tag{3.2.8}$$

We see that the matrix field $\mathbf{\Phi}$ belongs to $SU(2)$, up to a scale factor (provided $\mathbf{\Phi}$ is non-zero). In this notation, the Lagrangian (3.2.2) takes the form

$$\mathcal{L}(\mathbf{\Phi}, \partial_\mu \mathbf{\Phi}) = \frac{1}{4}\mathrm{Tr}\left[\partial_\mu \mathbf{\Phi}^\dagger \partial_\mu \mathbf{\Phi}\right] + \frac{m^2}{4}\mathrm{Tr}\left[\mathbf{\Phi}^\dagger \mathbf{\Phi}\right] + \frac{\lambda}{4!}\left(\frac{1}{2}\mathrm{Tr}\left[\mathbf{\Phi}^\dagger \mathbf{\Phi}\right]\right)^2 \ , \tag{3.2.9}$$

which is manifestly invariant under the global transformations

$$\mathbf{\Phi}(x)' = L\mathbf{\Phi}(x)R^\dagger \ , \quad L \in SU(2)_\mathrm{L} \ , \quad R \in SU(2)_\mathrm{R} \ . \tag{3.2.10}$$

The $SU(2)_\mathrm{R}$ symmetry is known as the *custodial symmetry*. By writing

$$R = \left(\begin{array}{cc} \exp\left(-\mathrm{i}\frac{g'}{2}\varphi\right) & 0 \\ 0 & \exp\left(\mathrm{i}\frac{g'}{2}\varphi\right) \end{array}\right) \ , \tag{3.2.11}$$

we can now identify $U(1)_Y$ as a subgroup of $SU(2)_{\mathrm{R}}$.

As a further alternative we introduce the 4-component vector notation

$$\begin{aligned}
&\vec{\phi}(x) = (\phi_1(x), \phi_2(x), \phi_3(x), \phi_4(x)) \ , \\
&\Phi^+(x) = \phi_2(x) + \mathrm{i}\ \phi_1(x) \ , \quad \Phi^0(x) = \phi_4(x) - \mathrm{i}\ \phi_3(x) \ , \\
&\mathbf{\Phi}(x) = \phi_4(x)\mathbf{1} + \mathrm{i}\left[\phi_1(x)\sigma^1 + \phi_2(x)\sigma^2 + \phi_3(x)\sigma^3\right] \ . \qquad (3.2.12)
\end{aligned}$$

Here σ^1, σ^2, and σ^3 are the Pauli matrices. In vector notation, the Lagrangian takes the form

$$\mathcal{L}(\phi, \partial_\mu \phi) = \frac{1}{2}\partial_\mu \vec{\phi} \cdot \partial_\mu \vec{\phi} + \frac{m^2}{2}\vec{\phi} \cdot \vec{\phi} + \frac{\lambda}{4!}\left(\vec{\phi} \cdot \vec{\phi}\right)^2 \ . \qquad (3.2.13)$$

This Lagrangian is manifestly $O(4)$-invariant under orthogonal rotations of the 4-component vector $\vec{\phi}$. This is precisely in agreement with the local isomorphism $SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}} \simeq O(4)$. We also see now from two perspectives that the global symmetry group has in total six generators.

As before, we distinguish between the symmetric and the broken phase.

• At the classical level, for $m^2 \geq 0$ there is a unique vacuum field configuration

$$\Phi(x) = \left(\begin{array}{c} 0 \\ 0 \end{array}\right) \ . \qquad (3.2.14)$$

In this case, the $SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}} \simeq O(4)$ symmetry is unbroken.

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

$$\Phi(x) = \left(\begin{array}{c} 0 \\ v \end{array}\right) \ , \quad v = \sqrt{-\frac{6m^2}{\lambda}} \in \mathbb{R}_+ \ , \qquad (3.2.15)$$

which implies

$$\mathbf{\Phi}(x) = v\mathbf{1} \ , \quad \vec{\phi}(x) = (0, 0, 0, v) \ . \qquad (3.2.16)$$

This vacuum configuration is not invariant under general $SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}}$ transformations. However, it is invariant under such transformations that obey $L = R$, which belong to the so-called vector subgroup $SU(2)_{\mathrm{L=R}}$. Hence, in this case the $SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}} \simeq O(4)$ symmetry is spontaneously broken down to the diagonal subgroup $SU(2)_{\mathrm{L=R}} \simeq O(3)$ which remains unbroken. The $U(1)_{\mathrm{em}}$ symmetry of electromagnetism will later be identified as a subgroup of $SU(2)_{\mathrm{L=R}}$.

Let us again expand around the vacuum configuration,

$$\Phi(x) = \begin{pmatrix} \pi_1(x) + \mathrm{i}\pi_2(x) \\ v + \sigma(x) + \mathrm{i}\pi_3(x) \end{pmatrix} . \tag{3.2.17}$$

To second order in the fluctuation fields σ and $\vec{\pi} = (\pi_1, \pi_2, \pi_3)$ we obtain

$$\begin{aligned} \frac{1}{2}\partial_\mu \Phi^\dagger \partial_\mu \Phi &= \frac{1}{2}\partial_\mu\sigma\partial_\mu\sigma + \frac{1}{2}\partial_\mu\vec{\pi}\cdot\partial_\mu\vec{\pi} \ , \\ V(\Phi) &= \frac{m^2}{2}\left[(v+\sigma)^2 + \vec{\pi}^{\,2}\right] + \frac{\lambda}{4!}\left[(v+\sigma)^2 + \vec{\pi}^{\,2}\right]^2 \\ &\approx \frac{m^2}{2}\left[v^2 + 2v\sigma + \sigma^2 + \vec{\pi}^{\,2}\right] + \frac{\lambda}{4!}\left[v^4 + 4v^3\sigma + 6v^2\sigma^2 + 2v^2\vec{\pi}^{\,2}\right] \\ &= \frac{1}{2}\left(m^2 + \frac{\lambda}{2}v^2\right)\sigma^2 + c \ , \end{aligned} \tag{3.2.18}$$

where c is again an irrelevant additive constant. Once more we find a massive σ-particle with

$$m_\sigma^2 = m^2 + \frac{\lambda}{2}v^2 = \frac{\lambda}{3}v^2 \ , \tag{3.2.19}$$

and in this case *three* massless Nambu-Goldstone bosons, π_1, π_2, and π_3. The massive σ-particle — a quantized fluctuation of the σ field — is known as the *Higgs particle*. While the Higgs particle is a singlet, the three Nambu-Goldstone bosons transform as a triplet under the unbroken $SU(2)_{\mathrm{L=R}} \simeq O(3)$ symmetry.

3.3 The Goldstone Theorem

In this chapter we have encountered a number of Nambu-Goldstone bosons. Let us now take a more general point of view and discuss the Goldstone theorem, which predicts the number of these massless particles for a general pattern of spontaneous breakdown of a continuous global symmetry.

As a prototype model, we consider an N-component real scalar field

$$\vec{\phi} = (\phi_1, \phi_2, \ldots, \phi_N) \tag{3.3.1}$$

with a potential $V(\vec{\phi})$ that is invariant under transformations of a symmetry group G. We further assume that this group has n_G generators T^a, with $a \in \{1, 2, \dots, n_G\}$, which are anti-symmetric $N \times N$ matrices. A general, infinitesimal symmetry transformation of the field takes the form

$$\vec{\phi}' = \exp(\omega_a T^a)\vec{\phi} \approx (1 + \omega_a T^a)\vec{\phi} \ , \tag{3.3.2}$$

for small angles ω_a. Let us assume that the potential has a set of degenerate vacua. We pick one spontaneously, $\vec{\phi} = \vec{v}$, and ask about the masses of fluctuations around this vacuum. First of all, since $\vec{v}$ is a minimum of the potential, we know that

$$\left.\frac{\partial V}{\partial \phi_i}\right|_{\vec{\phi}=\vec{v}} = 0 \ , \qquad i \in \{1, 2, \dots, N\} \ . \tag{3.3.3}$$

The matrix of second derivatives of the potential defines the masses,

$$M_{ij} = \left.\frac{\partial^2 V}{\partial \phi_i \partial \phi_j}\right|_{\vec{\phi}=\vec{v}} . \tag{3.3.4}$$

The eigenvalues of the matrix M are the squared masses of the physical particle fluctuations around the vacuum $\vec{v}$.

Let us now assume the vacuum to be invariant under the transformations in a subgroup H of G, $H \subset G$, which is generated by T^b with $b \in \{1, 2, \dots, n_H\}$ and $n_H \leq n_G$, *i.e.*

$$(1 + \omega_b T^b)\vec{v} = \vec{v} \quad \Rightarrow \quad T^b \vec{v} = 0 \ . \tag{3.3.5}$$

Invariance of the potential under the transformation group G implies for any vector $\vec{\phi}$

$$0 = V(\vec{\phi}') - V(\vec{\phi}) = \frac{\partial V}{\partial \phi_i} \ \omega_a T^a_{ij} \ \phi_j \ . \tag{3.3.6}$$

We differentiate this equation with respect to ϕ_k and evaluate it at $\vec{\phi} = \vec{v}$,

$$0 = \left.\frac{\partial^2 V}{\partial \phi_k \partial \phi_i}\right|_{\vec{\phi}=\vec{v}} \omega_a T^a_{ij} \ v_j + \left.\frac{\partial V}{\partial \phi_i}\right|_{\vec{\phi}=\vec{v}} \omega_a T^a_{ki} \quad \Rightarrow \quad M_{ki}(T^a \vec{v})_i = 0 \ . \tag{3.3.7}$$

Here we have used eq. (3.3.3). For the unbroken subgroup H, *i.e.* for $a \leq n_H$, this is trivially satisfied because $T^a \vec{v} = 0$, according to eq. (3.3.5). For the remaining generators with $a \in \{n_H + 1, n_H + 2, \ldots, n_G\}$, however, the equation implies that $T^a \vec{v}$ is an eigenvector of the matrix M with eigenvalue zero. Hence the difference $n_G - n_H$ is the degeneracy of the eigenvalue 0, *i.e.* the dimension of the manifold of vacuum states. Therefore, there are $n_G - n_H$ *massless modes* in a vacuum $\vec{v}$. Upon quantization, in a relativistic quantum field theory in more than two space-time dimensions, these modes turn into $n_G - n_H$ massless Nambu-Goldstone bosons.

For example, when Cooper pairs condense, the symmetry $G = U(1)$ that breaks spontaneously down to the trivial subgroup $H = \{1\}$ gives rise to $n_G - n_H = 1 - 0 = 1$ Nambu-Goldstone bosons. In the Higgs sector of the Standard Model, the symmetry $G = SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}} \simeq O(4)$ breaks spontaneously to the subgroup $H = SU(2)_{\mathrm{L=R}} \simeq O(3)$. Hence, in this case there are $n_G - n_H = 6 - 3 = 3$ Nambu-Goldstone bosons. In general, when a symmetry $G = O(N)$ breaks to $H = O(N-1)$ the number of broken generators is

$$n_G - n_H = \frac{1}{2}N(N-1) - \frac{1}{2}(N-1)(N-2) = N - 1 \ , \tag{3.3.8}$$

such that there are $N - 1$ Nambu-Goldstone bosons.

In *non-relativistic* theories the number of massless modes does not necessarily coincide with the number of Nambu-Goldstone bosons. For example, a quantum ferromagnet with a global symmetry $G = O(3)$ that is spontaneously broken down to the subgroup $H = O(2)$ has $n_G - n_H = 3 - 1 = 2$ massless modes, but only one Nambu-Goldstone particle — a magnetic spin-wave or magnon. Ferromagnetic magnons have a non-relativistic dispersion relation $E \propto |\vec{p}\,|^2$.[4] This results from the fact that the order parameter of the ferromagnet — the uniform magnetization, *i.e.* the total spin — is a conserved quantity. Quantum antiferromagnets also have an $O(3)$ symmetry that is spontaneously broken down to $O(2)$. However, other than in a ferromagnet, the staggered magnetization order parameter of an antiferromagnet is not a conserved quantity. As a consequence, antiferromagnetic magnons have a relativistic dispersion relation, $E \propto |\vec{p}|$, and in this case there are indeed two massless Nambu-Goldstone bosons.

[4] Despite the fact that these particles are massless, they do not obey the relativistic dispersion relation $E \propto |\vec{p}\,|$.

3.4 The Mermin-Wagner Theorem

It should be noted that Nambu-Goldstone bosons can only exist in more than two space-time dimensions. In the context of condensed matter physics David Mermin and Herbert Wagner, as well as Pierre Hohenberg were first to prove that the spontaneous breakdown of a continuous global symmetry cannot occur in two space-time dimensions. In the context of relativistic quantum field theories a corresponding theorem was proved by Sidney Coleman. The Mermin-Wagner theorem states that in two space-time dimensions an order parameter corresponding to a continuous global symmetry necessarily has a vanishing vacuum expectation value. Therefore massless Nambu-Goldstone bosons — which appear as a consequence of spontaneous symmetry breaking — cannot exist is this case. This behavior is due to infrared quantum fluctuations, which are particularly strong in lower dimensions. In more than two space-time dimensions, quantum fluctuations are more restricted because the field variables (*e.g.* on a lattice) are then coupled to a larger number of neighboring sites. The Mermin-Wagner theorem even applies to theories in $(2+1)$ dimensions, at least at non-zero temperature $T > 0$. In that case, the extent $\beta = 1/T$ of the Euclidean time dimension is finite, and thus there are only two infinitely extended dimensions. Consequently, in $(2+1)$ dimensions a continuous global symmetry can break spontaneously only at zero temperature $T = 0$. As first discussed by Sudip Chakravarty, Bertrand Halperin, and David Nelson, and studied in great detail by Peter Hasenfratz and Ferenc Niedermayer, due to non-perturbative effects, at small non-zero temperatures $T > 0$ the Nambu-Goldstone modes of a $(2+1)$-dimensional antiferromagnet obtain a finite correlation length, which is exponentially large in the inverse temperature, and thus diverges in the $T \to 0$ limit.

It should be noted that the Mermin-Wagner theorem does not exclude the existence of massless bosons in two dimensions, it just states that such particles cannot result from spontaneous symmetry breaking. For example, as first understood by Vadim Berezinskii, and independently by John Kosterlitz and David Thouless, at sufficiently low (but still non-zero) temperatures the 2-dimensional XY-model contains a massless boson. This boson exists, despite the fact that the Abelian continuous global $O(2)$ symmetry of the model is not spontaneously broken. In the XY-model, the low- and the high-temperature phase are separated by a so-called Berezinskii-

Kosterlitz-Thouless phase transition. Massless bosons can even arise in 2-dimensional systems with a non-Abelian $O(3)$ symmetry. For example, as was first derived by Hans Bethe, $(1+1)$-dimensional antiferromagnetic quantum spin chains with spin $\frac{1}{2}$ are gap-less. Duncan Haldane has pointed out that the low-energy effective field theory, which describes spin chains with half-integer spin, is a 2-dimensional $O(3)$-model with non-trivial vacuum angle $\theta = \pi$.[5] As was understood in detail by Ian Affleck, this system is a conformal field theory which contains massless bosons, despite the fact that the $O(3)$ symmetry is not spontaneously broken. Quantum spin chains with integer spins, on the other hand, correspond to $\theta = 0$. These systems have a gap and thus do not contain massless particles

The Mermin-Wagner theorem does not exclude either the spontaneous breakdown of a *discrete* symmetry in two dimensions. For example, at sufficiently low (but again non-zero) temperature the 2-dimensional Ising model has a spontaneously broken discrete $\mathbf{Z}(2)$ symmetry. Just as continuous symmetries cannot break spontaneously in two dimensions, discrete symmetries cannot break in one dimension. In a single space-time dimension there is just time and quantum field theory thus reduces to quantum mechanics of a finite number of degrees of freedom. Since spontaneous symmetry breaking is a collective phenomenon that necessarily involves an infinite number of degrees of freedom, it cannot arise in quantum mechanics (except for discrete symmetries at zero temperature).

3.5 Low-Energy Effective Field Theory

Since they are massless, Nambu-Goldstone bosons dominate the low-energy physics of any system with a spontaneously broken continuous global symmetry in more than two space-time dimensions. There is a general effective Lagrangian technique that describes the low-energy dynamics of the Nambu-Goldstone bosons. This approach was pioneered by Steven Weinberg and extended to a systematic method by Jürg Gasser and Heinrich Leutwyler for the pions — which represent the Nambu-Goldstone bosons of the spontaneously broken chiral symmetry of QCD. Hence, this method is known as

[5]In the framework of QCD, the vacuum angle θ will be discussed in detail in Chapter 10.

chiral perturbation theory. It is, however, generally applicable to any system of Nambu-Goldstone bosons.

We will now illustrate this technique for the Nambu-Goldstone bosons that arise in the Higgs sector of the Standard Model (before gauging the symmetry). As we have seen, the global symmetry of the Higgs sector is $G = SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}} \simeq O(4)$, which then breaks spontaneously down to a single $H = SU(2)_{\mathrm{L=R}} \simeq O(3)$ symmetry. Generally, in a low-energy effective theory the Nambu-Goldstone bosons are described by fields in the coset space G/H — the manifold of the group G with points being identified if they are connected by a symmetry transformation in the unbroken subgroup H. We saw that the dimension of the coset space, $n_G - n_H$, corresponds to the number of Nambu-Goldstone bosons. In the Higgs sector of the Standard Model, this coset space is

$$G/H = SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}}/SU(2)_{\mathrm{L=R}} = SU(2) \ , \tag{3.5.1}$$

or equivalently $G/H = O(4)/O(3) = S^3$. Hence, the three Nambu-Goldstone bosons can be described by a matrix-valued field $U(x) \in SU(2)$. It should be noted that the $SU(2)$ group manifold is isomorphic to a 3-dimensional sphere S^3. One may think of the field $U(x)$ as the "angular" degree of freedom of the Higgs field matrix of eq. (3.2.8), *i.e.*

$$\mathbf{\Phi}(x) = \begin{pmatrix} \Phi^0(x)^* & \Phi^+(x) \\ -\Phi^+(x)^* & \Phi^0(x) \end{pmatrix} = |\Phi(x)|\, U(x) \ , \quad |\Phi(x)|^2 = |\Phi^+(x)|^2 + |\Phi^0(x)|^2 . \tag{3.5.2}$$

Indeed, we have seen that the "radial" fluctuations of the magnitude $|\Phi|^2$ give rise to the massive Higgs particle, while the angular fluctuations along the vacuum manifold give rise to three massless Nambu-Goldstone bosons. From the $SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}}$ transformation rule of the Higgs field $\mathbf{\Phi}(x)$, eq. (4.3.2), one obtains

$$U'(x) = LU(x)R^\dagger \ , \quad L \in SU(2)_{\mathrm{L}} \ , \quad R \in SU(2)_{\mathrm{R}} \ . \tag{3.5.3}$$

The effective field theory is formulated as a systematic low-energy expansion. The low-energy physics of the Nambu-Goldstone bosons is dominated by those terms in the effective Lagrangian that contain a small number of derivatives. In Fourier space, a spatial derivative corresponds to a momentum and a temporal derivative corresponds to an energy. Hence, multiple-derivative terms are suppressed at low energies. All terms of the effective

Lagrangian must be invariant under all symmetries of the underlying microscopic system — in this case, of the Higgs sector of the Standard Model. In particular, the effective Lagrangian must be invariant under Lorentz transformations, and under the $SU(2)_\mathrm{L} \times SU(2)_\mathrm{R}$ transformations of eq. (3.5.3). The effective Lagrangian is constructed, order by order, by writing down all terms that are consistent with the symmetries. In the systematic derivative expansion, one starts with terms containing no derivatives. For example, the term $\mathrm{Tr}[U^\dagger U]$ is both Lorentz-invariant and $SU(2)_\mathrm{L} \times SU(2)_\mathrm{R}$-invariant. However, since $U \in SU(2)$ implies $U^\dagger U = \mathbf{1}$, this term is just a trivial constant. Indeed, there are no non-trivial terms without derivatives. Furthermore, there are no terms with just a single derivative, because its uncontracted Lorentz index would violate Lorentz invariance. The leading term of the effective Lagrangian therefore has two derivatives and is given by

$$\mathcal{L}(\partial_\mu U) = \frac{F^2}{4} \mathrm{Tr}\left[\partial_\mu U^\dagger \partial_\mu U\right] + \ldots \tag{3.5.4}$$

Higher order terms with four or six derivatives (represented by the dots) contribute less at low energies. Each term appears with a coefficient (F^2 in eq. (3.5.4)). These coefficients are called *low-energy constants*. They enter the effective theory as free parameters whose values cannot be deduced from symmetry considerations. Hence a theoretical prediction for them must be based on the underlying, fundamental theory. Comparing eq. (3.5.4) with eq. (3.2.9), using eq. (3.5.2), and setting $|\Phi|^2$ to its vacuum value v^2, one obtains a classical estimate of the low-energy parameter $F = v$. This estimate gives the correct order of magnitude but should not be taken too seriously. In order to properly identify the correct value of F based on the parameters of the Standard Model, one must take quantum effects into account.

The Higgs sector of the Standard Model — expressed as an N-component $\lambda\phi^4$-model (with $N = 4$) — is known as a *linear σ-model*. It is characterized by the Lagrangian

$$\begin{aligned}
\mathcal{L}(\vec{\phi}, \partial_\mu \vec{\phi}) &= \frac{1}{2} \partial_\mu \vec{\phi} \cdot \partial_\mu \vec{\phi} + \frac{m^2}{2} |\vec{\phi}|^2 + \frac{\lambda}{4!} |\vec{\phi}|^4 \\
&= \frac{1}{2} \partial_\mu \vec{\phi} \cdot \partial_\mu \vec{\phi} + \frac{\lambda}{4!} \left(|\vec{\phi}|^2 + \frac{6m^2}{\lambda} \right)^2 + c \\
&= \frac{1}{2} \partial_\mu \vec{\phi} \cdot \partial_\mu \vec{\phi} + \frac{\lambda}{4!} \left(|\vec{\phi}|^2 - v^2 \right)^2 + c \ ,
\end{aligned} \tag{3.5.5}$$

where c is an irrelevant constant. In the limit $\lambda \to \infty$, the action diverges unless $|\vec{\phi}(x)| = v$. Introducing the N-component unit-vector field $\vec{s}(x) = \vec{\phi}(x)/v$ (with $|\vec{s}(x)| = 1$ at any x), we arrive at the *non-linear σ-model*,[6]

$$\mathcal{L}(\partial_\mu \vec{s}) = \frac{v^2}{2} \partial_\mu \vec{s} \cdot \partial_\mu \vec{s} \ . \tag{3.5.6}$$

The non-linear constraint, $|\vec{s}(x)| = 1$, can be implemented in the measure of the path integral . Identifying

$$U(x) = s_4(x)\mathbf{1} + \mathrm{i} \ \left[s_1(x)\sigma^1 + s_2(x)\sigma^2 + s_3(x)\sigma^3 \right] \tag{3.5.7}$$

for $N = 4$, eq. (3.5.6) has just the structure of $\mathcal{L}(\partial_\mu U)$ in eq. (3.5.4). Hence, the low-energy effective theory for the Higgs sector is in fact a non-linear σ-model.

It is also interesting to anticipate that the very same effective Lagrangian describes the low-energy dynamics of the pions in QCD with two massless quarks. Only the value of the coupling constant F, which then corresponds to the pion decay constant, is different. The structure of the effective Lagrangian $\mathcal{L}(\partial_\mu U)$ solely depends on the symmetry and how it breaks spontaneously. Indeed, the chiral symmetry of two flavor QCD is again $SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}}$, which breaks spontaneously to $SU(2)_{\mathrm{L=R}}$.

Finally, we remark that the effective Lagrangian technique is still applicable if the spontaneous symmetry breaking is supplemented by a small amount of explicit breaking. If one adds a symmetry breaking term to the underlying microscopic Lagrangian, the Nambu-Goldstone bosons pick up a small mass. In that case, also the effective Lagrangian $\mathcal{L}(\partial_\mu U)$ contains symmetry breaking terms. It is then expanded in powers of the momenta *and* of the symmetry breaking parameter, according to some suitable counting rule. In the description of a classical ferromagnet, an explicit symmetry breaking term could represent a small external magnetic field. In two flavor QCD it corresponds to the small but finite quark masses of the flavors u and d, which then result in a small, non-zero mass for the pions (the pion is the lightest quark and gluon bound state). Chiral perturbation theory will be discussed further in Chapter **??**.

[6]For the special case $N = 1$, *i.e.* for a one-component scalar field on the lattice, this limit leads to the Ising model.

3.6 The Hierarchy Problem

Since λ is dimensionless[7], the parameter m^2 is the only dimensionful parameter in the Higgs sector of the Standard Model. We saw that this parameter determines the vacuum expectation value of the Higgs field

$$v = \sqrt{-\frac{6m^2}{\lambda}} , \tag{3.6.1}$$

which sets the energy scale for the breaking of the $SU(2)_L \times SU(2)_R$ symmetry. In fact, even after gauge and fermion fields have been added to the Standard Model, there will still only be this single dimensionful parameter. At the classical level, one could simply take the point of view that v is a truly fundamental energy scale in units of which all other dimensionful physical quantities can be expressed. The experimental value $v \approx 245$ GeV has been derived from the observed masses of the W- and Z-bosons.

At the quantum level, however, the situation is more complicated. Quantum field theories must be regularized and renormalized. Indeed, the ultraviolet cut-off Λ represents another energy scale that enters the quantum theory through the process of regularization. When one renormalizes the theory, one attempts to move the cut-off to infinity, keeping the physical masses and thus v fixed. As we will discuss in the next section, this is problematic in the Standard Model, because the Higgs sector is "trivial", *i.e.* it becomes a non-interacting theory in the $\Lambda \rightarrow \infty$ limit.

We know already that the Standard Model cannot be the "Theory of Everything" because it does not include gravity. The natural energy scale of gravity is the Planck scale

$$M_{\text{Planck}} = \frac{1}{\sqrt{G}} \approx 10^{19}\,\text{GeV} , \tag{3.6.2}$$

where G is Newton's constant. Even if we would assume (most likely quite unrealistically) that the Standard Model describes the physics correctly all the way up to the Planck scale, it would necessarily have to break down at that scale. In this sense, we can think of M_{Planck} as an ultimate ultra-violet cut-off of the Standard Model.

[7] Note that in d space-time dimensions the self-coupling constant λ has the dimension Mass^{d-4}. Hence, $d = 4$ is a special case with respect to its dimension.

Once we have appreciated the existence of the two fundamental scales v of the Standard Model and M_{Planck} of gravity with

$$\frac{v}{M_{\text{Planck}}} \approx 10^{-17}, \tag{3.6.3}$$

we are confronted with the hierarchy problem:

Why is the ratio of the electroweak scale and the Planck scale so small?

Although it does not extend over 120 orders of magnitude, this hierarchy problem represents a similar puzzle as the cosmological constant problem . One wonders whether there may be a dynamical mechanism that makes v naturally very much smaller than M_{Planck} or any other relevant ultra-violet cut-off scale Λ.

Let us discuss the hierarchy problem in the context of the lattice regularized scalar field theory. Nature must have found a concrete way to regularize the Higgs physics at ultra-short distances. Due to renormalizability and universality, only the symmetries, but not the details of this regularization should matter at low energies. For simplicity, we will use the regularization on a space-time lattice with spacing a as an admittedly oversimplified model of Nature at ultra-short distances. In other words, in this context we identify the lattice cut-off $\Lambda = 1/a$ with the Planck scale M_{Planck}. In the lattice regularization, the scalar field theory is characterized by the partition function

$$Z = \prod_x \int d\vec{\phi}_x \, \exp(-S_E[\vec{\phi}]) \ , \tag{3.6.4}$$

with the Euclidean lattice action given by

$$S_E[\vec{\phi}] = \sum_x a^4 \left[\frac{1}{2} \sum_\mu \left(\frac{\vec{\phi}_x - \vec{\phi}_{x+\hat{\mu}}}{a} \right)^2 + V(\vec{\phi}_x) \right] \ . \tag{3.6.5}$$

Here $\vec{\phi}_x$ is the scalar field at the lattice point x, and $\hat{\mu}$ is a vector of length a along the μ-direction. The first term is the finite difference analogue of the continuum expression $\frac{1}{2}\partial_\mu\vec{\phi} \cdot \partial_\mu\vec{\phi}$ and the potential

$$V(\vec{\phi}_x) = \frac{m^2}{2}|\vec{\phi}_x|^2 + \frac{\lambda}{4!}|\vec{\phi}_x|^4 \tag{3.6.6}$$

is the same as in the continuum theory. Just as the continuum theory (which is defined perturbatively), the lattice theory exists in two phases, one with and one without spontaneous symmetry breaking. The two phases are separated by a phase transition line $m_c^2(\lambda)$. For $m^2 < m_c^2(\lambda)$ the model is in the broken phase with three massless Nambu-Goldstone bosons, while for $m^2 > m_c^2(\lambda)$ it is in the massive symmetric phase. As intensive Monte Carlo simulations have shown, the phase transition is of second order. This implies that the correlation length in units of the lattice spacing a, $\xi_\sigma/a = 1/(m_\sigma a)$, corresponding to the inverse Higgs particle mass, diverges at the phase transition line. The Higgs mass (and thus the vacuum value v of the scalar field) behaves as

$$m_\sigma a = \frac{m_\sigma}{\Lambda} \sim |m^2 - m_c^2(\lambda)|^\nu \ . \tag{3.6.7}$$

Here ν is a critical exponent which takes the mean field theory value 1/2. If we want to identify the lattice cut-off Λ with the Planck scale M_{Planck} we must realize the hierarchy $m_\sigma/M_{\text{Planck}} \approx 10^{-16}$. To achieve this, for a given value of λ, one has to fine-tune the bare parameter m^2 to the critical value $m_c^2(\lambda)$ to many digits accuracy. This appears very unnatural. Explaining the hierarchy between the electroweak and the Planck scale without a need for fine-tuning is the challenge of the hierarchy problem.

For gauge fields a similar hierarchy problem does not exist. For example, the photon is naturally massless as a consequence of the unbroken gauge symmetry in the Coulomb phase of QED. Gluons, which are confined inside hadrons, also naturally exist at low energy scales, as a consequence of the property of asymptotic freedom of QCD. From a perturbative point of view, there is no hierarchy problem for fermions either, because fermion mass terms are forbidden by chiral symmetry. However, when considered *beyond* perturbation theory, fermions do suffer from a severe hierarchy problem. In particular, when regularized naively on a 4-dimensional space-time lattice, fermions suffer from species doubling. When the fermion doublers are removed by breaking chiral symmetry explicitly, without unnatural fine-tuning of the bare fermion mass, the renormalized mass flows to the cut-off scale. Remarkably, the non-perturbative hierarchy problem of fermions has been solved very elegantly by formulating the theory with an additional spatial dimension of finite extent. In five dimensions fermions may get localized on a 4-dimensional domain wall. Indeed, domain wall fermions are naturally light without any fine-tuning. Domain wall fermions provide a $(4+1)$-dimensional particle physics analogue of the chiral edge states of a

$(2+1)$-dimensional quantum Hall sample.

There have been attempts to solve the hierarchy problem of the Higgs sector of the Standard Model by postulating supersymmetry — a symmetry between bosons and fermions. In supersymmetric extensions of the Standard Model, each known particle has a superpartner. For example, the electron has a scalar superpartner — the so-called selectron — and the photon has a fermionic superpartner — the photino. Similarly, the Higgs particle has a fermionic Higgsino partner. By supersymmetry the Higgs mass (and thus the scale v) is then tied to the Higgsino mass, which is protected from running to the cut-off scale by chiral symmetry. Hence, with supersymmetry, elementary scalar particles (such as the Higgs particle) can be light without unnatural fine-tuning. At present supersymmetry is understood mostly perturbatively. Beyond perturbation theory, in particular on a lattice, it is highly non-trivial to construct supersymmetric theories. Since supersymmetry is intimately related to infinitesimal space-time translations, it is not surprising that discretizing space-time breaks supersymmetry explicitly. Indeed, obtaining supersymmetry in lattice theories often requires an unnatural fine-tuning of bare mass parameters, such that the hierarchy problem would remain unsolved.

Moreover, until now no superpartners have been observed. Hence, we do not know whether supersymmetry is realized in Nature. While the status of supersymmetry is yet unclear, ultimately the LHC will decide whether supersymmetry exists at the TeV scale. If supersymmetry is realized in Nature, it must be explicitly or spontaneously broken, otherwise superpartners should be degenerate with the known particles and should have long been detected. Since supersymmetry itself is not well understood beyond perturbation theory, it does not fit well into the concepts of this book. The same is true for other ideas for solving the hierarchy problem, *e.g.*, using extra dimensions.

An interesting non-perturbative approach to the hierarchy problem, which does fit well into the concepts of this book, is *technicolor*. In analogy to Cooper-pair condensation, in technicolor models the electroweak symmetry is spontaneously broken by the condensation of fermion pairs. Just as quarks are bound by strong color forces, the so-called techni-fermions are bound by very strong technicolor forces. In technicolor models the Higgs particle is a composite of two techni-fermions. While technicolor has its own problems, it remains to be seen whether the LHC will find evidence for this intriguing idea.

3.7 Triviality of the Standard Model

As we have seen, the Higgs sector of the Standard Model is a 4-component $\lambda\phi^4$-model. The Lagrangian contains the dimensionful parameter m^2, as well as the dimensionless scalar self-coupling λ. These parameters determine the vacuum expectation value

$$v = \sqrt{-\frac{6m^2}{\lambda}} \tag{3.7.1}$$

of the scalar field, as well as the Higgs mass

$$m_\sigma = \sqrt{-2m^2} = \sqrt{\frac{\lambda}{3}}\, v\ . \tag{3.7.2}$$

Hence, a *heavy* Higgs particle requires a *strongly* coupled scalar field (with a large value of λ). We have obtained these results essentially by considering the model just at the level of classical field theory. When the theory is quantized using perturbation theory, the bare parameters are renormalized, but remain free parameters. Thus the Higgs mass m_σ still appears arbitrary.

However, when a $\lambda\phi^4$-model is fully quantized beyond perturbation theory, a new feature arises. In the lattice regularization there is overwhelming evidence (albeit no rigorous proof) that the $\lambda\phi^4$-model — and hence the Standard Model — is *trivial* in $d \geq 4$. This means that the renormalized self-coupling λ goes to zero if one insists on sending the ultra-violet cut-off to infinity. In other words, the continuum limit $a \to 0$ of a lattice $\lambda\phi^4$-model is just a free field theory. How can we then use it to define the Standard Model as an interacting field theory beyond perturbation theory?

Indeed, one *should not* insist on completely removing the ultra-violet cut-off. This means that the Standard Model cannot possibly make sense at arbitrarily high energies (beyond the finite cut-off). Hence, it must be considered a low-energy effective theory, which must necessarily be replaced by something more fundamental at sufficiently high energies. In other words, the Standard Model could not even in principle be the "Theory of Everything". This is actually a remarkable property of the Standard Model: it kindly informs us about its own limitations and tells us that it will eventually break down. Non-trivial theories (like QCD), on the other hand, remain interacting even when the cut-off is removed completely. These theories could, in principle, be valid at arbitrarily high energy scales.

The triviality of the Standard Model leads to an estimate for an upper bound on the Higgs mass. A heavy Higgs boson corresponds to a large value of λ. On the other hand, we just pointed out that we get a free theory, $\lambda = 0$, when we remove the cut-off completely. Only when we leave the cut-off finite, we can get a heavy Higgs particle. However, the theory would clearly not make sense if it led to a Higgs mass similar to — or even larger than the ultra-violet cut-off. In the lattice regularization this puts an upper limit on the Higgs mass of around 600 GeV. Although this limit is not universal (it depends on the details of the regularization that one chooses), the triviality bound suggests that the (standard) Higgs particle should have a mass below about 600 GeV — or that the Standard Model is replaced by some new physics at that energy scale. Before the LHC, a systematic experimental search for the Higgs particle was performed up to around 110 GeV. Based on experimental data which were influenced by virtual Higgs effects, one expected to find the Higgs particle at energies between about 110 and 200 GeV. Indeed the LHC has found the Higgs partice at a mass of 126 GeV.

A (non-perturbatively renormalized) Higgs mass of $m_\sigma \approx 100$ GeV translates into a cut-off $\Lambda \approx 10^{36}$ GeV. This is far above the Planck scale (of about 10^{19} GeV), *i.e.* in a regime where physics is not understood at all. Hence a finite cut-off in this range is completely unproblematical. However, if we would move up to a Higgs mass of $m_\sigma \approx 600$ GeV, we would find the corresponding cut-off at about 6 TeV, *i.e.* at $\Lambda \approx 10 m_\sigma$. An even heavier Higgs particle would come too close to the cut-off scale to make any sense. Therefore 600 GeV appears as a reasonable magnitude for the theoretical upper bound. We note again that the upper bound is not universal, *e.g.* it depends on the short-distance details of the lattice action. In practice, using different regularizations that seem reasonable, this ambiguity has about a 10 % effect on the upper bound.

3.8 Electroweak Symmetry Restauration at High Temperature

Since the hot big bang, the Universe has undergone a dramatic evolution. The big bang itself represents a mathematical singularity in the solutions of classical general relativity, indicating our incomplete understanding of gravity. In the moment of the big bang, the energy density of the Universe was at the Planck scale where quantum effects of gravity are strong, and one would expect that the classical singularity is eliminated by quantum fluctuations of the space-time metric. Since we presently don't have an established theory of quantum gravity, we can only speculate about the time at and immediately after the big bang. However, since the Universe is expanding and thus cooling down, it soon reaches the energy scales of the Standard Model. Indeed, only about 10^{-14} sec after the big bang, the Universe has cooled down to temperatures in the TeV range, and its further evolution can then be understood based on the Standard Model combined with classical general relativity.

To a good approximation, the early Universe undergoes an adiabatic expansion (*i.e.* the total entropy is conserved), in which it remains in thermal equilibrium. When the Universe had a temperature of about 1 TeV, it contained an extremely hot gas of quarks, leptons, gauge bosons, Higgs particles, and perhaps other yet undetected particles, *e.g.* those forming the dark matter component of the Universe. At temperatures $T \gg v = 246$ GeV thermal fluctuations are so violent that the very early Universe was in an unbroken, symmetric phase. Just as the spontaneous magnetization of a ferromagnet is destroyed at high temperatures, the spontaneous order of the Higgs field cannot be maintained in the presence of strong thermal fluctuations. One often says that the $SU(2)_\mathrm{L} \times SU(2)_\mathrm{R}$ symmetry is restored in the early Universe. Of course, it would be more precise to say that it was not yet spontaneously broken. The high-temperature $SU(2)_\mathrm{L} \times SU(2)_\mathrm{R}$ symmetric phase that was realized in the early Universe should not be confused with the unbroken vacuum state (with $v = 0$) that would exist at zero temperature for $m^2 > 0$. As the Universe expands an cools, it eventually ends up in the broken symmetry vacuum (with $v \neq 0$) that we live in today.

It is interesting to ask how the expectation value $v(T)$ of the Higgs field depends on the temperature T. In particular, one expects a phase transition at some critical temperature T_c. For temperatures $T > T_c$ the early Universe is in the symmetric phase with $v(T) = 0$. When it expands and cools it goes through the phase transition and enters the broken phase in which $v(T) \neq 0$. The order of this so-called electroweak phase transition has an impact on the dynamics of the early Universe. In particular, a strong first order phase transition would have drastic consequences. Just like boiling water forms expanding bubbles of steam inside the liquid phase, a first order electroweak phase transition would also proceed via bubble nucleation. In this case, bubbles of broken phase would form inside the early symmetric phase. Since the bubble wall costs a finite amount of surface energy, the formation of these bubbles would be delayed by supercooling. Once the Universe has cooled sufficiently, bubbles of broken phase would suddenly nucleate and expand quickly, soon filling all of the Universe. The dynamics of a first order phase transition takes the system out of thermal equilibrium. As discussed by Andrei Sakharov in 1967, besides C and CP violation and the existence of baryon number violating processes, deviation from thermal equilibrium is a necessary prerequisite for dynamically generating the baryon asymmetry — the observed surplus of matter over anti-matter. As we will discuss in later chapters, the Standard Model indeed violates both C and CP as well as baryon number (at least at sufficiently high temperatures). In order to decide whether Standard Model physics alone might be able to explain the origin of the baryon asymmetry, it is thus vital to understand the nature of the electroweak phase transition.

3.9 Extended Model with Two Higgs Doublets

This section discusses physics beyond the Standard Model and may be skipped in a first reading.

Although the Higgs particle has been identified at the LHC at a mass of 126 GeV, the Higgs sector, which holds the key to the understanding of electroweak $SU(2)_\mathrm{L} \times SU(2)_\mathrm{R}$ symmetry breaking, is still the experimentally least well tested aspect of the Standard Model. The LHC has already produced the Higgs particle. Hopefully, it will also reveal exciting physics beyond the Standard Model, perhaps including supersymmetry, technicolor, or additional spatial dimensions.

Some theories beyond the Standard Model have an extended Higgs sector. For example, the minimal supersymmetric extension of the Standard Model (the so-called MSSM) contains two Higgs doublets. Also the Peccei-Quinn solution of the strong CP problem, which will be addressed in Chapter 11, relies on an extension of the Standard Model with two Higgs doublets. At present, it is not at all clear how many Higgs particles there are, and whether they are elementary or composite. All we know for sure is that there is a source of electroweak symmetry breaking. Without experimental guidance, it seems impossible to deduce the correct structure of the Higgs sector. The Standard Model assumes a minimal Higgs sector with just a single Higgs doublet.

In this section we consider an extension of the Standard Model by adding a second Higgs doublet $\widetilde{\Phi}$. We parameterize the two complex Higgs doublets as

$$\Phi(x) = \left(\begin{array}{c} \Phi^+(x) \\ \Phi^0(x) \end{array} \right) , \quad \widetilde{\Phi}(x) = \left(\begin{array}{c} \widetilde{\Phi}^0(x) \\ \widetilde{\Phi}^-(x) \end{array} \right) . \tag{3.9.1}$$

Under the $SU(2)_{\mathrm{L}}$ symmetry they transform as

$$\Phi'(x) = L\Phi(x) , \quad \widetilde{\Phi}'(x) = L\widetilde{\Phi}(x) , \tag{3.9.2}$$

and under $U(1)_Y$ as

$$\Phi'(x) = \exp\left(-\mathrm{i}\frac{g'}{2}\varphi\right)\Phi(x) , \quad \widetilde{\Phi}'(x) = \exp\left(\mathrm{i}\frac{g'}{2}\varphi\right)\widetilde{\Phi}(x) . \tag{3.9.3}$$

In addition to these symmetries, the extension of the Standard Model that we consider here has an additional $U(1)_{\mathrm{PQ}}$ symmetry — a so-called *Peccei-Quinn symmetry* — which acts as

$$\Phi'(x) = \exp\left(\mathrm{i}\alpha\right)\Phi(x) , \quad \widetilde{\Phi}'(x) = \exp\left(\mathrm{i}\alpha\right)\widetilde{\Phi}(x) . \tag{3.9.4}$$

The corresponding Lagrangian of the two Higgs doublet model takes the form

$$\mathcal{L}(\Phi, \partial_\mu\Phi, \widetilde{\Phi}, \partial_\mu\widetilde{\Phi}) = \frac{1}{2}\partial_\mu\Phi^\dagger\partial_\mu\Phi + \frac{1}{2}\partial_\mu\widetilde{\Phi}^\dagger\partial_\mu\widetilde{\Phi} + V(\Phi, \widetilde{\Phi}) . \tag{3.9.5}$$

There is no need to consider kinetic terms that mix the two scalar fields. If such terms were present, one could eliminate them by a field redefinition. However, in the potential $V(\Phi, \widetilde{\Phi})$ mixing terms may be present. The most general renormalizable potential invariant under $SU(2)_{\mathrm{L}}$, $U(1)_Y$, as well as $U(1)_{\mathrm{PQ}}$ is given by

$$\begin{aligned} V(\Phi, \widetilde{\Phi}) &= \frac{m^2}{2}|\Phi|^2 + \frac{\lambda}{4!}|\Phi|^4 + \frac{\widetilde{m}^2}{2}|\widetilde{\Phi}|^2 + \frac{\widetilde{\lambda}}{4!}|\widetilde{\Phi}|^4 + \frac{\kappa}{2}|\Phi^\dagger \widetilde{\Phi}|^2 \ , \\ |\Phi|^2 &= \Phi^{+*}\Phi^+ + \Phi^{0*}\Phi^0 \ , \quad |\widetilde{\Phi}|^2 = \widetilde{\Phi}^{0*}\widetilde{\Phi}^0 + \widetilde{\Phi}^{-*}\widetilde{\Phi}^- \ . \end{aligned} \tag{3.9.6}$$

In contrast to the Standard Model, the extended model does not have an additional $SU(2)_{\mathrm{R}}$ symmetry.

For $\kappa > 0$ the classical vacuum configurations obey $\Phi^\dagger \widetilde{\Phi} = 0$. For $m^2 < 0$, $\widetilde{m}^2 < 0$, one possible choice is

$$\Phi(x) = \begin{pmatrix} 0 \\ v \end{pmatrix} \ , \quad v = \sqrt{-\frac{6m^2}{\lambda}} \ , \quad \widetilde{\Phi}(x) = \begin{pmatrix} \widetilde{v} \\ 0 \end{pmatrix} \ , \quad \widetilde{v} = \sqrt{-\frac{6\widetilde{m}^2}{\widetilde{\lambda}}} \ . \tag{3.9.7}$$

This vacuum configuration is not invariant under either $SU(2)_{\mathrm{L}}$, $U(1)_Y$, or $U(1)_{\mathrm{PQ}}$. However, it is invariant against the $U(1)_{\mathrm{em}}$ subgroup of $SU(2)_{\mathrm{L}} \times U(1)_Y$, which acts as

$$\Phi'(x) = \begin{pmatrix} \exp(\mathrm{i}\varphi) & 0 \\ 0 & 0 \end{pmatrix} \Phi(x) \ , \quad \widetilde{\Phi}'(x) = \begin{pmatrix} 0 & 0 \\ 0 & \exp(-\mathrm{i}\varphi) \end{pmatrix} \widetilde{\Phi}(x) \ , \tag{3.9.8}$$

and will soon be identified as the symmetry of electromagnetism. Hence, the symmetry group $G = SU(2)_{\mathrm{L}} \times U(1)_Y \times U(1)_{\mathrm{PQ}}$ is spontaneously broken down to the subgroup $H = U(1)_{\mathrm{em}}$. According to the Goldstone theorem, in this case, there are $n_G - n_H = 3 + 1 + 1 - 1 = 4$ Nambu-Goldstone bosons. The additional fourth Nambu-Goldstone boson, which results from the spontaneous breakdown of the $U(1)_{\mathrm{PQ}}$ Peccei-Quinn symmetry, is known as the *axion*.

Let us again expand around the vacuum configuration by writing

$$\Phi(x) = \begin{pmatrix} \pi_1(x) + \mathrm{i}\pi_2(x) \\ v + \sigma(x) + \mathrm{i}\pi_3(x) \end{pmatrix} \ , \quad \widetilde{\Phi}(x) = \begin{pmatrix} \widetilde{v} + \widetilde{\sigma}(x) - \mathrm{i}\widetilde{\pi}_3(x) \\ -\widetilde{\pi}_1(x) + \mathrm{i}\widetilde{\pi}_2(x) \end{pmatrix} \ . \tag{3.9.9}$$

Up to quadratic order in the fluctuations, the potential then takes the form

$$\begin{aligned}
V(\Phi,\widetilde{\Phi}) &= \frac{m^2}{2}\left[(v+\sigma)^2+\pi_1^2+\pi_2^2+\pi_3^2\right]+\frac{\lambda}{4!}\left[(v+\sigma)^2+\pi_1^2+\pi_2^2+\pi_3^2\right]^2 \\
&+ \frac{\widetilde{m}^2}{2}\left[(\widetilde{v}+\widetilde{\sigma})^2+\widetilde{\pi}_1^2+\widetilde{\pi}_2^2+\widetilde{\pi}_3^2\right]+\frac{\widetilde{\lambda}}{4!}\left[(\widetilde{v}+\widetilde{\sigma})^2+\widetilde{\pi}_1^2+\widetilde{\pi}_2^2+\widetilde{\pi}_3^2\right]^2 \\
&+ \frac{\kappa}{2}|(\pi_1-\mathrm{i}\pi_2)(\widetilde{v}+\widetilde{\sigma}-\mathrm{i}\widetilde{\pi}_3)+(v+\sigma-\mathrm{i}\pi_3)(-\widetilde{\pi}_1+\mathrm{i}\widetilde{\pi}_2)|^2 \\
&\approx \frac{1}{2}\left(m^2+\frac{\lambda}{2}v^2\right)\sigma^2+\frac{1}{2}\left(\widetilde{m}^2+\frac{\widetilde{\lambda}}{2}\widetilde{v}^2\right)\widetilde{\sigma}^2 \\
&+ \frac{\kappa}{2}(v^2+\widetilde{v}^2)\left[\left(\frac{\widetilde{v}\pi_1-v\widetilde{\pi}_1}{\sqrt{v^2+\widetilde{v}^2}}\right)^2+\left(\frac{\widetilde{v}\pi_2-v\widetilde{\pi}_2}{\sqrt{v^2+\widetilde{v}^2}}\right)^2\right]+c\ , \qquad (3.9.10)
\end{aligned}$$

where c is once again an irrelevant constant. Indeed, there are four massive modes, *i.e.* four Higgs particles, σ, $\widetilde{\sigma}$, and

$$\rho_1=\frac{\widetilde{v}\pi_1-v\widetilde{\pi}_1}{\sqrt{v^2+\widetilde{v}^2}}\ , \quad \rho_2=\frac{\widetilde{v}\pi_2-v\widetilde{\pi}_2}{\sqrt{v^2+\widetilde{v}^2}}\ , \qquad (3.9.11)$$

with the corresponding mass squares

$$m_\sigma^2=\frac{\lambda}{3}v^2\ , \quad m_{\widetilde{\sigma}}^2=\frac{\widetilde{\lambda}}{3}\widetilde{v}^2\ , \quad m_{\rho_1}^2=m_{\rho_2}^2=\kappa(v^2+\widetilde{v}^2)\ , \qquad (3.9.12)$$

as well as four massless Nambu-Goldstone modes π_3, $\widetilde{\pi}_3$, and

$$\zeta_1=\frac{\widetilde{v}\pi_1+v\widetilde{\pi}_1}{\sqrt{v^2+\widetilde{v}^2}}\ , \quad \zeta_2=\frac{\widetilde{v}\pi_2+v\widetilde{\pi}_2}{\sqrt{v^2+\widetilde{v}^2}}\ . \qquad (3.9.13)$$

The modes σ, $\widetilde{\sigma}$, π_3, and $\widetilde{\pi}_3$ are neutral, whereas the modes $\rho_1\pm\mathrm{i}\rho_2$ and $\zeta_1\pm\mathrm{i}\zeta_2$ are charged under the unbroken subgroup $H=U(1)_{\mathrm{em}}$ of the symmetry $G=SU(2)_{\mathrm{L}}\times U(1)_Y\times U(1)_{\mathrm{PQ}}$.

Finally, let us construct the leading terms in the low-energy effective theory for the two Higgs doublet model. Following the general scheme, the fields describing the Nambu-Goldstone bosons parameterize the coset space $G/H=SU(2)_{\mathrm{L}}\times U(1)_Y\times U(1)_{\mathrm{PQ}}/U(1)_{\mathrm{em}}=SU(2)\times U(1)$, and hence take the form $U(x)\in SU(2)$ and $\exp(\mathrm{i}\theta(x))\in U(1)$. The leading terms of the effective Lagrangian are given by

$$\mathcal{L}(\partial_\mu V,\partial_\mu\theta) \;=\; \frac{F^2}{4}\mathrm{Tr}\left[\partial_\mu V^\dagger\partial_\mu V\right]+K\mathrm{Tr}\left[\partial_\mu V^\dagger\partial_\mu V\sigma^3\right]+\frac{\widetilde{F}^2}{2}\partial_\mu\theta\partial_\mu\theta$$

$$= \frac{F^2}{4}\mathrm{Tr}\left[\partial_\mu V^\dagger \partial_\mu V\right] + \frac{\widetilde{F}^2}{2}\partial_\mu\theta\partial_\mu\theta \ . \tag{3.9.14}$$

The term proportional to K seems to explicitly break the $SU(2)_\mathrm{R}$ symmetry down to $U(1)_Y$. However, this term simply vanishes. Consequently, despite the fact that there is no $SU(2)_\mathrm{R}$ symmetry in the two Higgs doublet model, at leading order the Lagrangian of the corresponding low-energy effective theory still has an $SU(2)_\mathrm{R}$ custodial symmetry. At higher order, on the other hand, $SU(2)_\mathrm{R}$ breaking terms do arise. Hence, the custodial symmetry is an *accidental* global symmetry. It arises only because no symmetry breaking terms exist in the leading low-energy Lagrangian.

The two Higgs doublet extension of the Standard Model was introduced by Roberto Peccei and Helen Quinn in 1977, in an attempt to solve the so-called strong CP-problem, which we will investigate in more detail in Chapter 11. In 1978 Steven Weinberg and Frank Wilczek realized independently that the spontaneous breakdown of the $U(1)_\mathrm{PQ}$ symmetry gives rise to a Nambu-Goldstone boson — the axion. Experimental axion searches have thus far been unsuccessful. Hence, it is still unclear whether the two Higgs doublet extension of the Standard Model is realized in Nature.

Chapter 4

From Superconductivity to Electroweak Gauge Bosons

In this chapter we introduce gauge fields mediating the electromagnetic and weak interactions. The weak interactions are responsible, for example, for the processes of radioactive decays. When *electroweak gauge fields* are included, the $SU(2)_\mathrm{L}$ symmetry — as well as the $U(1)_Y$ subgroup of $SU(2)_\mathrm{R}$ — turn into *local* symmetries. The electroweak $SU(2)_\mathrm{L} \times U(1)_Y$ gauge symmetry breaks spontaneously down to $U(1)_\mathrm{em}$ — the gauge group of electromagnetism. Due to the *Higgs mechanism* the W and Z gauge bosons become *massive*. The additional logitudinal polarization states of the three massive vector bosons, W^+, W^-, and Z are provided by three Nambu-Goldstone modes. One says that "the gauge bosons eat the Nambu-Goldstone bosons" and thus pick up a mass. The photon, on the other hand, remains massless as a consequence of the *unbroken* $U(1)_\mathrm{em}$ gauge symmetry of electromagnetism. The full gauge symmetry of the Standard Model is $SU(3)_\mathrm{c} \times SU(2)_\mathrm{L} \times U(1)_Y$, where the color gauge group $SU(3)_\mathrm{c}$ is associated with the strong interaction between quarks which is mediated by gluons. Since Higgs fields are color-neutral, before quarks are added, the gluons do not interact with Higgs bosons, W- and Z-bosons, or photons. We will add the gluons only later when we discuss the strong interaction.

To illustrate the basic ideas behind the Higgs mechanism, we first turn to a simpler model motivated by the condensed matter physics of superconductors — namely electrodynamics with a charged scalar field representing Cooper pairs. When Cooper pairs condense inside a superconductor, the $U(1)_{\mathrm{em}}$ gauge symmetry of electromagnetism undergoes the Higgs mechanism and the photon becomes massive.

4.1 Scalar Quantum Electrodynamics

We want to promote the global $U(1)$ symmetry discussed in Section 5.1 to a local one. This is a substantial enlargement of symmetry, since we proceed from one single symmetry parameter to one parameter at each space-time point. What we demand is a $U(1)$ invariance of the form

$$\Phi'(x) = \exp(\mathrm{i}e\varphi(x))\Phi(x) \ , \tag{4.1.1}$$

where $\varphi(x)$ is now a space-time dependent transformation parameter (which we assume to be differentiable). The potential is invariant already, $V(\Phi') = V(\Phi)$. The kinetic term, on the other hand, is not invariant as it stands, because

$$\partial_\mu\Phi'(x) = \exp(\mathrm{i}e\varphi(x))\left[\partial_\mu\Phi(x) + \mathrm{i}e\partial_\mu\varphi(x)\Phi(x)\right] \ . \tag{4.1.2}$$

In order to render it locally invariant, we must modify the derivative. To this end, we introduce a *gauge field* $A_\mu(x)$ and build a "covariant derivative"

$$\begin{aligned} D_\mu\Phi(x) &= \left[\partial_\mu - \mathrm{i}eA_\mu(x)\right]\Phi(x) \ , \\ D_\mu\Phi^*(x) &= \left[\partial_\mu + \mathrm{i}eA_\mu(x)\right]\Phi^*(x) \ . \end{aligned} \tag{4.1.3}$$

It should be noted that the covariant derivative D_μ takes different forms depending on the transformation properties of the field it acts on. For example, when D_μ acts on the complex conjugated field Φ^* the gauge field contribution in it also gets complex conjugated. The gauge field transforms such that the term $\partial_\mu\varphi$ in the covariant derivative is eliminated,

$$\begin{aligned} A'_\mu(x) &= A_\mu(x) + \partial_\mu\varphi(x) \ \Rightarrow \\ D_\mu\Phi'(x) &= \left[\partial_\mu - \mathrm{i}eA'_\mu(x)\right]\Phi'(x) = \exp(\mathrm{i}e\varphi(x))D_\mu\Phi(x) \ , \\ D_\mu\Phi^{*\prime}(x) &= \exp(-\mathrm{i}e\varphi(x))D_\mu\Phi^*(x) \ . \end{aligned} \tag{4.1.4}$$

Hence the operator D_μ is indeed gauge covariant. It can therefore be used to formulate a gauge invariant Lagrangian

$$\mathcal{L}(\Phi, \partial_\mu\Phi, A_\mu) = \frac{1}{2}D_\mu\Phi^* D_\mu\Phi + V(\Phi) \ . \tag{4.1.5}$$

The parameter e represents the electric charge of the scalar field, *i.e.* the strength of its coupling to A_μ. The anti-scalar, represented by the field Φ^*, has the opposite charge $-e$.

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

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) \ , \tag{4.1.6}$$

which is the obvious gauge invariant quantity to be built from first derivatives of $A_\mu(x)$,

$$\begin{aligned} F'_{\mu\nu}(x) &= \partial_\mu A'_\nu(x) - \partial_\nu A'_\mu(x) = \partial_\mu A_\nu(x) + \partial_\mu\partial_\nu\varphi(x) - \partial_\nu A_\mu(x) - \partial_\nu\partial_\mu\varphi(x) \\ &= F_{\mu\nu}(x) \ . \end{aligned} \tag{4.1.7}$$

The Lagrangian of the free electromagnetic field reads

$$\mathcal{L}(\partial_\mu A_\nu) = \frac{1}{4} F_{\mu\nu} F_{\mu\nu} = \frac{1}{2} \sum_{\mu>\nu} F_{\mu\nu} F_{\mu\nu} \ . \tag{4.1.8}$$

In the classical limit this Lagrangian leads to the inhomogeneous Maxwell equations[1]

$$\partial_\mu F_{\mu\nu} = 0 \ , \tag{4.1.9}$$

while the homogeneous Maxwell equations are automatically implied by the use of the 4-vector potential A_μ.

Thus the *total Lagrangian of scalar QED* takes the form

$$\boxed{\mathcal{L}(\Phi, \partial_\mu\Phi, A_\mu, \partial_\mu A_\nu) = \frac{1}{2} D_\mu\Phi^* D_\mu\Phi + V(\Phi) + \frac{1}{4} F_{\mu\nu} F_{\mu\nu}} \ . \tag{4.1.10}$$

It is *not* allowed to add an explicit mass term $\frac{m_\gamma^2}{2} A_\mu A_\mu$, because such a term would violate gauge invariance.

As in the case of the global $U(1)$ symmetry, we distinguish two cases:

• For $m^2 \geq 0$ the symmetry is unbroken, and we have a *Coulomb phase* with scalar particles of charge e and massless photons. In such a phase the

[1] In the general case this represents a current $\partial_\mu F_{\mu\nu} = j_\nu$, which — due to the anti-symmetry of $F_{\mu\nu}$ — obeys the continuity equation $\partial_\nu j_\nu = 0$.

electric charge is a conserved quantity. Indeed, in the vacuum of QED (and even of the full Standard Model) the $U(1)_{\rm em}$ symmetry of electrodynamics is realized in a Coulomb phase.

• Again, the broken phase (which corresponds to $m^2 < 0$ at the classical level) is particularly interesting. Inside a superconductor, the $U(1)_{\rm em}$ symmetry is spontaneously broken as a consequence of Cooper pair condensation. Since Cooper pairs are charged (they carry the charge $-2e$ of two electrons), their condensation implies that the vacuum itself contains an undetermined number of charges. As a consequence, electric charge is no longer locally conserved.[2] Once again, there are degenerate vacuum configurations, but they are now *related by gauge transformations*, as we see from Eqs. (4.1.1) and (4.1.4). Therefore they represent the same physical state. As a result, in contrast to systems with a spontaneously broken global symmetry, in a gauge theory "spontaneous symmetry breaking" does not lead to vacuum degeneracy. Strictly speaking, gauge symmetries cannot break spontaneously. In fact, they are not even symmetries of the physical world but merely redundancies in our theoretical description. Still, it is common practice to speak of "spontaneous gauge symmetry breaking". As we just did, in order to remind the reader of the subtleties related to this notion, also later we will always put "spontaneous gauge symmetry breaking" in inverted commas.

To take a closer look at this phase, it is helpful to fix the gauge, so that we obtain a reference point for an expansion. We choose the "unitary gauge"

$$\mathrm{Re}\,\Phi(x) = \phi_1(x) \geq 0 \ , \quad \mathrm{Im}\,\Phi(x) = \phi_2(x) = 0 \ . \tag{4.1.11}$$

Let us again investigate the fluctuations around the vacuum configuration $\phi_1(x) = v$. Due to gauge fixing, in this case we only deal with physical fluctuations, *i.e.*

$$\Phi(x) = v + \sigma(x) \ , \tag{4.1.12}$$

and thus there is no π-excitation. To $O(\sigma^2)$ we obtain

$$V(\Phi) \ = \ \frac{m^2}{2}(v+\sigma)^2 + \frac{\lambda}{4!}(v+\sigma)^4$$

[2]Of course, any real superconductor is a finite piece of material embedded in the Coulomb phase of the QED vacuum. Hence, the total charge of the whole superconductor still remains conserved.

$$\begin{aligned} &\approx \frac{m^2}{2}v^2 + m^2 v\sigma + \frac{m^2}{2}\sigma^2 + \frac{\lambda}{4!}(v^4 + 4v^3\sigma + 6v^2\sigma^2) \\ &= \frac{1}{2}(m^2 + \frac{\lambda}{2}v^2)\sigma^2 + c = \frac{\lambda}{6}v^2\sigma^2 + c \ . \end{aligned} \tag{4.1.13}$$

There is again a σ-particle with the same mass as in the case of the spontaneously broken global symmetry. However, the massless Nambu-Goldstone boson π has disappeared, since — as we mentioned above — the degeneracy of vacua is not physical any more.

What happened to the π degree of freedom? Let us consider the covariant kinetic term and expand it to second order in σ and A_μ,

$$\begin{aligned} \frac{1}{2}D_\mu\Phi^* D_\mu\Phi &= \frac{1}{2}\Big[(\partial_\mu + ieA_\mu)(v+\sigma)\Big]\ \Big[(\partial_\mu - ieA_\mu)(v+\sigma)\Big] \\ &= \frac{1}{2}(\partial_\mu\sigma + ieA_\mu v + ieA_\mu\sigma)(\partial_\mu\sigma - ieA_\mu v - ieA_\mu\sigma) \\ &\approx \frac{1}{2}\partial_\mu\sigma\partial_\mu\sigma + \frac{1}{2}e^2v^2 A_\mu A_\mu \ . \end{aligned} \tag{4.1.14}$$

Amazingly, we have obtained a *massive photon* with

$$\boxed{m_\gamma = ev} \ . \tag{4.1.15}$$

Therefore the missing degree of freedom (which was formerly identified as the π particle) has turned into an additional longitudinal polarization state of the photon.

This mechanism of mass generation is known as the *Higgs mechanism*. It is based on the "spontaneous breakdown" of a gauge symmetry. A phase in which the gauge symmetry is "spontaneously broken", so that the gauge bosons are massive, is called a *Higgs phase*. While the QED vacuum is in a Coulomb phase, inside a superconductor the $U(1)_{\mathrm{em}}$ gauge symmetry *is* "spontaneously broken", and the photon becomes massive. This mass can be measured, because it is related to the penetration depth of magnetic fields in the superconductor. This penetration falls off exponentially in proportion to $\exp(-m_\gamma r)$. One then identifies $1/m_\gamma$ as the *range* of the electromagnetic interaction. In the Coulomb phase, on the other hand, the electromagnetic interaction has an infinite range.

4.2 The Higgs Mechanism in the Electroweak Theory

Let us now turn to the electroweak gauge interactions in the Standard Model. Here we must promote the $SU(2)_{\mathrm{L}}$ symmetry as well as the $U(1)_Y$ subgroup of $SU(2)_{\mathrm{R}}$ to local symmetries. To begin with, we turn $SU(2)_{\mathrm{L}}$ into a gauge symmetry, *i.e.* we demand invariance of the Lagrangian against the gauge transformation

$$\Phi'(x) = L(x)\,\Phi(x)\ . \tag{4.2.1}$$

The potential $V(\Phi)$ is already invariant, but the kinetic term is not, because

$$\begin{aligned} \partial_\mu \Phi'(x) &= L(x)\partial_\mu \Phi(x) + \partial_\mu L(x)\Phi(x) \\ &= L(x)\left[\partial_\mu \Phi(x) + L(x)^\dagger \partial_\mu L(x)\Phi(x)\right]\ . \end{aligned} \tag{4.2.2}$$

As before, we want to compensate the additional term. For this purpose, we introduce a gauge field $W_\mu(x)$ and construct a covariant derivative of the form

$$D_\mu \Phi(x) = \left[\partial_\mu + W_\mu(x)\right]\Phi(x)\ . \tag{4.2.3}$$

The gauge field W_μ is a complex 2×2 matrix. In the kinetic term, the above covariant derivative is multiplied by

$$D_\mu \Phi(x)^\dagger = \partial_\mu \Phi(x)^\dagger + \Phi(x)^\dagger W_\mu^\dagger(x) = \partial_\mu \Phi(x)^\dagger - \Phi(x)^\dagger W_\mu \tag{4.2.4}$$

In the last step, we have taken W_μ to be *anti-Hermitian*, $W_\mu^\dagger = -W_\mu$. In this way we make sure that the kinetic term in the Lagrangian is real. In this form, W_μ is also a natural generalization of the term $\mathrm{i}eA_\mu$, which entered the covariant derivative in the gauging of a single complex scalar field (in the previous Section). Hence this gauge field can be written as

$$W_\mu(x) = \mathrm{i}g W_\mu^a(x)\frac{\sigma^a}{2}\ , \quad a = 1,2,3\ , \tag{4.2.5}$$

where σ^a are the Pauli matrices given in eq. (**??**) (which are Hermitian), and the factor $1/2$ is a convention. The parameter g is the gauge coupling constant; it characterises the strength of the coupling between the Higgs field and the gauge field W_μ.

For the behavior of the new matrix-valued gauge field under a gauge transformation we make the ansatz

$$W'_\mu(x) = L(x)\,[W_\mu(x) + \partial_\mu]\,L(x)^\dagger\ . \tag{4.2.6}$$

The virtue of this ansatz is that it leads to the simple relation

$$\begin{aligned} D_\mu \Phi'(x) &= \left[\partial_\mu + W'_\mu(x)\right] \Phi'(x) \\ &= L(x)\left[\partial_\mu \Phi(x) + L(x)^\dagger \partial_\mu L(x) \Phi(x)\right. \\ &+ \left. W_\mu(x) L(x)^\dagger L(x) \Phi(x) + \partial_\mu L(x)^\dagger L(x) \Phi(x)\right] \\ &= L(x)\,[\partial_\mu + W_\mu(x)]\,\Phi(x) = L(x) D_\mu \Phi(x)\ . \end{aligned} \tag{4.2.7}$$

Similarly we obtain

$$D_\mu \Phi'(x)^\dagger = D_\mu \Phi(x)^\dagger L(x)^\dagger\ . \tag{4.2.8}$$

Thus we arrive at the desired gauge invariant Lagrangian

$$\mathcal{L}(\Phi, \partial_\mu \Phi, W_\mu) = \frac{1}{2} D_\mu \Phi^\dagger D_\mu \Phi + V(\Phi)\ . \tag{4.2.9}$$

So far the gauge field is external. We still have to add its own kinetic term. The field strength tensor of a non-Abelian gauge field is given by

$$W_{\mu\nu} = D_\mu W_\nu - D_\nu W_\mu = \partial_\mu W_\nu - \partial_\nu W_\mu + [W_\mu, W_\nu]\ , \tag{4.2.10}$$

and it transforms as

$$W'_{\mu\nu}(x) = L(x) W_{\mu\nu}(x) L(x)^\dagger\ . \tag{4.2.11}$$

We see that it is natural to add the commutator term to $W_{\mu\nu}$, since it transforms in the same way as the other terms. Moreover, it is consistent to use the covariant derivative also for the formulation of the field strength. Hence we may consider this as the general form of a field strength. The case of a $U(1)$ gauge field that we discussed before in eq.(4.1.6) was just the special situation where the commutator vanishes. The presence of a commutator term in $W_{\mu\nu}$ has important consequences: in contrast to Abelian gauge fields, non-Abelian gauge fields are *charged* themselves; hence they interact among each other, even without other charged fields present.

In analogy to the Abelian gauge theory, eq. (4.1.8), we write

$$\mathcal{L}(W_{\mu\nu}) = \frac{1}{4g^2}\, W^a_{\mu\nu} W^a_{\mu\nu} = -\frac{1}{2g^2}\, \mathrm{Tr}\, W_{\mu\nu} W_{\mu\nu} \ , \tag{4.2.12}$$

which is indeed gauge invariant, and $W_{\mu\nu}(x) = \mathrm{i} g W^a_{\mu\nu}(x)\frac{\sigma^a}{2}$. The structure of non-Abelian gauge fields was first described in an unpublished letter of Wolfgang Pauli to Abraham Pais in 1953. The first paper introducing $SU(2)$ gauge theories is the ground-breaking work of Chen-Ning Yang and Richard Mills in 1954.

Thus far, we have limited the gauging to the $SU(2)_{\mathrm{L}}$ transformations, and therefore to transformations with the determinant 1. Now we want to gauge the extra $U(1)$ transformations related to the determinant. The group of these transformations is again $U(1)_Y$. The Higgs field then transforms as

$$\Phi'(x) = \exp\Big(-\mathrm{i}\frac{g'}{2}\varphi(x)\Big)\Phi(x) \ . \tag{4.2.13}$$

Here g' is a new coupling constant — the *weak hypercharge* (and the factor $-\frac{1}{2}$ is purely conventional as in eq.(4.2.5)). As we discussed in Section 5.2, the $U(1)_Y$ symmetry is actually a subgroup of $SU(2)_{\mathrm{R}}$ with

$$R(x) = \begin{pmatrix} \exp(-\mathrm{i}\frac{g'}{2}\varphi(x)) & 0 \\ 0 & \exp(\mathrm{i}\frac{g'}{2}\varphi(x)) \end{pmatrix} . \tag{4.2.14}$$

It should be emphasized again that only the $U(1)_Y$ subgroup and not the whole $SU(2)_{\mathrm{R}}$ symmetry is gauged. Gauging solely the $U(1)_Y$ subgroup implies an explicit breaking of the global $SU(2)_{\mathrm{R}}$ symmetry. Therefore we are not going to consider the remaining two generators of $SU(2)_{\mathrm{R}}$. The $U(1)_Y$ gauge field transforms as

$$B'_\mu(x) = B_\mu(x) + \partial_\mu\varphi(x) \ . \tag{4.2.15}$$

This new gauge field contributes an additional term to the covariant derivative,

$$\begin{aligned} D_\mu\Phi(x) &= \left[\partial_\mu + W_\mu(x) + \mathrm{i}\frac{g'}{2}B_\mu(x)\right]\Phi(x) \\ &= \left[\partial_\mu + \mathrm{i}W^a_\mu(x)\frac{\sigma^a}{2} + \mathrm{i}\frac{g'}{2}B_\mu(x)\right]\begin{pmatrix} \Phi^+(x) \\ \Phi^0(x) \end{pmatrix} , \end{aligned} \tag{4.2.16}$$

which is anti-Hermitian as well. We use again the Abelian gauge invariant field strength

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \ , \tag{4.2.17}$$

to add another pure gauge term, and we arrive at the total Lagrangian

$$\boxed{\mathcal{L}(\Phi, W_\mu, B_\mu) = \frac{1}{2} D_\mu \Phi^\dagger D_\mu \Phi + V(\Phi) - \frac{1}{2g^2} \mathrm{Tr}(W_{\mu\nu} W_{\mu\nu}) + \frac{1}{4} B_{\mu\nu} B_{\mu\nu}} \ . \tag{4.2.18}$$

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

$$\Phi(x) = \left(\begin{array}{c} 0 \\ v \end{array} \right), \ v \in \mathbf{R}_+ \ . \tag{4.2.19}$$

The vacuum state (4.2.19) is invariant under $U(1)$ gauge transformations of the type

$$\Phi'(x) = \left(\begin{array}{cc} \exp(\mathrm{i}e\varphi(x)) & 0 \\ 0 & 1 \end{array} \right) \Phi(x) \ , \tag{4.2.20}$$

which have a $U(1)_Y$ hypercharge part, along with a diagonal $SU(2)_{\mathrm{L}}$ part,

$$\left(\begin{array}{cc} \exp(\mathrm{i}e\varphi) & 0 \\ 0 & 1 \end{array} \right) = \left(\begin{array}{cc} \exp(\mathrm{i}e\varphi/2) & 0 \\ 0 & \exp(\mathrm{i}e\varphi/2) \end{array} \right) \left(\begin{array}{cc} \exp(\mathrm{i}e\varphi/2) & 0 \\ 0 & \exp(-\mathrm{i}e\varphi/2) \end{array} \right) . \tag{4.2.21}$$

Hence the choice of the vacuum state does not "break" the $SU(2)_{\mathrm{L}} \times U(1)_Y$ symmetry completely. Instead, there is a remaining $U(1)$ symmetry, which we denote as $U(1)_{\mathrm{em}}$, because we will soon identify it with the electromagnetic gauge group. Since that symmetry remains *unbroken*, despite the Higgs mechanism, there will be one massless gauge boson — the photon. All other gauge bosons "eat up" a Nambu-Goldstone boson and become massive. To see this, we consider again the fluctuations in the unitary gauge,

$$\Phi(x) = \left(\begin{array}{c} 0 \\ v + \sigma(x) \end{array} \right) \ . \tag{4.2.22}$$

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

$$\frac{1}{2} D_\mu \Phi^\dagger D_\mu \Phi = \frac{1}{2} \left| \left(\partial_\mu + \mathrm{i} g W_\mu^a \frac{\sigma^a}{2} + \mathrm{i} \frac{g'}{2} B_\mu \right) \left(\begin{array}{c} 0 \\ v + \sigma \end{array} \right) \right|^2$$

$$= \frac{1}{2} \partial_\mu \sigma \partial_\mu \sigma + \frac{(v+\sigma)^2}{2} \ (0,1) \ \left[\left(g W_\mu^a \frac{\sigma^a}{2} + \frac{g'}{2} B_\mu \right) \left(g W_\mu^b \frac{\sigma^b}{2} + \frac{g'}{2} B_\mu \right) \right] \left(\begin{array}{c} 0 \\ 1 \end{array} \right)$$

$$= \frac{1}{2}\partial_\mu\sigma\partial_\mu\sigma + \frac{1}{8}(v+\sigma)^2 \Big[g^2 W_\mu^1 W_\mu^1 + g^2 W_\mu^2 W_\mu^2 + (gW_\mu^3 - g' B_\mu)(gW_\mu^3 - g' B_\mu)\Big] \ . \tag{4.2.23}$$

In addition, we have the usual potential term

$$V(\Phi) = \frac{m^2}{2}(v+\sigma)^2 + \frac{\lambda}{4}(v+\sigma)^4 = -m^2\sigma^2 + \dots \ , \tag{4.2.24}$$

hence there is once more a Higgs particle with

$$m_\sigma^2 = -2m^2 \ . \tag{4.2.25}$$

Moreover, there are two W-*bosons* of mass

$$m_W = \frac{1}{2} g v \ . \tag{4.2.26}$$

Furthermore, we introduce the linear combination[3]

$$Z_\mu = \frac{gW_\mu^3 - g' B_\mu}{\sqrt{g^2 + g'^2}} \ , \tag{4.2.27}$$

which represents the Z-*boson* with the mass

$$m_Z = \frac{1}{2}\sqrt{g^2 + g'^2}\ v \ . \tag{4.2.28}$$

The remaining orthonormal linear combination

$$A_\mu = \frac{g' W_\mu^3 + g B_\mu}{\sqrt{g^2 + g'^2}} \tag{4.2.29}$$

remains massless and describes the *photon*.

We introduce the *Weinberg angle* (or weak mixing angle) $\theta_{\rm W}$ to write down these linear combinations as

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_{\rm W} & \sin\theta_{\rm W} \\ -\sin\theta_{\rm W} & \cos\theta_{\rm W} \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} , \tag{4.2.30}$$

[3]In the space spanned by the basis vectors W_μ^3 and B_μ we observe in eq. (4.2.23) the doublet $(g, -g')$, which we still normalise.

such that

$$\frac{g}{\sqrt{g^2 + g'^2}} = \cos\theta_\mathrm{W} \ , \qquad \frac{g'}{\sqrt{g^2 + g'^2}} = \sin\theta_\mathrm{W} \ , \tag{4.2.31}$$

and therefore

$$\frac{m_W}{m_Z} = \frac{g}{\sqrt{g^2 + g'^2}} = \cos\theta_\mathrm{W} \ . \tag{4.2.32}$$

The W- and Z-bosons have indeed been discovered in the UA1 and UA2 high energy experiments at the Super Proton Synchrotron (SPS) accelerator at CERN in 1983. The *experimental* values for the masses are

$$m_W \simeq 80.399(23)\ \mathrm{GeV}\ , \quad m_Z \simeq 91.1876(21)\ \mathrm{GeV} \quad \Rightarrow \quad \sin^2\theta_\mathrm{W} \simeq 0.23116(13)\ . \tag{4.2.33}$$

There are a number of ways to measure $\sin^2\theta_\mathrm{W}$ in high energy experiments, and the results based on different methods agree with the value obtained from the ratio m_W/m_Z within the errors. This is a nice confirmation of the consistency of the Standard Model. On the down-side, θ_W is one of the parameters which are completely free in the Standard Model — a prediction for its value would require a superior theory.[4]

The coupling constant of the photon is the charge e. On the other hand, the corresponding covariant derivative of the scalar field reads

$$\begin{aligned}
D_\mu\Phi &= \left[\partial_\mu + \mathrm{i}gW^1_\mu\frac{\sigma^1}{2} + \mathrm{i}gW^2_\mu\frac{\sigma^2}{2} + \mathrm{i}gW^3_\mu\frac{\sigma^3}{2} + \mathrm{i}\frac{g'}{2}B_\mu\right]\begin{pmatrix}\Phi_+\\ \Phi_0\end{pmatrix} \\
&= \left[\partial_\mu + \mathrm{i}gW^1_\mu\frac{\sigma^1}{2} + \mathrm{i}gW^2_\mu\frac{\sigma^2}{2}\right. \\
&+ \left.\frac{\mathrm{i}}{2}\begin{pmatrix} gW^3_\mu + g'B_\mu & 0 \\ 0 & -gW^3_\mu + g'B_\mu\end{pmatrix}\right]\begin{pmatrix}\Phi_+\\ \Phi_0\end{pmatrix} \\
&= \left[\partial_\mu + \mathrm{i}gW^1_\mu\frac{\sigma^1}{2} + \mathrm{i}gW^2_\mu\frac{\sigma^2}{2}\right. \\
&+ \left.\mathrm{i}\begin{pmatrix} \frac{g^2-g'^2}{2\sqrt{g^2+g'^2}}Z_\mu + \frac{gg'}{\sqrt{g^2+g'^2}}A_\mu & 0 \\ 0 & \sqrt{g^2+g'^2}Z_\mu\end{pmatrix}\right]\begin{pmatrix}\Phi_+\\ \Phi_0\end{pmatrix} .
\end{aligned} \tag{4.2.34}$$

[4]The same holds for the individual values of m_W and m_Z.

We can now read off the electric charge as

$$e = \frac{gg'}{\sqrt{g^2 + g'^2}} \ . \tag{4.2.35}$$

We see that indeed only Φ_+ couples to the electromagnetic field A_μ. It has charge e, while Φ_0 is neutral (as we anticipated in footnote 3 of Chapter 5). At the same time, we see that the Z-boson is electrically neutral, hence it is often denoted as Z^0.[5] The rôle of the W-bosons in this respect will be illuminated in the next Chapter, when we introduce electroweak couplings between fermions.

4.3 Accidental Custodial Symmetry

Let us again consider the matrix representation of the Higgs field

$$\mathbf{\Phi}(x) = \left(\begin{array}{cc} \Phi^0(x)^* & \Phi^+(x) \\ -\Phi^+(x)^* & \Phi^0(x) \end{array} \right) \ , \tag{4.3.1}$$

which would transforms as

$$\mathbf{\Phi}'(x) = L(x)\mathbf{\Phi}(x)R(x)^\dagger \ , \quad L(x) \in SU(2)_{\mathrm{L}} \ , \quad R(x) \in SU(2)_{\mathrm{R}} \tag{4.3.2}$$

under $SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}}$ gauge transformations. As we have seen in Setion 6.2, only the $U(1)_Y$ subgroup of $SU(2)_{\mathrm{R}}$ is gauged in the Standard Model. Still, one could also imagine to turn the full $SU(2)_{\mathrm{R}}$ symmetry into a gauge symmetry. In that case, the corresponding covariant derivative would take the form

$$D_\mu \mathbf{\Phi}(x) = \partial_\mu \mathbf{\Phi}(x) + W_\mu(x)\mathbf{\Phi}(x) - \mathbf{\Phi}(x)X_\mu(x) \ . \tag{4.3.3}$$

Here $X_\mu(x) = \mathrm{i}g' X_\mu^a(x)\frac{\sigma^a}{2}$ is a hypothetical non-Abelian gauge field that transforms as

$$X'_\mu(x) = R(x)\left[X_\mu + \partial_\mu\right] R(x)^\dagger \tag{4.3.4}$$

under $SU(2)_{\mathrm{R}}$ gauge transformations. Since only the $U(1)_Y$ subgroup of $SU(2)_{\mathrm{R}}$ is gauged in the Standard Model, the hypothetical non-Abelian

[5] Flipping the sign of either g or g' changes the sign of the electric charge e of Φ_+. However, such sign flips do not affect the coupling of the Higgs field to Z_μ, which suggests that the Z-boson is electrically neutral.

gauge field X_μ is then reduced to the Abelian Standard Model gauge field B_μ, *i.e.*

$$X_\mu(x) = \mathrm{i}g' B_\mu(x) \frac{\sigma^3}{2} \ . \tag{4.3.5}$$

Before the $U(1)_Y$ subgroup of $SU(2)_\mathrm{R}$ is gauged (or equivalently when one puts $g' = 0$), $SU(2)_\mathrm{R}$ is an exact global symmetry, known as the custodial symmetry. Once $U(1)_Y$ is gauged, the custodial symmetry is explicitly violated and thus turns into an approximate global symmetry.

Gauge theories contain redundant unphysical degrees of freedom which do not affect the physics due to gauge invariance. Hence, in order to maintain only the physically relevant degrees of freedom, gauge symmetries must not be broken explicitly. Global symmetries, on the other hand, are usually only approximate and arise due to some hierarchy of energies scales, whose origin may or may not be understood. Let us now ask whether we understand the origin of the approximate custodial symmetry. In particular, we can ask whether there may be other sources of custodial symmetry breaking besides the weak $U(1)_Y$ gauge interactions. While such symmetry breaking terms can always be constructed using sufficiently many derivatives or field values, here we limit ourselves to perturbatively renormalizable interactions, which are the ones that dominate the physics at low energies. Since

$$\mathbf{\Phi}^\dagger \mathbf{\Phi} = \mathbf{\Phi} \mathbf{\Phi}^\dagger = |\Phi^0|^2 + |\Phi^+|^2 \tag{4.3.6}$$

is proportional to the unit-matrix, one cannot construct any $SU(2)_\mathrm{L} \times U(1)_Y$-invariant terms without derivatives that explicitly break the custodial $SU(2)_\mathrm{R}$ symmetry. For example, the terms $\mathrm{Tr}\left[\mathbf{\Phi}^\dagger \mathbf{\Phi} \sigma^3\right]$ and $\mathrm{Tr}\left[\mathbf{\Phi}^\dagger \mathbf{\Phi} \mathbf{\Phi}^\dagger \mathbf{\Phi} \sigma^3\right]$ simply vanish, and

$$\mathrm{Tr}\left[\mathbf{\Phi}^\dagger \mathbf{\Phi} \sigma^3 \mathbf{\Phi}^\dagger \mathbf{\Phi} \sigma^3\right] = 2\left(|\Phi^0|^2 + |\Phi^+|^2\right)^2 \tag{4.3.7}$$

just reduces to the standard $SU(2)_\mathrm{L} \times SU(2)_\mathrm{R}$-invariant quartic self-coupling. Using two covariant derivatives one can also construct the term $\mathrm{Tr}\left[D_\mu \mathbf{\Phi}^\dagger D_\mu \mathbf{\Phi} \sigma^3\right]$, which may seem to explicitly break $SU(2)_\mathrm{L} \times SU(2)_\mathrm{R}$ down to $SU(2)_\mathrm{L} \times U(1)_Y$. If this were indeed the case, this term should also be included in the Standard Model Lagrangian with an adjustable prefactor. If this prefactor would not be unnaturally small, the custodial symmetry should be strongly explicitly broken and would not even remain a useful approximate symmetry. Only if the prefactor of such a term would be small (perhaps due to some not yet understood hierarchy of energy scales), the symmetry would remain only weakly broken. It would then be puzzling why the symmetry is not more strongly broken. Interestingly, no such puzzle exists for the custodial symmetry of the Standard Model. In particular, although this may not be obvious, one can show that the term $\mathrm{Tr}\left[D_\mu \mathbf{\Phi}^\dagger D_\mu \mathbf{\Phi} \sigma^3\right]$ again simply vanishes. Indeed, besides the gauge coupling g', in the gauge-Higgs sector there is no other renormalizable interaction that explicitly breaks the custodial symmetry.[6]. One says that the custodial symmetry is *accidental*. It simply arises because g' is relatively small and no other renormalizable symmetry breaking terms exist.

4.4 Lattice Gauge-Higgs Models

The previous discussion of gauge theories was essentially at a classical level. The quantization of gauge theories is a delicate issue. In Appendix B the simplest gauge theory — an Abelian theory of free photons — is quantized canonically. However, in order to avoid subtleties related to Dirac's quantization with "first and second class constraints", we have already slightly simplified the presentation. When non-Abelian gauge fields are concerned, canonical quantization becomes even more complicated. Our method of choice, instead, is the quantization using the functional integral. In perturbation theory, the quantization of non-Abelian gauge fields using the functional integral requires gauge fixing, which leads to the introduction of "Faddeev-Popov ghost fields". This is a non-trivial procedure, which is well explained in the textbook literature. Here we follow a non-perturbative approach to the problem by regularizing the theory on a space-time lattice. Lattice gauge theories were introduced by Franz Wegner in the context of classical statistical mechanics and by Kenneth Wilson, as well as independently by Jan Smit, in the context of quantum field theory. Non-Abelian lattice gauge theories do not require gauge fixing and are thus conceptually simpler than their continuum counterparts, usually treated with dimensional regularization. As was shown analytically by Thomas Reisz, in perturbation theory lattice gauge theories define the same continuum limit as the dimensional regularization. In contrast to perturbative approaches,

[6]In Chapter 9 we will couple the Higgs field to leptons and quarks, which gives rise to additional terms that also break the custodial symmetry

lattice gauge theory can address physics at finite and even strong coupling. In particular, when we will discuss the strong interaction in Chapters 9 and 10, we will make use of the lattice regularization. Lattice QCD has become a quantitative tool that allows us to compute the properties of strongly interacting particles using Monte Carlo simulations. In this section, we use the lattice regularization to investigate the *phase structure* of Abelian and non-Abelian gauge theories. We will encounter a *Coulomb phase* with massless photons, as well as *Higgs phases* with massive gauge bosons. In addition, there are *confined phases* which may or may not be distinguishable from Higgs phases. In particular, in the Standard Model, the weakly coupled Higgs phase is analytically connected with a strongly coupled confined phase. Hence, in this case Higgs and confined phases are indistinguishable.

Let us first discuss the lattice version of scalar electrodynamics; we considered its continuum version before in Section 6.1. Then there is a complex scalar field $\Phi_x \in \mathbb{C}$ defined at the sites x of a 4-dimensional hyper-cubic lattice with spacing a. In addition, there is an Abelian lattice gauge field $A_{x,\mu} \in \mathbb{R}$, which is naturally defined on the links (x,μ) connecting neighboring lattice sites x and $x+\hat{\mu}$ (where $\hat{\mu}$ is a vector of length a pointing in the μ-direction). The gauge transformations φ_x are defined at the lattice sites x and act as

$$\Phi'_x = \exp(\mathrm{i}e\varphi_x)\Phi_x \ , \quad A'_{\mu,x} = A_{\mu,x} - \frac{\varphi_{x+\hat{\mu}} - \varphi_x}{a} \ . \tag{4.4.1}$$

In the continuum limit $a \to 0$ the second equation turns into the continuum relation $A'_\mu(x) = A_\mu(x) - \partial_\mu \varphi(x)$. The corresponding field strength

$$F_{\mu\nu,x} = \frac{A_{\nu,x+\hat{\mu}} - A_{\nu,x}}{a} - \frac{A_{\mu,x+\hat{\nu}} - A_{\mu,x}}{a} \ , \tag{4.4.2}$$

is naturally associated with the elementary lattice *plaquettes*. Just as in the continuum, on the lattice the Abelian field strength is gauge invariant, *i.e.* $F'_{\mu\nu,x} = F_{\mu\nu,x}$. In the continuum limit one recovers $F_{\mu\nu,x} \to \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) = F_{\mu\nu}(x)$. Let us also define a *parallel transporter* link variable

$$U_{\mu,x} = \exp(\mathrm{i}eA_{\mu,x}a) \ , \tag{4.4.3}$$

which transforms as

$$U'_{\mu,x} = \exp(\mathrm{i}e\varphi_x) U_{\mu,x} \exp(-\mathrm{i}e\varphi_{x+\hat{\mu}}) \ . \tag{4.4.4}$$

On the lattice, we distinguish left- and right-handed covariant derivatives, which are defined as

$$D^{\mathrm{L}}_\mu \Phi_x = \frac{U_{\mu,x}\Phi_{x+\hat\mu} - \Phi_x}{a} \ , \quad D^{\mathrm{R}}_\mu \Phi_{x+\hat\mu} = \frac{\Phi_{x+\hat\mu} - U^*_{\mu,x}\Phi_x}{a} \ . \tag{4.4.5}$$

Under gauge transformations they transform as

$$D^{\mathrm{L}}_\mu \Phi'_x = \exp(\mathrm{i}e\varphi_x) D^{\mathrm{L}}_\mu \Phi_x \ , \quad D^{\mathrm{R}}_\mu \Phi'_{x+\hat\mu} = \exp(\mathrm{i}e\varphi_{x+\hat\mu}) D^{\mathrm{R}}_\mu \Phi_{x+\hat\mu} \ . \tag{4.4.6}$$

The two derivatives are related by

$$D^{\mathrm{L}}_\mu \Phi_x = U_{\mu,x} D^{\mathrm{R}}_\mu \Phi_{x+\hat\mu} \ , \tag{4.4.7}$$

such that

$$D^{\mathrm{L}}_\mu \Phi^*_x D^{\mathrm{L}}_\mu \Phi_x = D^{\mathrm{R}}_\mu \Phi^*_{x+\hat\mu} D^{\mathrm{R}}_\mu \Phi_{x+\hat\mu} = \frac{1}{a^2}\left[|\Phi_{x+\hat\mu}|^2 + |\Phi_x|^2 - 2\mathrm{Re}\left(\Phi^*_x U_{\mu,x}\Phi_{x+\hat\mu}\right)\right] \ . \tag{4.4.8}$$

The Euclidean lattice action of scalar QED is given by

$$S[\Phi, A_\mu] = \sum_x a^4 \left[\frac{1}{2} D^{\mathrm{L}}_\mu \Phi^*_x D^{\mathrm{L}}_\mu \Phi_x + \frac{m^2}{2}|\Phi_x|^2 + \frac{\lambda}{4!}|\Phi_x|^4 + \frac{1}{4} F_{\mu\nu,x} F_{\mu\nu,x}\right] \ . \tag{4.4.9}$$

In order to make the regularized functional integral finite, in this so-called non-compact formulation of scalar lattice QED, one must fix the gauge.[7] Here we choose the Lorenz gauge $\partial_\mu A_\mu(x) = 0$, whose lattice version takes the form

$$\delta A_x = \sum_\mu \frac{1}{a}\left(A_{\mu,x+\hat\mu} - A_{\mu,x}\right) = 0 \ . \tag{4.4.10}$$

The resulting functional integral is then given by

$$Z = \int \mathcal{D}\Phi \int \mathcal{D}A_\mu \ \exp(-S[\Phi, A_\mu]) \prod_x \delta(\delta A_x) \ , \tag{4.4.11}$$

where the δ-function enforces the Lorenz gauge condition. The measures of the functional integration over the scalar and gauge field configurations are given by

$$\int \mathcal{D}\Phi = \prod_x \int_{\mathbb{C}} d\Phi_x \ , \quad \int \mathcal{D}A_\mu = \prod_{x,\mu} \int_{\mathbb{R}} dA_{\mu,x} \ . \tag{4.4.12}$$

[7]Gauge fixing is unnecessary in the compact formulation in terms of link variables $U_{\mu,x}$ which is also used in non-Abelian gauge theories. This will be discussed below.

The phase diagram of the lattice model, which has been obtained using both analytic (cite Florian Nill) and numerical methods, is illustrated schematically in Figure ??? for a fixed value of λ. At sufficiently negative values of m^2 and sufficiently small values of e, there is a Higgs phase with a massive photon which is separated from a Coulomb phase in which the photon is massless. The two phases are separated by a first order phase transition (cf. Appendix ???) which becomes second order at $e \to 0$. Near this *critical point* one can take a continuum limit of scalar QED, which is likely to be a trivial (i.e. non-interacting) theory.

Let us now turn to *non-Abelian* lattice gauge theories applied to the gauge-Higgs sector of the Standard Model. For simplicity, we gauge only the $SU(2)_{\mathrm{L}}$ and not also the $U(1)_Y$ symmetry. On the lattice, the Higgs field is again a complex doublet

$$\Phi_x = \left(\begin{array}{c} \Phi_{+,x} \\ \Phi_{0,x} \end{array} \right) , \tag{4.4.13}$$

which is associated with the lattice sites x. The non-Abelian lattice gauge field is defined in terms of parallel transporter link variables $U_{\mu,x}$ which are 2×2 matrices taking values in the gauge group $SU(2)_{\mathrm{L}}$. Unlike in the non-compact lattice formulation of scalar QED, one does not introduce a lattice variant of the non-Abelian vector potential $W_\mu(x)$. Still, in the classical continuum limit $a \to 0$ one can identify

$$U_{\mu,x} = \exp(W_\mu(x)a) = \exp\left(\mathrm{i}gW_\mu^a(x)\frac{\sigma^a}{2}a\right) . \tag{4.4.14}$$

Under non-Abelian lattice gauge transformations $L_x \in SU(2)_{\mathrm{L}}$ the fields transform as

$$\Phi_x' = L_x \Phi_x \ , \quad U_{\mu,x}' = L_x U_{\mu,x} L_{x+\hat\mu}^\dagger \ . \tag{4.4.15}$$

Just as in the Abelian theory, the covariant left- and right-derivatives are given by

$$D_\mu^{\mathrm{L}} \Phi_x = \frac{U_{\mu,x}\Phi_{x+\hat\mu} - \Phi_x}{a} \ , \quad D_\mu^{\mathrm{R}} \Phi_{x+\hat\mu} = \frac{\Phi_{x+\hat\mu} - U_{\mu,x}^\dagger \Phi_x}{a} \ , \tag{4.4.16}$$

which now transform as

$$D_\mu^{\mathrm{L}} \Phi_x' = L_x D_\mu^{\mathrm{L}} \Phi_x \ , \quad D_\mu^{\mathrm{R}} \Phi_{x+\hat\mu}' = L_x D_\mu^{\mathrm{R}} \Phi_{x+\hat\mu} \ . \tag{4.4.17}$$

In analogy to eq.(4.4.7), one obtains

$$D^{\rm L}_\mu \Phi_x = U_{\mu,x} D^{\rm R}_\mu \Phi_{x+\hat\mu} \ , \tag{4.4.18}$$

such that

$$\begin{aligned} &D^{\rm L}_\mu \Phi^\dagger_x D^{\rm L}_\mu \Phi_x = D^{\rm R}_\mu \Phi^\dagger_{x+\hat\mu} D^{\rm R}_\mu \Phi_{x+\hat\mu} = \\ &\frac{1}{a^2}\left[\Phi^\dagger_{x+\hat\mu}\Phi_{x+\hat\mu} + \Phi^\dagger_x \Phi_x - \Phi^\dagger_x U_{\mu,x}\Phi_{x+\hat\mu} - \Phi^\dagger_{x+\hat\mu} U^\dagger_{\mu,x}\Phi_x\right] \ . \end{aligned} \tag{4.4.19}$$

The lattice action of the $SU(2)_{\rm L}$-invariant gauge-Higgs model takes the form

$$\begin{aligned} S[\Phi, U_\mu] &= \sum_x a^4 \left[\frac{1}{2} D^{\rm L}_\mu \Phi^\dagger_x D^{\rm L}_\mu \Phi_x + \frac{m^2}{2}\Phi^\dagger_x \Phi_x + \frac{\lambda}{4!}|\Phi^\dagger_x \Phi_x|^2\right. \\ &+ \left.\frac{1}{4g^2a^2}\left(2 - {\rm Tr}\left(U_{\mu,x} U_{\nu,x+\hat\mu} U^\dagger_{\mu,x+\hat\nu} U^\dagger_{\nu,x}\right)\right)\right] \ . \end{aligned} \tag{4.4.20}$$

The last term is the kinetic and self-interaction term of the W-bosons. It is built from a product of link parallel transporters around an elementary lattice plaquette. It is instructive to show that this term is gauge invariant and that it turns into ${\rm Tr}(W_{\mu\nu}W_{\mu\nu})/4g^2$ in the continuum limit. The functional integral describing the $SU(2)_{\rm L}$ gauge-Higgs model is given by

$$Z = \int \mathcal{D}\Phi \int \mathcal{D}U_\mu \ \exp(-S[\Phi, U_\mu]) \ . \tag{4.4.21}$$

In this case the measures of the functional integrations are given by

$$\int \mathcal{D}\Phi = \prod_x \int_{\mathbb{C}^2} d\Phi_x \ , \quad \int \mathcal{D}U_\mu = \prod_{x,\mu} \int_{SU(2)} dU_{\mu,x} \exp(-S[\Phi, U_\mu]) \ . \tag{4.4.22}$$

Here $dU_{\mu,x}$ is the so-called *Haar measure*, which is invariant under gauge transformations both on the left and on the right end of the link, *i.e.*

$$dU'_{\mu,x} = L_x dU_{\mu,x} L^\dagger_{x+\hat\mu} \ . \tag{4.4.23}$$

The group manifold of $SU(2)$ is a sphere S^3 with a 3-dimensional surface. The Haar measure of $SU(2)$ is just the natural isotropic measure on S^3.

The phase diagram of the lattice gauge-Higgs model, which has been obtained using numerical simulations, is illustrated schematically in Figure ??? for a fixed value of λ. For sufficiently negative m^2 and sufficiently small g, there is a Higgs phase with a massive W-boson. However, unlike in scalar QED, there is no massless Coulomb phase. Instead, 4-dimensional non-Abelian gauge theories have a confined phase.

4.5 From Electroweak to Grand Unification

This section discusses physics beyond the Standard Model and may be skipped in a first reading. At this point we return to continuum notation.

As we have seen, the gauge groups $SU(2)_{\mathrm{L}}$ and $U(1)_Y$ give rise to two distinct gauge couplings g and g'. Hence, in the Standard Model the electroweak interactions are not truly unified. In the next chapter we will also include the strong interaction with the gauge group $SU(3)_{\mathrm{c}}$ which is associated with yet another gauge coupling g_s. Hence, the full gauge group of the Standard Model $SU(3)_{\mathrm{c}} \times SU(2)_{\mathrm{L}} \times U(1)_Y$ has three gauge couplings. In the framework of Grand Unified Theories (GUT), which are an extension of the Standard Model, one embeds the electroweak and strong interactions in one simple gauge group (*e.g.* $SU(5)$, $SO(10)$), or the exceptional group $E(6)$, which leads to a relation between g, g', and g_s.

The symmetries $SU(5)$ or $SO(10)$ are too large to be realized at low temperatures. They must be spontaneously broken to the $SU(3)_{\mathrm{c}} \times SU(2)_{\mathrm{L}} \times U(1)_Y$ symmetry of the Standard Model. In Grand Unified Theories this happens at temperatures about 10^{14} GeV, which were realized in the Universe about 10^{-34} sec after the Big Bang. At present (and in the foreseeable future) these energy scales cannot be probed experimentally. Hence, we now rely on theoretical arguments, and sometimes on speculation.

To illustrate the idea behind GUTs, let us first unify the electroweak gauge interactions by embedding $SU(2)_{\mathrm{L}} \times U(1)_Y$ into one single gauge group. Since the group $SU(2)_{\mathrm{L}} \times U(1)_Y$ has two commuting generators, *i.e.* its rank is $1 + 1 = 2$, the embedding unified group must also have a rank of at least 2. There are two so-called "simple" Lie groups of rank 2 — the special unitary group $SU(3)$ and the exceptional group $G(2)$, which contains $SU(3)$ as a subgroup. Hence, the minimal unifying group that contains $SU(2)_{\mathrm{L}} \times U(1)_Y$ is $SU(3)$ (not to be confused with the color gauge group $SU(3)_{\mathrm{c}}$), which has 8 generators. When the electroweak interactions of the Standard Model are embedded in $SU(3)$, half of the gauge bosons can be identified with known particles: $SU(2)_{\mathrm{L}}$ has 3 W-bosons, and $U(1)_Y$ has one B-boson which, together with W^3, forms the Z-boson and the photon. The remaining 4 gauge bosons of $SU(3)$ are new hypothetical particles, which we call X and Y. In order to make these particles heavy, the $SU(3)$

symmetry must be spontaneously broken to $SU(2)_{\rm L} \times U(1)_Y$. Again, this is achieved via the Higgs mechanism, in this case using an 8-component scalar field transforming under the adjoint representation of $SU(3)$. We write

$$\mathbf{\Phi}(x) = \Phi_a(x)\lambda^a \ , \quad a \in \{1, 2, \dots, 8\} \ . \tag{4.5.1}$$

The λ^a are the eight Gell-Mann matrices — generators of $SU(3)$ — described in Appendix E. They are Hermitean, traceless 3×3 matrices, analogous to the 3 Pauli matrices which generate $SU(2)$. Under gauge transformations $\Omega \in SU(3)$ the scalar field transforms as

$$\mathbf{\Phi}'(x) = \Omega(x)\mathbf{\Phi}(x)\Omega(x)^\dagger \ . \tag{4.5.2}$$

We introduce a potential of the form

$$V(\mathbf{\Phi}) = \frac{m^2}{4}{\rm Tr}\left(\mathbf{\Phi}^2\right) + \frac{\mu}{3!}{\rm Tr}\left(\mathbf{\Phi}^3\right) + + \frac{\lambda}{4!}{\rm Tr}\left(\mathbf{\Phi}^4\right) \ . \tag{4.5.3}$$

The potential is gauge invariant due to the cyclic nature of the trace. It is interesting to note (and straightforward to check) that

$$\left({\rm Tr}\left(\mathbf{\Phi}^2\right)\right)^2 = 2{\rm Tr}\left(\mathbf{\Phi}^4\right) \ ,$$

$$\det\mathbf{\Phi} = -\frac{1}{3}{\rm Tr}\left(\mathbf{\Phi}^3\right) \ , \tag{4.5.4}$$

which implies that the quartic potential of eq. (4.5.3 represents the most general $SU(3)$-invariant and renormalizable form. Since the term $({\rm Tr}\,(\mathbf{\Phi}^2))$ is actually $SO(8)$ rather than just $SU(3)$ invariant, eq. (4.5.4) also implies that, for $\mu = 0$, the potential $V(\mathbf{\Phi})$ has an enlarged $SO(8)$ symmetry. For $m^2 < 0$, this symmetry breaks spontaneously down to $SO(7)$, thus leading to 7 massless Nambu-Goldstone bosons. When $\mu \neq 0$, on the other hand, the symmetry of the potential is just $SU(3)$, which can break spontaneously to $SU(2) \times U(1)$ or to $U(1) \times U(1)$, leading to $8-3-1 = 4$ or $8-1-1 = 6$ massless Nambu-Goldstone bosons, respectively. To investigate the pattern of symmetry breaking, we choose a unitary gauge, in which the scalar field is diagonal (one uses the unitary transformation $\Omega(x)$ to diagonalize the Hermitean matrix $\mathbf{\Phi}(x)$)

$$\mathbf{\Phi}(x) = \Phi_3(x)\lambda^3 + \Phi_8(x)\lambda^8 = \begin{pmatrix} \Phi_3(x) + \frac{1}{\sqrt{3}}\Phi_8(x) & 0 & 0 \\ 0 & -\Phi_3(x) + \frac{1}{\sqrt{3}}\Phi_8(x) & 0 \\ 0 & 0 & -\frac{1}{\sqrt{3}}\Phi_8(x) \end{pmatrix} . \tag{4.5.5}$$

The potential then takes the form

$$V(\mathbf{\Phi}) = \frac{m^2}{2}\sum_i \Phi_i^2 + \frac{\lambda_1}{4!}\left(\sum_i \Phi_i^2\right)^2 + \frac{\lambda_2}{4!}\sum_i \Phi_i^4 \; . \tag{4.5.6}$$

The potential then takes the form

$$V(\mathbf{\Phi}) = \frac{m^2}{2}\sum_i \Phi_i^2 + \frac{\lambda_1}{4!}\left(\sum_i \Phi_i^2\right)^2 + \frac{\lambda_2}{4!}\sum_i \Phi_i^4 \; . \tag{4.5.7}$$

The minima of the potential are characterized by

$$\frac{\partial V}{\partial \Phi_i} = m^2\Phi_i + \Phi_i\frac{\lambda_1}{6}\sum_j \Phi_j^2 + \frac{\lambda_2}{6}\Phi_i^3 = c \; . \tag{4.5.8}$$

Here c is a Lagrange multiplier that implements the constraint $\sum_i \Phi_i = 0$. We are interested in minima with an unbroken $SU(2)_{\mathrm{L}} \times U(1)_Y$ symmetry, for which $\Phi_1 = \Phi_2$. Hence, we can write

$$\mathbf{\Phi}(x) = v\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \tag{4.5.9}$$

such that

$$\begin{aligned}
& m^2 v + 4\lambda_1 v^2(3 + 2\frac{9}{4})v + 4\lambda_2 v^3 = C, \\
& -\frac{3}{2}m^2 v - 4\lambda_1 v^2(3 + 2\frac{9}{4})\frac{3}{2}v - 4\lambda_2 v^3\frac{27}{8} = C \;\Rightarrow \\
& C = \frac{4}{5}\lambda_2 v^3(3 - \frac{27}{4}) = -3\lambda_2 v^3 \;\Rightarrow \\
& m^2 v + \lambda_1 30 v^3 + \lambda_2 7 v^3 = 0 \;\Rightarrow\; v = \sqrt{-\frac{m^2}{30\lambda_1 + 7\lambda_2}}.
\end{aligned} \tag{4.5.10}$$

The value of the potential at the minimum is given by

$$\begin{aligned}
V(\Phi) &= \frac{1}{2}m^2 v^2(3 + 2\frac{9}{4}) + \lambda_1 v^4(3 + 2\frac{9}{4})^2 + \lambda_2 v^4(3 + 2\frac{81}{16}) \\
&= \frac{1}{2}m^2 v^2\frac{15}{2} + \lambda_1 v^4\frac{225}{4} + \lambda_2 v^4\frac{105}{8}) \\
&= -\frac{15}{4}v^4(30\lambda_1 + 7\lambda_2) + \lambda_1 v^4\frac{225}{4} + \lambda_2 v^4\frac{105}{8}) \\
&= v^4(-\frac{225}{4}\lambda_1 - \frac{105}{8}\lambda_2) = -m^4\frac{15}{8}\frac{1}{30\lambda_1 + 7\lambda_2}.
\end{aligned} \tag{4.5.11}$$

For $\lambda_1, \lambda_2 > 0$ the value of the potential is negative, indicating that the $SU(3)$ symmetric phase at $\Phi = 0$ with $V(\Phi) = 0$ is not the true vacuum. It is instructive to convince oneself that other symmetry breaking patterns — for example to $U(1) \times U(1)$ — are not dynamically preferred over $SU(2)_{\rm L} \times U(1)_Y$ breaking.

Let us now consider the $SU(3)$ unified gauge field

$$V_\mu(x) = \mathrm{i} g_3 V_\mu^a(x) \lambda_a \ . \tag{4.5.12}$$

Under non-Abelian gauge transformations we have

$$V'_\mu(x) = \Omega(x)(V_\mu(x) + \partial_\mu)\Omega(x)^\dagger \ . \tag{4.5.13}$$

For an adjoint Higgs field the covariant derivative takes the form

$$D_\mu \mathbf{\Phi}(x) = \partial_\mu \mathbf{\Phi}(x) + [V_\mu(x), \mathbf{\Phi}(x)] \ . \tag{4.5.14}$$

It is instructive to show that this indeed transforms covariantly. Introducing the field strength tensor

$$V_{\mu\nu}(x) = \partial_\mu V_\nu(x) - \partial_\nu V_\mu(x) + [V_\mu(x), V_\nu(x)] \ , \tag{4.5.15}$$

the bosonic part of the $SU(3)$ GUT Lagrangian takes the form

$$\mathcal{L}(\mathbf{\Phi}, \partial_\mu \mathbf{\Phi}, V_\mu, \partial_\mu V_\nu) = \frac{1}{2} \mathrm{Tr}\,(D_\mu \mathbf{\Phi} D_\mu \mathbf{\Phi}) + V(\mathbf{\Phi}) + \frac{1}{4} \mathrm{Tr}\,(V_{\mu\nu} V_{\mu\nu}) \ . \tag{4.5.16}$$

We now insert the vacuum value of the scalar field to obtain the mass terms for the gauge field

$$\frac{1}{2} \mathrm{Tr}\,(D_\mu \mathbf{\Phi} D_\mu \mathbf{\Phi}) = \mathrm{Tr}\,([V^\mu, \mathbf{\Phi}][V_\mu, \mathbf{\Phi}]) \ . \tag{4.5.17}$$

We introduce the X- and Y-bosons via

$$V_\mu(x) = \begin{pmatrix} & & & X_\mu^r & Y_\mu^r \\ & G_\mu & & X_\mu^g & Y_\mu^g \\ & & & X_\mu^b & Y_\mu^b \\ X_\mu^{r*} & X_\mu^{g*} & X_\mu^{b*} & & \\ Y_\mu^{r*} & Y_\mu^{g*} & Y_\mu^{b*} & & W_\mu \end{pmatrix} . \tag{4.5.18}$$

The X- and Y-boson form an electroweak doublet. They are the fields that become massive after the spontaneous breakdown of $SU(3)$ down to $SU(2)_\mathrm{L} \times U(1)_Y$, because one obtains

$$
\begin{aligned}
[V_\mu, \mathbf{\Phi}] &= v \begin{pmatrix} & G_\mu \mathbf{1} & & \begin{matrix} -\frac{3}{2}X_\mu^r & -\frac{3}{2}Y_\mu^r \\ -\frac{3}{2}X_\mu^g & -\frac{3}{2}Y_\mu^g \\ -\frac{3}{2}X_\mu^b & -\frac{3}{2}Y_\mu^b \end{matrix} \\ \begin{matrix} X_\mu^{r*} \\ Y_\mu^{r*} \end{matrix} & \begin{matrix} X_\mu^{g*} \\ Y_\mu^{g*} \end{matrix} & \begin{matrix} X_\mu^{b*} \\ Y_\mu^{b*} \end{matrix} & -\frac{3}{2}W_\mu \mathbf{1} \end{pmatrix} \\
&- v \begin{pmatrix} & G_\mu \mathbf{1} & & \begin{matrix} X_\mu^r & Y_\mu^r \\ X_\mu^g & Y_\mu^g \\ X_\mu^b & Y_\mu^b \end{matrix} \\ \begin{matrix} -\frac{3}{2}X_\mu^{r*} \\ -\frac{3}{2}Y_\mu^{r*} \end{matrix} & \begin{matrix} -\frac{3}{2}X_\mu^{g*} \\ -\frac{3}{2}Y_\mu^{g*} \end{matrix} & \begin{matrix} -\frac{3}{2}X_\mu^{b*} \\ -\frac{3}{2}Y_\mu^{b*} \end{matrix} & -\frac{3}{2}W_\mu \mathbf{1} \end{pmatrix} \\
&= v \begin{pmatrix} & 0 & & \begin{matrix} -\frac{5}{2}X_\mu^r & -\frac{5}{2}Y_\mu^r \\ -\frac{5}{2}X_\mu^g & -\frac{5}{2}Y_\mu^g \\ -\frac{5}{2}X_\mu^b & -\frac{5}{2}Y_\mu^b \end{matrix} \\ \begin{matrix} -\frac{5}{2}X_\mu^{r*} \\ -\frac{5}{2}Y_\mu^{r*} \end{matrix} & \begin{matrix} -\frac{5}{2}X_\mu^{g*} \\ -\frac{5}{2}Y_\mu^{g*} \end{matrix} & \begin{matrix} -\frac{5}{2}X_\mu^{b*} \\ -\frac{5}{2}Y_\mu^{b*} \end{matrix} & 0 \end{pmatrix} ,
\end{aligned}
\tag{4.5.19}
$$

and hence

$$\mathrm{Tr}\left([V_\mu, \mathbf{\Phi}][V_\mu, \mathbf{\Phi}]\right) = -\frac{9}{2} v^2 (X_\mu^* X_\mu + Y_\mu^* Y_\mu) \ . \tag{4.5.20}$$

The X- and Y-bosons thus pick up the mass

$$m_X^2 = m_Y^2 = \frac{9}{2} g_3^2 v^2 \ . \tag{4.5.21}$$

These 4 gauge bosons become massive by eating 4 Nambu-Goldstone bosons. Indeed, when the grand unified group $G = SU(3)$ breaks spontaneously down to the subgroup $H = SU(2)_\mathrm{L} \times U(1)_Y$, according to the Goldstone theorem, there are $8 - 3 - 1 = 4$ Nambu-Goldstone bosons.

The full Standard Model gauge group is $SU(3)_\mathrm{c} \times SU(2)_\mathrm{L} \times U(1)_Y$. The group $SU(n)$ has rank $n-1$, *i.e.* $n-1$ of the n^2-1 generators commute with each other. The rank of the group $U(1)$ is 1. Thus, the rank of the Standard Model group is $2 + 1 + 1 = 4$. Hence, if we want to embed that group in a simple Lie group, its rank must be at least 4. The smallest Lie group (i.e. the one with the smallest number of generators) with that property is $SU(5)$, which has rank 4 and $5^2 - 1 = 24$ generators. Consequently, in an $SU(5)$ gauge theory there are 24 gauge bosons. When the Standard Model is embedded in $SU(5)$, half of the gauge bosons can be identified with known particles: $SU(3)_c$ has $3^2 - 1 = 8$ gluons, $SU(2)_L$ has $2^2 - 1 = 3$ W-bosons, and $U(1)_Y$ has one B-boson. The remaining 12 gauge bosons of $SU(5)$ are hypothetical particles, again called X and Y. In order to make these unobserved particles sufficiently heavy, the $SU(5)$ symmetry must be spontaneously broken down to $SU(3)_c \times SU(2)_L \times U(1)_Y$. Again, this can be achieved using the Higgs mechanism, now with a scalar field transforming under the 24-dimensional adjoint representation of $SU(5)$. The X- and Y-bosons are color triplets and electroweak doublets. These 12 gauge bosons become massive by eating 12 Nambu-Goldstone bosons. Indeed when $G = SU(5)$ breaks spontaneously down to the subgroup $H = SU(3)_c \times SU(2)_L \times U(1)_Y$, according to the Goldstone theorem there are $n_G - n_H = 24 - 8 - 3 - 1 = 12$ Nambu-Goldstone bosons.

In the $SU(5)$ GUT there is only one gauge coupling g_5 to which the three standard model gauge couplings g, g', and g_s are related. In an $SU(5)$ symmetric phase one has

$$g = g_s = g_5, \quad g' = \sqrt{\frac{3}{5}} g_5 \ . \tag{4.5.22}$$

Hence, the Weinberg angle would then take the form

$$\sin^2 \theta_W = \frac{g'^2}{g^2 + g'^2} = \frac{3g_5^2}{5g_5^2 + 3g_5^2} = \frac{3}{8} \ . \tag{4.5.23}$$

This is not in agreement with the experimental value $\sin^2 \theta_W = 0.23119(14)$. However, we do not live in an $SU(5)$ symmetric world. One can use the renormalization group to run the above relations from the GUT scale, where they apply, down to our low energy scales. One obtains realistic values for the coupling constants when one puts the GUT scale at about $v = 10^{15}$ GeV. The masses of the X- and Y-bosons are also in that range. The GUT scale is significantly below the Planck scale 10^{19} GeV, which justifies neglecting gravity in the above considerations. In order to achieve simultaneous unification of all three couplings g, g', and g_s at the GUT scale, one must add further matter degrees of freedom beyond the quarks and leptons of the

Standard Model. For example, the minimal supersymmetric extension of the Standard Model achieves this property.

As we will discuss in Chapter 15, Grand Unified Theories predict the decay of the proton at least at some rate. Despite numerous experimental efforts, proton decay has never been observed, *i.e.* as far as we know today, the proton is a stable particle. To be explicit, its life-time exceeds 2.1×10^{29} years (with 90 percent confidence level). Indeed, the minimal $SU(5)$ model has been ruled out experimentally, because the proton lives longer than this model predicts. Other GUTs based on the orthogonal group $SO(10)$ or the exceptional group $E(6)$ predict proton decay at a slower rates, which are not ruled out experimentally.

Chapter 5

One Generation of Leptons and Quarks

In this chapter we add the *fermions* to the Lagrangian of the Standard Model. The fermions are *leptons* and *quarks*. The leptons participate in the electroweak gauge interactions, whereas the quarks are affected by both, electroweak and strong interactions. It is interesting that we need to add leptons *and* quarks at the same time; a simplification of the Standard Model without quarks would be mathematically inconsistent. This is because the quarks cancel anomalies, which would explicitly break the gauge symmetry in a purely leptonic model at the quantum level. Cancellation of *anomalies in gauge symmetries* is absolutely necessary, both perturbatively and beyond perturbation theory. *Anomalies in global symmetries*, on the other hand, are a perfectly acceptable form of explicit symmetry breaking. In fact, they are necessary to correctly describe some aspects of the physics. In this chapter, we will limit ourselves to *one single generation* of fermions. Until recently, the corresponding lepton fields would have included only left-handed electrons and neutrinos as well as right-handed electrons, but no right-handed neutrinos. By now we know that neutrinos have a small mass, which motivates the addition of a right-handed neutrino field. Still, we will follow our strategy of adding fields step by step, and so we will first work with left-handed neutrinos only. In this chapter, we will also limit ourselves to *one single generation* of fermions.

5.1 Weyl and Dirac Spinors

The 4-dimensional Euclidean space-time is invariant against translations by 4-vectors as well as against $SO(4)$ space-time rotations. Together this constitutes Euclidean Poincaré invariance. Ihe internal $O(4)$ symmetry of the Higgs sector contains an $SO(4)$ subgroup which factorizes into the two internal symmetries $SU(2)_\mathrm{L}$ and $SU(2)_\mathrm{R}$, as we have seen in Section 5.2. Since the group theory is identical, the same is true for the space-time rotation symmetry $SO(4) = SU(2)_\mathrm{L} \times SU(2)_\mathrm{R}$. The fields of the Standard Model must transform appropriately under space-time rotations. Their transformation behavior can be characterized by specifying the representation of $SO(4)$, or equivalently of $SU(2)_\mathrm{L}$ and $SU(2)_\mathrm{R}$. Since $SU(2)$ representations are characterized by a "spin" $S = 0, \frac{1}{2}, 1, \ldots$, the transformation behavior of the Standard Model fields under $SO(4)$ space-time rotations can be characterized by a pair $(S_\mathrm{L}, S_\mathrm{R})$. Scalar fields are invariant under space-time rotations and thus transform in the $(0,0)$ representation of $SO(4)$. Vector fields, on the other hand, are 4-vectors and transform as $(\frac{1}{2}, \frac{1}{2})$.

We will soon introduce the fermion fields of the Standard Model. The fundamental fermion fields of the Standard Model are left- or right-handed Weyl fermions, which transform as $(\frac{1}{2}, 0)$ or $(0, \frac{1}{2})$, respectively. A Dirac fermion, on the other hand, is described by two Weyl fermion fields, one left- and one right-handed, and thus transforms in the reducible representation $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$. In Euclidean space-time, the Dirac matrices γ_μ are Hermitean and obey the anti-commutation relation

$$\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu} \ , \ \gamma_\mu^\dagger = \gamma_\mu \ . \tag{5.1.1}$$

In addition, we define

$$\gamma_5 = \gamma_1 \gamma_2 \gamma_3 \gamma_4 \ , \tag{5.1.2}$$

which implies

$$\{\gamma_\mu, \gamma_5\} = 0 \ , \ \gamma_5^\dagger = \gamma_5 \ , \ \gamma_5^2 = 1 \ . \tag{5.1.3}$$

In the *chiral basis* (also known as the Weyl basis), in which γ_5 is diagonal, the Dirac matrices take the form

$$\gamma_i = \sigma_2 \otimes \sigma_i = \begin{pmatrix} 0 & -\mathrm{i}\sigma_i \\ \mathrm{i}\sigma_i & 0 \end{pmatrix} \ , \ \gamma_4 = \sigma_1 \otimes \mathbf{1} = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \ , \ \gamma_5 = \sigma_3 \otimes \mathbf{1} = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix} \ , \tag{5.1.4}$$

where σ_i with $i \in \{1, 2, 3\}$ are the Pauli matrices and $\mathbf{1}$ is the 2×2 unit-matrix. It is convenient to introduce projection operators on the left- and right-handed components of a Dirac spinor

$$P_{\mathrm{R}} = \frac{1}{2}(1 + \gamma_5) = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & 0 \end{pmatrix} , \quad P_{\mathrm{L}} = \frac{1}{2}(1 - \gamma_5) = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{1} \end{pmatrix} , \tag{5.1.5}$$

which obey

$$P_{\mathrm{R}}^2 = P_{\mathrm{R}} , \quad P_{\mathrm{L}}^2 = P_{\mathrm{L}} , \quad P_{\mathrm{R}} + P_{\mathrm{L}} = 1 , \quad P_{\mathrm{R}} P_{\mathrm{L}} = P_{\mathrm{L}} P_{\mathrm{R}} = 0 . \tag{5.1.6}$$

In the Euclidean functional integral, fermions are described by anti-commuting Grassmann variables, which are discussed in Appendix E. A right-handed (or left-handed) *Weyl spinor* $\psi_{\mathrm{R}}(x)$ (or $\psi_{\mathrm{L}}(x)$) consists of two Grassmann numbers $\psi_{\mathrm{R}}^1(x)$ and $\psi_{\mathrm{R}}^2(x)$ (or $\psi_{\mathrm{L}}^1(x)$ and $\psi_{\mathrm{L}}^2(x)$). Two Weyl spinors can be combined to form a 4-component *Dirac spinor*

$$\psi(x) = \begin{pmatrix} \psi_{\mathrm{R}}^1(x) \\ \psi_{\mathrm{R}}^2(x) \\ \psi_{\mathrm{L}}^1(x) \\ \psi_{\mathrm{L}}^2(x) \end{pmatrix} . \tag{5.1.7}$$

By applying the projection operators, we recover the Weyl spinors

$$\psi_{\mathrm{R}}(x) = P_{\mathrm{R}} \psi(x) = \begin{pmatrix} \psi_{\mathrm{R}}^1(x) \\ \psi_{\mathrm{R}}^2(x) \\ 0 \\ 0 \end{pmatrix} , \quad \psi_{\mathrm{L}}(x) = P_{\mathrm{L}} \psi(x) = \begin{pmatrix} 0 \\ 0 \\ \psi_{\mathrm{L}}^1(x) \\ \psi_{\mathrm{L}}^2(x) \end{pmatrix} . \tag{5.1.8}$$

In order to account for fermions and anti-fermions, we also introduce the spinors $\bar{\psi}_{\mathrm{L}}(x)$ and $\bar{\psi}_{\mathrm{R}}(x)$, which consist of additional independent Grassmann numbers $\bar{\psi}_{\mathrm{L}}^1(x)$, $\bar{\psi}_{\mathrm{L}}^2(x)$ and $\bar{\psi}_{\mathrm{R}}^1(x)$, $\bar{\psi}_{\mathrm{R}}^2(x)$.[1] Again these can be combined to form the Dirac spinor

$$\bar{\psi}(x) = \left(\bar{\psi}_{\mathrm{L}}^1(x), \bar{\psi}_{\mathrm{L}}^2(x), \bar{\psi}_{\mathrm{R}}^1(x), \bar{\psi}_{\mathrm{R}}^2(x) \right) . \tag{5.1.9}$$

[1] In the Hamiltonian formalism, the field operators $\hat{\psi}(x)$ and $\hat{\bar{\psi}}(x) = \hat{\psi}^\dagger(x)\gamma_0$ are related, while the corresponding Grassmann-valued fields $\psi(x)$ and $\bar{\psi}(x)$ in the functional integral are independent variables.

By applying the chiral projection operators we recover the Weyl spinors

$$\bar{\psi}_{\mathrm{R}}(x) = \bar{\psi}(x)P_{\mathrm{L}} = \big(0, 0, \bar{\psi}^1_{\mathrm{R}}(x), \bar{\psi}^2_{\mathrm{R}}(x)\big) \ ,$$
$$\bar{\psi}_{\mathrm{L}}(x) = \bar{\psi}(x)P_{\mathrm{R}} = \big(\bar{\psi}^1_{\mathrm{L}}(x), \bar{\psi}^2_{\mathrm{L}}(x), 0, 0\big) \ . \tag{5.1.10}$$

One can now construct separate Lorentz-invariant Lagrangians for free massless left- or right-handed Weyl fermions

$$\mathcal{L}_{0\mathrm{R}}(\bar{\psi}_{\mathrm{R}}, \psi_{\mathrm{R}}) = \bar{\psi}_{\mathrm{R}}\gamma_\mu\partial_\mu\psi_{\mathrm{R}} \ , \quad \mathcal{L}_{0\mathrm{L}}(\bar{\psi}_{\mathrm{L}}, \psi_{\mathrm{L}}) = \bar{\psi}_{\mathrm{L}}\gamma_\mu\partial_\mu\psi_{\mathrm{L}} \ . \tag{5.1.11}$$

A massive free Dirac fermion, on the other hand, requires both left- and right-handed components and is described by the Lagrangian

$$\mathcal{L}_0(\bar{\psi}, \psi) = \bar{\psi}\gamma_\mu\partial_\mu\psi + m\bar{\psi}\psi = \bar{\psi}_{\mathrm{R}}\gamma_\mu\partial_\mu\psi_{\mathrm{R}} + \bar{\psi}_{\mathrm{L}}\gamma_\mu\partial_\mu\psi_{\mathrm{L}} + m(\bar{\psi}_{\mathrm{R}}\psi_{\mathrm{L}} + \bar{\psi}_{\mathrm{L}}\psi_{\mathrm{R}}) \ . \tag{5.1.12}$$

In particular, the mass term couples left- and right-handed fields. In the free theory these fields decouple only in the *chiral limit*, $m = 0$.

5.2 Parity, Charge Conjugation, and Time-Reversal

Parity and charge conjugation are important discrete symmetries that exchange left- and right-handed Weyl fermions. In Euclidean space-time, *parity* acts as a spatial inversion, which replaces $x = (\vec{x}, x_4)$ with $(-\vec{x}, x_4)$, combined with multiplication by a matrix P in Dirac space, *i.e.*

$${}^{\mathrm{P}}\psi(\vec{x}, x_4) = P\psi(-\vec{x}, x_4) \ , \quad {}^{\mathrm{P}}\bar{\psi}(\vec{x}, x_4) = \bar{\psi}(-\vec{x}, x_4)P^{-1} \ . \tag{5.2.1}$$

The matrix P obeys

$$P^{-1}\gamma_i P = -\gamma_i \ , \quad P^{-1}\gamma_4 P = \gamma_4 \ , \tag{5.2.2}$$

and in the chiral basis it takes the form

$$P = P^{-1} = \gamma_4 = \sigma_1 \otimes \mathbf{1} = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \ . \tag{5.2.3}$$

As a consequence, parity exchanges left- and right-handed fields, *i.e.*

$$\begin{aligned}
{}^{\mathrm{P}}\psi_{\mathrm{R}}(\vec{x},x_4) &= P_{\mathrm{R}}\gamma_4\psi(-\vec{x},x_4) = \gamma_4 P_{\mathrm{L}}\psi(-\vec{x},x_4) = P\psi_{\mathrm{L}}(-\vec{x},x_4) \ , \\
{}^{\mathrm{P}}\psi_{\mathrm{L}}(\vec{x},x_4) &= P_{\mathrm{L}}\gamma_4\psi(-\vec{x},x_4) = \gamma_4 P_{\mathrm{R}}\psi(-\vec{x},x_4) = P\psi_{\mathrm{R}}(-\vec{x},x_4) \ , \\
{}^{\mathrm{P}}\bar{\psi}_{\mathrm{R}}(\vec{x},x_4) &= \bar{\psi}(-\vec{x},x_4)\gamma_4 P_{\mathrm{L}} = \bar{\psi}(-\vec{x},x_4)P_{\mathrm{R}}\gamma_4 = \bar{\psi}_{\mathrm{L}}(-\vec{x},x_4)P^{-1} \ , \\
{}^{\mathrm{P}}\bar{\psi}_{\mathrm{L}}(\vec{x},x_4) &= \bar{\psi}(-\vec{x},x_4)\gamma_4 P_{\mathrm{R}} = \bar{\psi}(-\vec{x},x_4)P_{\mathrm{L}}\gamma_4 = \bar{\psi}_{\mathrm{R}}(-\vec{x},x_4)P^{-1} \ ,
\end{aligned} \tag{5.2.4}$$

which implies that a theory with fermions of just one chirality explicitly violates parity.

The Lagrangian depends on fields which are functions of x. Since under parity $x = (\vec{x}, x_4)$ turns into $(-\vec{x}, x_4)$, the Lagrangian itself can not be P-invariant. What may be invariant, however, is the action. Let us hence apply parity to the action of a free right-handed fermion

$$\begin{aligned}
S_{0\mathrm{R}}[{}^{\mathrm{P}}\bar{\psi}_{\mathrm{R}}, {}^{\mathrm{P}}\psi_{\mathrm{R}}] &= \int d^4x \ \mathcal{L}_{0\mathrm{R}}({}^{\mathrm{P}}\bar{\psi}_{\mathrm{R}}, {}^{\mathrm{P}}\psi_{\mathrm{R}}) = \int d^4x \ {}^{\mathrm{P}}\bar{\psi}_{\mathrm{R}}(\vec{x},x_4)\gamma_\mu\partial_\mu \ {}^{\mathrm{P}}\psi_{\mathrm{L}}(\vec{x},x_4) \\
&= \int d^4x \ \bar{\psi}_{\mathrm{L}}(-\vec{x},x_4)P^{-1}\gamma_\mu\partial_\mu P\psi_{\mathrm{L}}(-\vec{x},x_4) \\
&= \int d^4x \ \bar{\psi}_{\mathrm{L}}(-\vec{x},x_4)(-\gamma_i\partial_i + \gamma_4\partial_4)\psi_{\mathrm{L}}(-\vec{x},x_4) \\
&= \int d^4x \ \bar{\psi}_{\mathrm{L}}(\vec{x},x_4)\gamma_\mu\partial_\mu\psi_{\mathrm{L}}(\vec{x},x_4) = S_{0\mathrm{L}}[\bar{\psi}_{\mathrm{L}}, \psi_{\mathrm{L}}] \ .
\end{aligned} \tag{5.2.5}$$

In the last step we have made a change of variables from $-\vec{x}$ to $\vec{x}$. As we see, under parity the action of a right-handed fermion turns into the one of a left-handed fermion. In particular, each individual action is not invariant against P.

Let us now consider *charge conjugation,* which exchanges particles and anti-particles. In the Euclidean functional integral, charge conjugation acts as

$${}^{\mathrm{C}}\psi(x) = C\bar{\psi}(x)^{\mathsf{T}} \ , \quad {}^{\mathrm{C}}\bar{\psi}(x) = -\psi(x)^{\mathsf{T}}C^{-1} \ , \tag{5.2.6}$$

where T denotes transpose and the charge conjugation matrix in Dirac space satisfies

$$C^{-1}\gamma_\mu C = -\gamma_\mu^{\mathsf{T}} \ . \tag{5.2.7}$$

In the chiral basis it is given by

$$C = C^{-1} = \mathrm{i}\gamma_2\gamma_4 = \sigma_3 \otimes \sigma_2 = \begin{pmatrix} \sigma_2 & 0 \\ 0 & -\sigma_2 \end{pmatrix} \ . \tag{5.2.8}$$

This implies that also charge conjugation exchanges left- and right-handed fermions

$$\begin{aligned}
{}^{\mathrm{C}}\psi_{\mathrm{R}}(x) &= P_{\mathrm{R}} C \bar{\psi}(x)^{\mathsf{T}} = C P_{\mathrm{R}} \bar{\psi}(x)^{\mathsf{T}} = C[\bar{\psi}(x) P_{\mathrm{R}}]^{\mathsf{T}} = C \bar{\psi}_{\mathrm{L}}(x)^{\mathsf{T}} \ , \\
{}^{\mathrm{C}}\psi_{\mathrm{L}}(x) &= P_{\mathrm{L}} C \bar{\psi}(x)^{\mathsf{T}} = C P_{\mathrm{L}} \bar{\psi}(x)^{\mathsf{T}} = C[\bar{\psi}(x) P_{\mathrm{L}}]^{\mathsf{T}} = C \bar{\psi}_{\mathrm{R}}(x)^{\mathsf{T}} \ , \\
{}^{\mathrm{C}}\bar{\psi}_{\mathrm{R}}(x) &= -\psi(x)^{\mathsf{T}} C^{-1} P_{\mathrm{L}} = -\psi(x)^{\mathsf{T}} P_{\mathrm{L}} C^{-1} = -[P_{\mathrm{L}} \psi(x)]^{\mathsf{T}} C^{-1} = -\psi_{\mathrm{L}}(x)^{\mathsf{T}} C^{-1} \ , \\
{}^{\mathrm{C}}\bar{\psi}_{\mathrm{L}}(x) &= -\psi(x)^{\mathsf{T}} C^{-1} P_{\mathrm{R}} = -\psi(x)^{\mathsf{T}} P_{\mathrm{R}} C^{-1} = -[P_{\mathrm{R}} \psi(x)]^{\mathsf{T}} C^{-1} = -\psi_{\mathrm{R}}(x)^{\mathsf{T}} C^{-1} \ .
\end{aligned} \tag{5.2.9}$$

Hence, a theory that contains only right- or only left-handed fermions also explicitly breaks charge conjugation.

We now apply charge conjugation to the action of a right-handed free fermion

$$\begin{aligned}
S_{0\mathrm{R}}[{}^{\mathrm{C}}\bar{\psi}_{\mathrm{R}}, {}^{\mathrm{C}}\psi_{\mathrm{R}}] &= \int d^4x \ {}^{\mathrm{C}}\bar{\psi}_{\mathrm{R}}(x) \gamma_\mu \partial_\mu \ {}^{\mathrm{C}}\psi_{\mathrm{L}}(x) \\
&= -\int d^4x \ \psi_{\mathrm{L}}(x)^{\mathsf{T}} C^{-1} \gamma_\mu \partial_\mu C \bar{\psi}_{\mathrm{L}}(x)^{\mathsf{T}} \\
&= \int d^4x \ \psi_{\mathrm{L}}(x)^{\mathsf{T}} \gamma_\mu^{\mathsf{T}} \partial_\mu \bar{\psi}_{\mathrm{L}}(x)^{\mathsf{T}} = -\int d^4x \ [\partial_\mu \bar{\psi}_{\mathrm{L}}(x) \gamma_\mu \psi_{\mathrm{L}}(x)]^{\mathrm{T}} \\
&= \int d^4x \ \bar{\psi}_{\mathrm{L}}(x) \gamma_\mu \partial_\mu \psi_{\mathrm{L}}(x) = S_{0\mathrm{L}}[\bar{\psi}_{\mathrm{L}}, \psi_{\mathrm{L}}] \ .
\end{aligned} \tag{5.2.10}$$

In the last two steps we have used the anti-commutation rules of Grassmann variables and we have performed a partial integration. Also charge conjugation exchanges the actions of left- and right-handed fermions.

Let us also consider the combination of charge conjugation and parity CP. We then have

$$\begin{aligned}
{}^{\mathrm{CP}}\psi(\vec{x}, x_4) &= C[\bar{\psi}(-\vec{x}, x_4) P^{-1}]^{\mathsf{T}} = C P \bar{\psi}(-\vec{x}, x_4)^{\mathsf{T}} \ , \\
{}^{\mathrm{CP}}\bar{\psi}(\vec{x}, x_4) &= -[P \psi(-\vec{x}, x_4)]^{\mathsf{T}} C^{-1} = -\psi(-\vec{x}, x_4)^{\mathsf{T}} P^{\mathsf{T}} C^{-1} \ ,
\end{aligned} \tag{5.2.11}$$

which implies

$$\begin{aligned}
&{}^{\mathrm{CP}}\psi_{\mathrm{R}}(\vec{x}, x_4) = C P \bar{\psi}_{\mathrm{R}}(-\vec{x}, x_4)^{\mathsf{T}} \ , \quad {}^{\mathrm{CP}}\psi_{\mathrm{L}}(\vec{x}, x_4) = C P \bar{\psi}_{\mathrm{L}}(-\vec{x}, x_4)^{\mathsf{T}} \ , \\
&{}^{\mathrm{CP}}\bar{\psi}_{\mathrm{R}}(\vec{x}, x_4) = -\psi_{\mathrm{R}}(-\vec{x}, x_4)^{\mathsf{T}} P^{\mathsf{T}} C^{-1} \ , \quad {}^{\mathrm{CP}}\bar{\psi}_{\mathrm{L}}(\vec{x}, x_4) = -\psi_{\mathrm{L}}(-\vec{x}, x_4)^{\mathsf{T}} P^{\mathsf{T}} C^{-1} .
\end{aligned} \tag{5.2.12}$$

Since both C and P exchange the actions of left- and right-handed fermions, CP leaves these actions invariant, *i.e.*

$$S_{0\mathrm{R}}[{}^{\mathrm{CP}}\bar{\psi}_{\mathrm{R}}, {}^{\mathrm{CP}}\psi_{\mathrm{R}}] = S_{0\mathrm{R}}[\bar{\psi}_{\mathrm{R}}, \psi_{\mathrm{R}}] \ , \quad S_{0\mathrm{L}}[{}^{\mathrm{CP}}\bar{\psi}_{\mathrm{L}}, {}^{\mathrm{CP}}\psi_{\mathrm{L}}] = S_{0\mathrm{L}}[\bar{\psi}_{\mathrm{L}}, \psi_{\mathrm{L}}] \ . \tag{5.2.13}$$

Finally, let us consider *Euclidean time-reversal* which acts as

$${}^{\mathrm{T}}\psi(\vec{x}, x_4) = T\bar{\psi}(\vec{x}, -x_4)^{\mathsf{T}} \ , \quad {}^{\mathrm{T}}\bar{\psi}(\vec{x}, x_4) = -\psi(\vec{x}, -x_4)^{\mathsf{T}} T^{-1} \ . \tag{5.2.14}$$

Here the superscript T on the left refers to time-reversal and the superscript T on the right denotes transpose, while the prefactor T is a matrix in Dirac space that obeys

$$T^{-1}\gamma_i T = -\gamma_i^{\mathsf{T}} \ , \quad T^{-1}\gamma_4 T = \gamma_4^{\mathsf{T}} \ . \tag{5.2.15}$$

In the chiral basis, it takes the form

$$T = \gamma_2\gamma_5 = \mathrm{i}\sigma_1 \otimes \sigma_2 = \begin{pmatrix} 0 & \mathrm{i}\sigma_2 \\ \mathrm{i}\sigma_2 & 0 \end{pmatrix} \ . \tag{5.2.16}$$

This implies

$$\begin{aligned} &{}^{\mathrm{T}}\psi_{\mathrm{R}}(\vec{x}, x_4) = T\bar{\psi}_{\mathrm{R}}(\vec{x}, -x_4)^{\mathsf{T}} \ , \quad {}^{\mathrm{T}}\psi_{\mathrm{L}}(\vec{x}, x_4) = T\bar{\psi}_{\mathrm{L}}(\vec{x}, -x_4)^{\mathsf{T}} \ , \\ &{}^{\mathrm{T}}\bar{\psi}_{\mathrm{R}}(\vec{x}, x_4) = -\psi_{\mathrm{R}}(\vec{x}, -x_4)^{\mathsf{T}} T^{-1} \ , \quad {}^{\mathrm{T}}\bar{\psi}_{\mathrm{L}}(\vec{x}, x_4) = -\psi_{\mathrm{L}}(\vec{x}, -x_4)^{\mathsf{T}} T^{-1} \ . \end{aligned} \tag{5.2.17}$$

Under T, the action of a free right-handed fermion then transforms as

$$\begin{aligned} S_{0\mathrm{R}}[{}^{\mathrm{T}}\bar{\psi}_{\mathrm{R}}, {}^{\mathrm{T}}\psi_{\mathrm{R}}] &= \int d^4x \ {}^{\mathrm{T}}\bar{\psi}_{\mathrm{R}}(\vec{x}, x_4)\gamma_\mu\partial_\mu \ {}^{\mathrm{T}}\psi_{\mathrm{R}}(\vec{x}, x_4) \\ &= -\int d^4x \ \psi_{\mathrm{R}}(\vec{x}, -x_4)^{\mathsf{T}} T^{-1}\gamma_\mu\partial_\mu T\bar{\psi}_{\mathrm{R}}(\vec{x}, -x_4)^{\mathsf{T}} \\ &= \int d^4x \ \psi_{\mathrm{R}}(\vec{x}, -x_4)^{\mathsf{T}}(\gamma_i\partial_i - \gamma_4\partial_4)^{\mathsf{T}}\bar{\psi}_{\mathrm{R}}(\vec{x}, -x_4)^{\mathsf{T}} \\ &= -\int d^4x \ [(\gamma_i\partial_i - \gamma_4\partial_4)\bar{\psi}_{\mathrm{R}}(\vec{x}, -x_4)\psi_{\mathrm{R}}(\vec{x}, -x_4)]^{\mathsf{T}} \\ &= -\int d^4x \ (\gamma_i\partial_i + \gamma_4\partial_4)\bar{\psi}_{\mathrm{R}}(\vec{x}, x_4)\psi_{\mathrm{R}}(\vec{x}, x_4) \\ &= \int d^4x \ \bar{\psi}_{\mathrm{R}}(\vec{x}, x_4)\gamma_\mu\partial_\mu\psi_{\mathrm{R}}(\vec{x}, x_4) = S_{0\mathrm{R}}[\bar{\psi}_{\mathrm{R}}, \psi_{\mathrm{R}}] \end{aligned} \tag{5.2.18}$$

In the last three steps we have used the anti-commutation rules of Grassmann variables, we have substituted $-x_4$ by x_4, and we have performed a partial integration. Similarly, for left-handed fermions one obtains $S_{0L}[{}^{\mathrm{T}}\bar{\psi}_{\mathrm{L}}, {}^{\mathrm{T}}\psi_{\mathrm{L}}] = S_{0L}[\bar{\psi}_{\mathrm{L}}, \psi_{\mathrm{L}}]$.

As was first shown by Wolfgang Pauli, the combination CPT is a symmetry of any relativistic quantum field theory. This is the *CPT theorem* [?].[2] On a fermion field, the CPT symmetry acts as

$$^{\mathrm{CPT}}\psi(x) = -\mathrm{i}\gamma_5\psi(-x) \ , \quad ^{\mathrm{CPT}}\bar{\psi}(x) = \mathrm{i}\bar{\psi}(-x)\gamma_5 \ , \tag{5.2.19}$$

which implies

$$\begin{aligned} ^{\mathrm{CPT}}\psi_{\mathrm{R}}(x) &= -\mathrm{i}\psi_{\mathrm{R}}(-x) \ , \quad ^{\mathrm{CPT}}\psi_{\mathrm{L}}(x) = \mathrm{i}\psi_{\mathrm{L}}(-x) \ , \\ ^{\mathrm{CPT}}\bar{\psi}_{\mathrm{R}}(x) &= -\mathrm{i}\bar{\psi}_{\mathrm{R}}(-x) \ , \quad ^{\mathrm{CPT}}\bar{\psi}_{\mathrm{L}}(x) = \mathrm{i}\bar{\psi}_{\mathrm{L}}(-x) \ . \end{aligned} \tag{5.2.20}$$

It is interesting to note that, as one would expect, parity, charge conjugation, and time-reversal square to the identity, i.e.

$$\mathrm{P}^2 = \mathrm{C}^2 = \mathrm{T}^2 = 1 \ , \tag{5.2.21}$$

while they do not all commute with one another. In particular, in the chiral basis one obtains

$$\mathrm{P\ C} = -\mathrm{C\ P} \ , \quad \mathrm{C\ T} = -\mathrm{T\ C} \ , \quad \mathrm{T\ P} = \mathrm{P\ T} \ . \tag{5.2.22}$$

5.3 Electrons and Left-handed Neutrinos

The leptons of the *first generation* are electrons and their neutrinos. We start with left-handed neutrinos and right-handed anti-neutrinos only. We denote the spinor fields of these leptons as $\nu_{\mathrm{L}}(x), \bar{\nu}_{\mathrm{L}}(x), e_{\mathrm{L}}(x), e_{\mathrm{R}}(x), \bar{e}_{\mathrm{L}}(x)$, and $\bar{e}_{\mathrm{R}}(x)$. Before we introduce right-handed neutrino fields, the neutrinos are massless, while the electrons will pick up a mass through the Higgs mechanism. However, before we introduce couplings between the lepton fields and the Higgs field, even the electrons are massless.

[2] While the CPT theorem applies to all relativistic local quantum field theories, it does not always apply beyond this framework, *e.g.* in string theory which violates strict locality.

At this point — without mass or interaction terms — the free lepton Lagrangian

$$\mathcal{L}_0(\bar{\nu},\nu,\bar{e},e) = \bar{\nu}_{\mathrm{L}}\gamma_\mu\partial_\mu\nu_{\mathrm{L}} + \bar{e}_{\mathrm{L}}\gamma_\mu\partial_\mu e_{\mathrm{L}} + \bar{e}_{\mathrm{R}}\gamma_\mu\partial_\mu e_{\mathrm{R}} \tag{5.3.1}$$

has several global symmetries. First of all, all lepton fields can be multiplied by the same phase $\chi \in \mathbb{R}$

$$\begin{aligned}
\nu'_{\mathrm{L}}(x) &= \exp(\mathrm{i}\chi)\nu_{\mathrm{L}}(x)\ , \quad \bar{\nu}'_{\mathrm{L}}(x) = \bar{\nu}_{\mathrm{L}}(x)\exp(-\mathrm{i}\chi)\ , \\
e'_{\mathrm{L}}(x) &= \exp(\mathrm{i}\chi)e_{\mathrm{L}}(x)\ , \quad \bar{e}'_{\mathrm{L}}(x) = \bar{e}_{\mathrm{L}}(x)\exp(-\mathrm{i}\chi)\ , \\
e'_{\mathrm{R}}(x) &= \exp(\mathrm{i}\chi)e_{\mathrm{R}}(x)\ , \quad \bar{e}'_{\mathrm{R}}(x) = \bar{e}_{\mathrm{R}}(x)\exp(-\mathrm{i}\chi)\ .
\end{aligned} \tag{5.3.2}$$

The corresponding global symmetry $U(1)_L$ is associated with lepton number conservation.[3] This symmetry is *vector-like* because it affects left- and right-handed lepton fields in the same way.

The free lepton Lagrangian also has another global Abelian symmetry, which is promoted to the local $U(1)_Y$ symmetry in the Standard Model

$$\begin{aligned}
\nu'_{\mathrm{L}}(x) &= \exp\left(\mathrm{i}Y_{l_{\mathrm{L}}}g'\varphi(x)\right)\nu_{\mathrm{L}}(x)\ , \quad \bar{\nu}'_{\mathrm{L}}(x) = \bar{\nu}_{\mathrm{L}}(x)\exp\left(-\mathrm{i}Y_{l_{\mathrm{L}}}g'\varphi(x)\right)\ , \\
e'_{\mathrm{L}}(x) &= \exp\left(\mathrm{i}Y_{l_{\mathrm{L}}}g'\varphi(x)\right)e_{\mathrm{L}}(x)\ , \quad \bar{e}'_{\mathrm{L}}(x) = \bar{e}_{\mathrm{L}}(x)\exp\left(-\mathrm{i}Y_{l_{\mathrm{L}}}g'\varphi(x)\right)\ , \\
e'_{\mathrm{R}}(x) &= \exp\left(\mathrm{i}Y_{e_{\mathrm{R}}}g'\varphi(x)\right)e_{\mathrm{R}}(x)\ , \quad \bar{e}'_{\mathrm{R}}(x) = \bar{e}_{\mathrm{R}}(x)\exp\left(-\mathrm{i}Y_{e_{\mathrm{R}}}g'\varphi(x)\right)
\end{aligned} \tag{5.3.3}$$

Here we assign weak hypercharges $Y_{l_{\mathrm{L}}}$ and $Y_{e_{\mathrm{R}}}$ to the left-handed leptons and the right-handed electron, respectively. Later, we will adjust the values of $Y_{l_{\mathrm{L}}}$ and $Y_{e_{\mathrm{R}}}$ such that the observed electric charges of electrons and neutrinos are reproduced correctly.

The left-handed neutrino and electron fields form an $SU(2)_{\mathrm{L}}$ *doublet*

$$l_{\mathrm{L}}(x) = \begin{pmatrix} \nu_{\mathrm{L}}(x) \\ e_{\mathrm{L}}(x) \end{pmatrix}\ , \qquad \bar{l}_{\mathrm{L}}(x) = \left(\bar{\nu}_{\mathrm{L}}(x), \bar{e}_{\mathrm{L}}(x)\right)\ . \tag{5.3.4}$$

The free lepton Lagrangian has another global symmetry which rotates the left-handed neutrino and electron fields into each other. In the Standard Model, this symmetry is again promoted to a local one

$$\begin{aligned}
l'_{\mathrm{L}}(x) &= \begin{pmatrix} \nu'_{\mathrm{L}}(x) \\ e'_{\mathrm{L}}(x) \end{pmatrix} = L(x)\begin{pmatrix} \nu_{\mathrm{L}}(x) \\ e_{\mathrm{L}}(x) \end{pmatrix} = L(x)l_{\mathrm{L}}(x)\ , \\
\bar{l}'_{\mathrm{L}}(x) &= \left(\bar{\nu}'_{\mathrm{L}}(x), \bar{e}'_{\mathrm{L}}(x)\right) = \left(\bar{\nu}_{\mathrm{L}}(x), \bar{e}_{\mathrm{L}}(x)\right)L(x)^\dagger = \bar{l}_{\mathrm{L}}(x)L(x)^\dagger\ ,
\end{aligned} \tag{5.3.5}$$

[3]It should be noted that the subscript L on ν_{L} and e_{L} refers to left, while in $U(1)_L$ it refers to the lepton number L.

with $L(x) \in SU(2)_{\mathrm{L}}$. The right-handed component of the electron field, $e_{\mathrm{R}}(x)$, on the other hand, is an $SU(2)_{\mathrm{L}}$ *singlet*, *i.e.* it remains invariant under $SU(2)_{\mathrm{L}}$ transformations

$$e'_{\mathrm{R}}(x) = e_{\mathrm{R}}(x) \ . \tag{5.3.6}$$

Since left- and right-handed fields transform differently under $SU(2)_{\mathrm{L}}$, also the $SU(2)_{\mathrm{L}}$ gauge symmetry is chiral.

In analogy to spin, one introduces a "*weak isospin*" which acts on the left-handed doublet as $T^3_{\mathrm{L}} = \frac{1}{2}\sigma^3$. The leptons have the following 3-components of the weak isospin

$$T^3_{\mathrm{L}\nu_{\mathrm{L}}} = \frac{1}{2} \ , \quad T^3_{\mathrm{L}e_{\mathrm{L}}} = -\frac{1}{2} \ , \quad T^3_{\mathrm{L}e_{\mathrm{R}}} = 0 \ . \tag{5.3.7}$$

Analogously, we introduce a generator T^3_{R} which takes the values

$$T^3_{\mathrm{R}\nu_{\mathrm{L}}} = 0 \ , \quad T^3_{\mathrm{R}e_{\mathrm{L}}} = 0 \ , \quad T^3_{\mathrm{R}e_{\mathrm{R}}} = -\frac{1}{2} \ . \tag{5.3.8}$$

This operator generates an Abelian subgroup of $SU(2)_{\mathrm{R}}$. Later we will also introduce a right-handed neutrino field $\nu_{\mathrm{R}}(x)$ for which

$$T^3_{\mathrm{L}\nu_{\mathrm{R}}} = 0 \ , \quad T^3_{\mathrm{R}\nu_{\mathrm{R}}} = \frac{1}{2} \ . \tag{5.3.9}$$

In the Standard model the $SU(2)_{\mathrm{L}}$ and $U(1)_Y$ (but not the full $SU(2)_{\mathrm{R}}$) symmetries are promoted to gauge symmetries. Just as in the gauge-Higgs Lagrangian of the Standard Model, this is achieved by substituting ordinary derivatives ∂_μ by covariant derivatives D_μ. For the left-handed lepton doublet the covariant derivative takes the form

$$D_\mu \begin{pmatrix} \nu_{\mathrm{L}}(x) \\ e_{\mathrm{L}}(x) \end{pmatrix} = \left[\partial_\mu + \mathrm{i}Y_{l_{\mathrm{L}}} g' B_\mu(x) + \mathrm{i}g W^a_\mu(x) \frac{\sigma^a}{2}\right] \begin{pmatrix} \nu_{\mathrm{L}}(x) \\ e_{\mathrm{L}}(x) \end{pmatrix} \ . \tag{5.3.10}$$

It should be noted that the derivative ∂_μ — as well as the gauge field term containing B_μ — act as unit 2×2 matrices in the flavor space. Using $W_\mu(x) = \mathrm{i}g W^a_\mu(x)\sigma^a/2$, the previous equation can also be written as

$$D_\mu l_{\mathrm{L}}(x) = [\partial_\mu + \mathrm{i}Y_{l_{\mathrm{L}}} g' B_\mu(x) + W_\mu(x)]\, l_{\mathrm{L}}(x) \ . \tag{5.3.11}$$

For the right-handed electron singlet the covariant derivative takes the form

$$D_\mu e_{\rm R}(x) = [\partial_\mu + {\rm i} Y_{e_{\rm R}} g' B_\mu(x)]\, e_{\rm R}(x) \ . \qquad (5.3.12)$$

The Lagrangian describing the propagation of the leptons as well as their interactions with the $U(1)_Y$ and $SU(2)_{\rm L}$ gauge fields then takes the form

$$\begin{aligned} \mathcal{L}(\bar\nu, \nu, \bar e, e, B_\mu, W_\mu) &= \bar l_{\rm L} \gamma_\mu D_\mu l_{\rm L} + \bar e_{\rm R} \gamma_\mu D_\mu e_{\rm R} \\ &= (\bar\nu_{\rm L}, \bar e_{\rm L}) \gamma_\mu D_\mu \left(\begin{array}{c} \nu_{\rm L}(x) \\ e_{\rm L}(x) \end{array} \right) + \bar e_{\rm R} \gamma_\mu D_\mu e_{\rm R} (5.3.13) \end{aligned}$$

In order to ensure gauge invariance, the $SU(2)_{\rm L}$ gauge coupling g has to take the same universal value as in the gauge-Higgs sector, which was discussed in Chapter 6.

It is important to note that a direct mass term $m_e(\bar e_{\rm L} e_{\rm R} + \bar e_{\rm R} e_{\rm L})$ is not gauge invariant, because the left- and right-handed electron fields transform differently under both $SU(2)_{\rm L}$ and $U(1)_Y$ gauge transformation. Consequently, direct mass terms are forbidden in the Standard Model. This is a nice feature of chiral gauge theories, because it protects the fermions from additive mass renormalization. Thus, in contrast to the scalar Higgs field, there is no hierarchy problem for chiral fermions, at least at the level of perturbation theory. Later we will construct Yukawa interaction terms between the fermions and the Higgs field. After spontaneous symmetry breaking, *i.e.* when the Higgs field picks up a non-zero vacuum expectation value v, such terms give rise to dynamically generated fermion masses. In this way, in the Standard Model with massless neutrinos all fermion masses are tied to the electroweak symmetry breaking scale v.

5.4 CP and T Invariance of Gauge Interactions

As we have seen, left-handed electrons and neutrinos are $SU(2)_{\rm L}$ doublets while right-handed electrons are singlets. Right-handed neutrino fields are not even introduced in the minimal version of the Standard Model. Consequently, left- and right-handed particles have different physical properties, which makes the Standard Model a *chiral* gauge theory. As a result of this asymmetric treatment of left- and right-handed degrees of freedom, parity P and charge conjugation C are explicitly broken in the Standard Model. Parity violation was predicted by Tsung-Dao Lee and Chen-Ning Yang in 1956 and indeed observed in weak interaction processes by Madame Chien-Shiung Wu in 1957. As we will now discuss, the gauge interactions still respect the combined discrete symmetry CP as well as the time-reversal T.[4]

Let us introduce the transformation behavior of the gauge fields under the discrete symmetries P, C, and T. The Abelian gauge field B_μ transforms as

$$\begin{aligned}
&{}^{\mathrm{P}}B_i(\vec{x},x_4) = -B_i(-\vec{x},x_4) \ , \quad {}^{\mathrm{P}}B_4(\vec{x},x_4) = B_4(-\vec{x},x_4) \ , \\
&{}^{\mathrm{C}}B_\mu(x) = -B_\mu(x) \ , \\
&{}^{\mathrm{T}}B_i(\vec{x},x_4) = -B_i(\vec{x},-x_4) \ , \quad {}^{\mathrm{T}}B_4(\vec{x},x_4) = B_4(\vec{x},-x_4) \ . \qquad (5.4.1)
\end{aligned}$$

Consequently, the combined transformations CP and CPT take the form

$$\begin{aligned}
&{}^{\mathrm{CP}}B_i(\vec{x},x_4) = B_i(-\vec{x},x_4) \ , \quad {}^{\mathrm{CP}}B_4(\vec{x},x_4) = -B_4(-\vec{x},x_4) \ , \\
&{}^{\mathrm{CPT}}B_\mu(x) = -B_\mu(-x) \ . \qquad (5.4.2)
\end{aligned}$$

Similarly, the non-Abelian gauge field W_μ transforms as

$$\begin{aligned}
&{}^{\mathrm{P}}W_i(\vec{x},x_4) = -W_i(-\vec{x},x_4) \ , \quad {}^{\mathrm{P}}W_4(\vec{x},x_4) = W_4(-\vec{x},x_4) \ , \\
&{}^{\mathrm{C}}W_\mu(x) = W_\mu(x)^* \ , \\
&{}^{\mathrm{T}}W_i(\vec{x},x_4) = W_i(\vec{x},-x_4)^* \ , \quad {}^{\mathrm{T}}W_4(\vec{x},x_4) = -W_4(\vec{x},-x_4)^* \ , \qquad (5.4.3)
\end{aligned}$$

which implies

$$\begin{aligned}
&{}^{\mathrm{CP}}W_i(\vec{x},x_4) = -W_i(-\vec{x},x_4)^* \ , \quad {}^{\mathrm{CP}}W_4(\vec{x},x_4) = W_4(-\vec{x},x_4)^* \ , \\
&{}^{\mathrm{CPT}}W_\mu(x) = -W_\mu(-x) \ . \qquad (5.4.4)
\end{aligned}$$

Let us now investigate the CP transformation properties of the interaction terms that couple the right-handed electron to the $U(1)_Y$ gauge field.

[4]As we will discuss in Chapter 10, with three or more generations of fermions, the fermion-Higgs couplings explicitly violate CP and T. Furthermore, as we will discuss in Chapter ???, the QCD vacuum angle θ is another source of explicit CP and T breaking. However, in Nature this parameter is consistent with zero.

Using the CP transformation rules for the fermions of Eq.(5.2.12), we obtain

$$\begin{aligned}
S[{}^{\mathrm{CP}}\bar{e}_{\mathrm{R}}, {}^{\mathrm{CP}}e_{\mathrm{R}}, {}^{\mathrm{CP}}B_\mu] &= \int d^4x \ {}^{\mathrm{CP}}\bar{e}_{\mathrm{R}}(\vec{x}, x_4)\gamma_\mu \mathrm{i}Y_{l_{\mathrm{L}}} g' \ {}^{\mathrm{CP}}B_\mu(\vec{x}, x_4) {}^{\mathrm{CP}}e_{\mathrm{R}}(\vec{x}, x_4) \\
&= -\int d^4x \ e_{\mathrm{R}}(-\vec{x}, x_4)^{\mathsf{T}} P^{\mathsf{T}} C^{-1} \left[-\gamma_i \mathrm{i}Y_{l_{\mathrm{L}}} g' B_i(-\vec{x}, x_4)\right. \\
&+ \left. \gamma_4 \mathrm{i}Y_{l_{\mathrm{L}}} g' B_4(-\vec{x}, x_4)\right] CP\bar{e}_{\mathrm{R}}(-\vec{x}, x_4)^{\mathsf{T}} \\
&= \int d^4x \ e_{\mathrm{R}}(-\vec{x}, x_4)^{\mathsf{T}} \left[\gamma_i^{\mathsf{T}} \mathrm{i}Y_{l_{\mathrm{L}}} g' B_i(-\vec{x}, x_4)\right. \\
&+ \left. \gamma_4^{\mathsf{T}} \mathrm{i}Y_{l_{\mathrm{L}}} g' B_4(-\vec{x}, x_4)\right] \bar{e}_{\mathrm{R}}(-\vec{x}, x_4)^{\mathsf{T}} \\
&= \int d^4x \ \bar{e}_{\mathrm{R}}(-\vec{x}, x_4)\gamma_\mu \mathrm{i}Y_{l_{\mathrm{L}}} g' B_\mu(-\vec{x}, x_4) e_{\mathrm{R}}(-\vec{x}, x_4) \\
&= S[\bar{e}_{\mathrm{R}}, e_{\mathrm{R}}, B_\mu] \ . \qquad (5.4.5)
\end{aligned}$$

In the same manner, one can show that the couplings of the left-handed leptons are also CP-invariant. Due to the CPT theorem, the interaction terms are automatically CPT- and thus (due to CP invariance) also T-invariant.

Finally, we list the C, P, and T transformation properties of the Higgs field

$$\begin{aligned}
&{}^{\mathrm{P}}\Phi(\vec{x}, x_4) = \Phi(-\vec{x}, x_4) \ , \\
&{}^{\mathrm{C}}\Phi(x) = \Phi(x)^* \ , \\
&{}^{\mathrm{T}}\Phi(\vec{x}, x_4) = \Phi(\vec{x}, -x_4)^* \ , \qquad (5.4.6)
\end{aligned}$$

which then implies

$$\begin{aligned}
&{}^{\mathrm{CP}}\Phi(\vec{x}, x_4) = \Phi(-\vec{x}, x_4)^* \ , \\
&{}^{\mathrm{CPT}}\Phi(x) = \Phi(-x) \ . \qquad (5.4.7)
\end{aligned}$$

Using these transformation rules, it is straightforward to return to the gauge-Higgs sector and show that the action

$$S[\Phi, W_\mu, B_\mu] = \int d^4x \left[\frac{1}{2} D_\mu \Phi^\dagger D_\mu \Phi + V(\Phi) - \frac{1}{2g^2} \mathrm{Tr}(W_{\mu\nu} W_{\mu\nu}) + \frac{1}{4} B_{\mu\nu} B_{\mu\nu}\right] \qquad (5.4.8)$$

is invariant separately under C, P, and T.

5.5 Fixing the Lepton Weak Hypercharges

We know that the electron carries electric charge $-e$, while the neutrino is electrically neutral. In the Lagrangian (5.3.13) we recognize off-diagonal terms that couple leptons of different electric charge, associated with W^1 and W^2. In order to preserve the electric charge under interactions, these gauge bosons must be charged themselves. In particular, we find a positive and a negative W-boson given by

$$W_\mu^\pm(x) = \frac{1}{\sqrt{2}} \left(W_\mu^1(x) \mp \mathrm{i} W_\mu^2(x) \right) \ , \tag{5.5.1}$$

which implies

$$W_\mu^1(x)\frac{\sigma^1}{2} + W_\mu^2(x)\frac{\sigma^2}{2} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & W_\mu^+(x) \\ W_\mu^-(x) & 0 \end{pmatrix} = \frac{1}{\sqrt{2}} (W_\mu^+(x)\sigma^+ + W_\mu^-(x)\sigma^-) \ , \tag{5.5.2}$$

where $\sigma^\pm = \frac{1}{2}(\sigma^1 \pm \mathrm{i}\sigma^2)$.

We observed before (in the Higgs sector) that the electrically neutral gauge fields, *i.e.* the flavor diagonal fields, split physically into a massless photon and a massive Z-boson

$$A_\mu(x) = \frac{g' W_\mu^3(x) + g B_\mu(x)}{\sqrt{g^2 + g'^2}} \ , \ Z_\mu(x) = \frac{g W_\mu^3(x) - g' B_\mu(x)}{\sqrt{g^2 + g'^2}} \ . \tag{5.5.3}$$

They are natural to consider after spontaneous symmetry breaking. Inserting the inverse relations

$$W_\mu^3(x) = \frac{g' A_\mu(x) + g Z_\mu(x)}{\sqrt{g^2 + g'^2}} \ , \ B_\mu(x) = \frac{g A_\mu(x) - g' Z_\mu(x)}{\sqrt{g^2 + g'^2}} \ , \tag{5.5.4}$$

we can write the lepton-gauge coupling terms in the Lagrangian (5.3.13) as

$$\begin{aligned} \mathcal{L}(\bar\nu, \nu, \bar e, e, A_\mu, Z_\mu) &= (\bar\nu_\mathrm{L}, \bar e_\mathrm{L})\gamma_\mu \Big[\partial_\mu + \mathrm{i} \begin{pmatrix} X_\mu^1 & \frac{g}{\sqrt{2}} W_\mu^+ \\ \frac{g}{\sqrt{2}} W_\mu^- & X_\mu^2 \end{pmatrix} \Big] \begin{pmatrix} \nu_\mathrm{L} \\ e_\mathrm{L} \end{pmatrix} \\ &+ \ \bar e_\mathrm{R} \gamma_\mu \Big[\partial_\mu + \mathrm{i} \frac{Y_{e_\mathrm{R}} g'}{\sqrt{g^2 + g'^2}} (g A_\mu - g' Z_\mu)\Big] e_R \ , \end{aligned} \tag{5.5.5}$$

where

$$X_\mu^1(x) = \frac{1}{\sqrt{g^2+g'^2}}\left[gg'\left(\frac{1}{2}+Y_{l_\mathrm{L}}\right)A_\mu(x)+\left(\frac{1}{2}g^2-Y_{l_\mathrm{L}}g'^2\right)Z_\mu(x)\right] ,$$
$$X_\mu^2(x) = \frac{1}{\sqrt{g^2+g'^2}}\left[gg'\left(-\frac{1}{2}+Y_{l_\mathrm{L}}\right)A_\mu(x)+\left(-\frac{1}{2}g^2-Y_{l_\mathrm{L}}g'^2\right)Z_\mu(x)\right] . \tag{5.5.6}$$

Since the neutrino does not couple to the photon field A_μ, the term X_μ^1 must not contain a contribution from A_μ. This implies $Y_{l_\mathrm{L}} = -1/2$, and therefore

$$\begin{aligned} X_\mu^1(x) &= \frac{\sqrt{g^2+g'^2}}{2}Z_\mu(x) , \\ X_\mu^2(x) &= \frac{1}{\sqrt{g^2+g'^2}}\left[\frac{g'^2-g^2}{2}Z_\mu(x)-gg'A_\mu(x)\right] \\ &= -\frac{\sqrt{g^2+g'^2}}{2}\left[\cos(2\theta_\mathrm{W})Z_\mu(x)+\sin(2\theta_\mathrm{W})A_\mu(x)\right] . \end{aligned} \tag{5.5.7}$$

Here θ_W is the Weinberg angle introduced in Eq. (4.2.31). Again, we identify

$$e = \frac{gg'}{\sqrt{g^2+g'^2}} \tag{5.5.8}$$

as the unit of electric charge, in exact agreement with Eq. (4.2.35). Indeed, $-e$ is the correct electric charge of the left-handed electron. In order to obtain the same value $-e$ also for the right-handed electron, we now adjust its weak hypercharge to $Y_{e_\mathrm{R}} = -1$. We now see that Y_{l_L} and Y_{e_R} are different. Consequently, not only the $SU(2)_\mathrm{L}$ but also the $U(1)_Y$ gauge couplings are chiral.

At this point, we observe a simple relation between the weak hypercharge Y, (*i.e.* the coupling to B_μ in units of g'), the generator T_R^3, which was introduced in Eq. (5.3.8), and the lepton number L,

$$Y = T_\mathrm{R}^3 - \frac{1}{2}L . \tag{5.5.9}$$

For the left-handed neutrino and the left-handed electron, which both have lepton number $L = 1$, this equation takes the form

$$Y_{l_\mathrm{L}} = 0 - \frac{1}{2} = -\frac{1}{2} , \tag{5.5.10}$$

and for the right-handed electron, which again has $L = 1$, it reads

$$Y_{e_\mathrm{R}} = -\frac{1}{2} - \frac{1}{2} = -1 \ . \tag{5.5.11}$$

Furthermore, the electric charge Q (in units of e) is related to Y and the third component of the weak isospin T^3_L, which was introduced in Eq. (5.3.7)

$$Q = T^3_\mathrm{L} + Y = T^3_\mathrm{L} + T^3_\mathrm{R} - \frac{1}{2}L \ . \tag{5.5.12}$$

For the left-handed neutrino this equation takes the form

$$Q_{\nu_\mathrm{L}} = T^3_{\mathrm{L}\nu_\mathrm{L}} + Y_{l_\mathrm{L}} = \frac{1}{2} - \frac{1}{2} = 0 \ , \tag{5.5.13}$$

for the left-handed electron it reads

$$Q_{e_\mathrm{L}} = T^3_{\mathrm{L}e_\mathrm{L}} + Y_{l_\mathrm{L}} = -\frac{1}{2} - \frac{1}{2} = -1 \ , \tag{5.5.14}$$

and at last for the right-handed electron

$$Q_{e_\mathrm{R}} = T^3_{\mathrm{L}e_\mathrm{R}} + Y_{e_\mathrm{R}} = 0 - 1 = -1 \ . \tag{5.5.15}$$

In the following, relation (5.5.12) will be given a prominent status. We remark here that its validity is also a consequence of the parameter choice $Y_{l_\mathrm{L}} = -1/2$ that we made in order to decouple the neutrino from A_μ.

We can also interpret these expressions in terms of gauge couplings to *fermionic currents*. Generally, currents are 4-vectors $j_\mu(x)$ obeying the continuity equation $\partial_\mu j_\mu = 0$, at least at the classical level. From the Lagrangian of a free fermion, we obtain the Noether current $\bar{\psi}\gamma_\mu\psi$. Its continuity can also be derived from the free Dirac equation (**??**) and its adjoint, Eq. (**??**). According to the interpretation elaborated by Wolfgang Pauli and Victor Weisskopf, we should consider currents of charge instead of probability. The electromagnetic current of the electron with charge $-e$ amounts to

$$j^\mathrm{em}_\mu = -e\left(\bar{e}_\mathrm{L}\gamma_\mu e_\mathrm{L} + \bar{e}_\mathrm{R}\gamma_\mu e_\mathrm{R}\right) \ . \tag{5.5.16}$$

In these terms, the photon field couples to the electromagnetic current via the term $\mathrm{i}A_\mu j^\mathrm{em}_\mu$ in the Lagrangian. It should be noted that this current is *neutral*. This means that there is no change in the charge between the initial and the final fermion states, *i.e.* the state before and after the scattering on the photon. Clearly, the charge cannot change because the photon is neutral.

Besides the couplings to the photon, we also have *weak current interactions*. The *weak neutral current* j^0_μ couples to the Z-boson field,

$$\begin{aligned} j^0_\mu &= \frac{\sqrt{g^2+g'^2}}{2}\bar{\nu}_\mathrm{L}\gamma_\mu\nu_\mathrm{L} + \frac{g'^2-g^2}{2\sqrt{g^2+g'^2}}\bar{e}_\mathrm{L}\gamma_\mu e_\mathrm{L} + \frac{g'^2}{\sqrt{g^2+g'^2}}\bar{e}_\mathrm{R}\gamma_\mu e_\mathrm{R} \\ &= \frac{\sqrt{g^2+g'^2}}{2}\left(\bar{\nu}_\mathrm{L}\gamma_\mu\nu_\mathrm{L} - \cos(2\theta_\mathrm{W})\bar{e}_\mathrm{L}\gamma_\mu e_\mathrm{L} + \sin^2\theta_\mathrm{W}\bar{e}_\mathrm{R}\gamma_\mu e_\mathrm{R}\right) \end{aligned} \quad (5.5.17)$$

The *charged currents*, on the other hand, couple to the charged gauge bosons $W^\pm$ and take the form

$$j^+_\mu = \frac{g}{2}\bar{\nu}_\mathrm{L}\gamma_\mu e_\mathrm{L}\,, \quad j^-_\mu = \frac{g}{2}\bar{e}_\mathrm{L}\gamma_\mu\nu_\mathrm{L}\ . \quad (5.5.18)$$

For the current j^+_μ (or j^-_μ) the charge increases (or decreases) in the transition from the initial to the final state. Note that the neutral currents contain both left- and right-handed contributions, while the charged currents are purely left-handed.

This set of charged and neutral currents enables a number of physical transitions, such as the decays $W^- \to e_\mathrm{L} + \bar{\nu}_\mathrm{L}$, $Z \to \nu_\mathrm{L} + \bar{\nu}_\mathrm{L}$, or $Z \to e_\mathrm{L} + \bar{e}_\mathrm{L}$. As a general principle, particles tend to decay into lighter particles if there is no conservation law preventing such a decay. Nevertheless, for instance the decay $Z \to e_\mathrm{L} + \bar{e}_\mathrm{L}$ can also be inverted: an electron-positron pair collides at very high energy and generates a Z-boson, which will very soon again decay into leptons. In the electron-positron scattering amplitude this channel is then visible as a resonance at the energy that is needed for the Z-mass. The coupling of Z to the weak neutral current j^0_μ also describes the scattering of a neutrino or an electron off a Z-boson.

5.6 Gauge Anomalies in the Lepton Sector

As it stands, the Standard Model with just electrons and neutrinos is inconsistent because it suffers from *anomalies* in its gauge interactions. Anomalies represent a form of explicit symmetry breaking due to quantum effects, while at the classical level the corresponding symmetry is exact. Anomalies usually arise from a non-invariance of the functional measure, while the action of the theory is invariant. *Gauge anomalies* represent an explicit violation of gauge invariance. Since gauge invariance is vital for eliminating redundant gauge-dependent degrees of freedom, theories with an explicitly broken gauge symmetry are mathematically and physically inconsistent. In order to render the Standard Model consistent, the gauge anomalies of the leptons must be cancelled by other fields. The quarks, which participate in both the electroweak and the strong interactions, serve this purpose. As we will see later, in contrast to gauge anomalies, anomalies in global symmetries are perfectly acceptable and even necessary to describe the physics correctly.

In the Standard Model, there are different types of gauge anomalies that must be cancelled. First, there is a triangle anomaly in the $U(1)_Y$ gauge interaction which manifests itself already within (but also beyond) perturbation theory. One considers the interaction with a triangle built from fermion propagators; each corner is a vertex with a coupling to an external gauge field B_μ. The interaction at each vertex is proportional to the weak hypercharge Y of the fermion that propagates around the triangle. Hence the total contribution is proportional to Y^3. The full amplitude of these triangle diagrams must be symmetric if we perform an overall flip from left- to right-handedness, *i.e.* the antisymmetric quantity

$$A = \sum_{\mathrm{L}} Y^3 - \sum_{\mathrm{R}} Y^3 \ , \tag{5.6.1}$$

which is proportional to the anomaly, is supposed to vanish. The sums extend over the left- and right-handed degrees of freedom, respectively.[5] Since the left-handed neutrino and electron carry the weak hypercharge $Y_{l_\mathrm{L}} = -1/2$, while the right-handed electron has $Y_{e_\mathrm{R}} = -1$, the $U(1)_Y$ triangle anomaly in the lepton sector is given by

$$A_l = 2Y_{l_\mathrm{L}}^3 - Y_{e_\mathrm{R}}^3 = 2\left(-\frac{1}{2}\right)^3 - (-1)^3 = \frac{3}{4} \neq 0 \ . \tag{5.6.2}$$

As we will see later, this non-zero anomaly in the lepton sector will be cancelled by a corresponding anomaly in the quark sector.

[5] We do not present an evaluation of this triangle diagram. A detailed explanation of this calculation can be found *e.g.* in [**?**].

The general expression for $SU(2)_\mathrm{L} \times U(1)_Y$ triangle anomalies is given by

$$A^{abc} = \mathrm{Tr}_\mathrm{L}\left[(T^aT^b + T^bT^a)T^c\right] - \mathrm{Tr}_\mathrm{R}\left[(T^aT^b + T^bT^a)T^c\right] \ . \tag{5.6.3}$$

The T^a with $a \in \{1,2,3\}$ refer to the generators of $SU(2)_\mathrm{L}$ and $T^4 = Y$. If all three indices a, b, and c are equal to 4, we recover the $U(1)_Y$ anomaly discussed above, *i.e.* $A^{444} = 2A$. If one index belongs to $\{1,2,3\}$ and the other two are equal to 4, the tracelessness of the $SU(2)_\mathrm{L}$ generators leads to a vanishing anomaly. Similarly, if all three indices belong to $\{1,2,3\}$, the Pauli matrix identity,

$$\mathrm{Tr}\left[(\sigma^a\sigma^b + \sigma^b\sigma^a)\sigma^c\right] = 2\delta_{ab}\mathrm{Tr}\sigma^c = 0 \ , \tag{5.6.4}$$

again leads to a vanishing anomaly. However, if two indices belong to $\{1,2,3\}$ while the third one, say c, is equal to 4, the anomaly takes the form

$$A^{ab4} = \mathrm{Tr}_\mathrm{L}\left[\frac{1}{4}(\sigma^a\sigma^b + \sigma^b\sigma^a)Y\right] = \delta_{ab}\mathrm{Tr}_\mathrm{L}Y \ . \tag{5.6.5}$$

Here, we have used the fact that in the Standard Model the left-handed fermions are $SU(2)_\mathrm{L}$ doublets (*i.e.* $T^a = \frac{1}{2}\sigma^a$), while the right-handed fermions are $SU(2)_\mathrm{L}$ singlets (*i.e.* $T^a = 0$). In the lepton sector, the corresponding anomaly is given by

$$A_l^{ab4} = \delta_{ab} 2Y_{l_\mathrm{L}} = -\delta_{ab} \neq 0 \ , \tag{5.6.6}$$

which thus gives rise to another inconsistency.

In addition to the triangle anomalies there is a *"global gauge anomaly"* in the $SU(2)_\mathrm{L}$ gauge interactions, which was discovered by Edward Witten [**?**]. It should be stressed that here "global" does not refer to a global symmetry. Instead it refers to the global topological properties of $SU(2)_\mathrm{L}$ gauge transformation. Two gauge transformations are considered topologically equivalent if they can be deformed continuously into one another. The corresponding equivalence classes are known as homotopy groups.[6] In four space-time dimensions the homotopy group of $SU(2)_\mathrm{L}$ gauge transformations is

$$\Pi_4[SU(2)] = \Pi_4[S^3] = \mathbf{Z}(2) \ , \tag{5.6.7}$$

[6]Homotopy groups are discussed in some detail in Appendix ???.

i.e. these gauge transformations fall into two distinct topological classes. The topologically trivial class contains all gauge transformations that can be continuously deformed into the gauge transformation $L(x) = \mathbf{1}$. The topologically non-trivial class contains all other gauge transformations. As Edward Witten first realized, the fermionic measure of each doublet in an $SU(2)_{\mathrm{L}}$ gauge theory changes sign under topologically non-trivial $SU(2)_{\mathrm{L}}$ gauge transformations. Hence, in order to be gauge invariant, the theory must contain an even number of doublets. Since the lepton sector of the first generation of the Standard Model contains a single $SU(2)_{\mathrm{L}}$ doublet (consisting of the left-handed electron and neutrino), it suffers from Witten's global gauge anomaly. In order to cancel this anomaly, we must add an *odd number of* $SU(2)_{\mathrm{L}}$ *doublets*. Since it is associated with "large" gauge transformations, which are not located in the neighborhood of the identity, the global gauge anomaly manifests itself only beyond perturbation theory. In particular, it is not visible in perturbative triangle diagrams.

5.7 Up and Down Quarks

In the first fermion generation there are also the *up and down quarks* which will come to our rescue and cancel both, the triangle anomalies and the global gauge anomaly. The quarks are massive and thus require the introduction of *left- and right-handed fields*.[7] In addition to the electroweak interaction, the quarks participate in the strong interactions and thus they carry an $SU(N_{\mathrm{c}})$ *color charge*. The color index on a quark field then takes values $c \in \{1, 2, \dots, N_{\mathrm{c}}\}$. In the real world the number of colors is $N_{\mathrm{c}} = 3$. However, as we will see, a consistent variant of the Standard Model can be formulated with any odd number of colors. There are some misconceptions about this issue in most of the textbook literature. In order to illuminate this point, we keep N_{c} general in this and some other chapters. The left-handed up and down quark fields (with color index $c \in \{1, 2, \dots, N_{\mathrm{c}}\}$) then

[7]For some time there was a controversy whether the up quark mass might vanish. However, this turned out to be inconsistent with experiment. Still, even if some quarks were massless, one would need to introduce both, left- and right-handed quark fields, in order to achieve anomaly cancellation.

form N_c different $SU(2)_L$ doublets

$$q^c_L(x) = \begin{pmatrix} u^c_L(x) \\ d^c_L(x) \end{pmatrix} , \quad \bar{q}^c_L(x) = \left(\bar{u}^c_L(x), \bar{d}^c_L(x)\right) , \tag{5.7.1}$$

or equivalently two $SU(N_c)$ color N_c-plets. The right-handed quarks u^c_R and d^c_R form again two $SU(N_c)$ color N_c-plets, but they are $SU(2)_L$ singlets. Since we have added N_c left-handed $SU(2)_L$ quark doublets, in order to cancel the global gauge anomaly of the lepton sector, the number of colors N_c must be odd in the Standard Model.[8] In complete analogy to the lepton sector, in the quark sector the generators T^3_L and T^3_R take the values

$$\begin{aligned} &T^3_{\mathrm{L}u_\mathrm{L}} = \frac{1}{2} \, , \; T^3_{\mathrm{L}d_\mathrm{L}} = -\frac{1}{2} \, , \; T^3_{\mathrm{L}u_\mathrm{R}} = 0 \, , \; T^3_{\mathrm{L}d_\mathrm{R}} = 0 \, , \\ &T^3_{\mathrm{R}u_\mathrm{L}} = 0 \, , \; T^3_{\mathrm{R}d_\mathrm{L}} = 0 \, , \; T^3_{\mathrm{R}u_\mathrm{R}} = \frac{1}{2} \, , \; T^3_{\mathrm{R}d_\mathrm{R}} = -\frac{1}{2} \, . \end{aligned} \tag{5.7.2}$$

Using an Einstein summation convention for the color index c, the Lagrangian for free massless quarks,

$$\mathcal{L}_0(\bar{u}, u, \bar{d}, d) = \bar{u}^c_L \gamma_\mu \partial_\mu u^c_L + \bar{u}^c_R \gamma_\mu \partial_\mu u^c_R + \bar{d}^c_L \gamma_\mu \partial_\mu d^c_L + \bar{d}^c_R \gamma_\mu \partial_\mu d^c_R , \tag{5.7.3}$$

has a global $U(1)_B$ symmetry which acts by multiplying all quark fields by the same phase

$$\begin{aligned} u^c_L{}'(x) &= \exp(\mathrm{i}\rho/N_c) u^c_L(x) , \quad \bar{u}^c_L{}'(x) = \bar{u}^c_L(x) \exp(-\mathrm{i}\rho/N_c) , \\ u^c_R{}'(x) &= \exp(\mathrm{i}\rho/N_c) u^c_R(x) , \quad \bar{u}^c_R{}'(x) = \bar{u}^c_R(x) \exp(-\mathrm{i}\rho/N_c) , \\ d^c_L{}'(x) &= \exp(\mathrm{i}\rho/N_c) d^c_L(x) , \quad \bar{d}^c_L{}'(x) = \bar{d}^c_L(x) \exp(-\mathrm{i}\rho/N_c) , \\ d^c_R{}'(x) &= \exp(\mathrm{i}\rho/N_c) d^c_R(x) , \quad \bar{d}^c_R{}'(x) = \bar{d}^c_R(x) \exp(-\mathrm{i}\rho/N_c) . \end{aligned} \tag{5.7.4}$$

Analogous to lepton number, the corresponding conserved charge is the quark number, or equivalently the baryon number B. Each baryon contains N_c confined quarks, and hence the baryon number of a quark is $B = 1/N_c$.

We still need to assign weak hypercharges to the quark fields. $SU(2)_L$ gauge invariance requires that the left-handed up and down quarks carry

[8] Here we assume that the anomalies are cancelled within a single generation of fermions. If the number of generations would be even, the global gauge anomaly would also cancel for even N_c. However, since baryons (which consist of N_c quarks) would then be bosons, the resulting physics would be drastically different from the real world.

the same charge Y_{q_L}. On the other hand, since the right-handed quarks are $SU(2)_\mathrm{L}$ singlets, up and down may then carry different hypercharges Y_{u_R} and Y_{u_R}. The $U(1)_Y$ gauge transformations then act as

$$\begin{aligned}
u_\mathrm{L}^c{}'(x) &= \exp(\mathrm{i}Y_{q_\mathrm{L}}g'\varphi(x))u_\mathrm{L}^c(x) \ , \quad \bar u_\mathrm{L}^c{}'(x) = \bar u_\mathrm{L}(x)^c \exp(-\mathrm{i}Y_{q_\mathrm{L}}g'\varphi(x)) \ , \\
u_\mathrm{R}^c{}'(x) &= \exp(\mathrm{i}Y_{u_\mathrm{R}}g'\varphi(x))u_\mathrm{R}^c(x) \ , \quad \bar u_\mathrm{R}^c{}'(x) = \bar u_\mathrm{R}^c(x) \exp(-\mathrm{i}Y_{u_\mathrm{R}}g'\varphi(x)) \ , \\
d_\mathrm{L}^c{}'(x) &= \exp(\mathrm{i}Y_{q_\mathrm{L}}g'\varphi(x))d_\mathrm{L}^c(x) \ , \quad \bar d_\mathrm{L}^c{}'(x) = \bar d_\mathrm{L}^c(x) \exp(-\mathrm{i}Y_{q_\mathrm{L}}g'\varphi(x)) \ , \\
d_\mathrm{R}^c{}'(x) &= \exp(\mathrm{i}Y_{d_\mathrm{R}}g'\varphi(x))d_\mathrm{R}^c(x) \ , \quad \bar d_\mathrm{R}^c{}'(x) = \bar d_\mathrm{R}^c(x) \exp(-\mathrm{i}Y_{d_\mathrm{R}}g'\varphi(x))
\end{aligned} \tag{5.7.5}$$

Under $SU(2)_\mathrm{L}$ gauge transformations the quark fields transform as

$$\begin{aligned}
q_\mathrm{L}^c{}'(x) &= \left(\begin{array}{c} u_\mathrm{L}^c{}'(x) \\ d_\mathrm{L}^c{}'(x) \end{array}\right) = L(x) \left(\begin{array}{c} u_\mathrm{L}^c(x) \\ d_\mathrm{L}^c(x) \end{array}\right) = L(x) q_\mathrm{L}^c(x) \ , \\
\bar q_\mathrm{L}^c{}'(x) &= \left(\bar u_\mathrm{L}^c{}'(x), \bar d_\mathrm{L}^c{}'(x)\right) = \left(\bar u_\mathrm{L}^c(x), \bar d_\mathrm{L}^c(x)\right) L(x)^\dagger = \bar q_\mathrm{L}^c(x) L(x)^\dagger \ , \\
u_\mathrm{R}^c{}'(x) &= u_\mathrm{R}^c(x) \ , \ d_\mathrm{R}^c{}'(x) = d_\mathrm{R}^c(x) \ , \\
\bar u_\mathrm{R}^c{}'(x) &= \bar u_\mathrm{R}^c(x) \ , \ \bar d_\mathrm{R}^c{}'(x) = \bar d_\mathrm{R}^c(x) \ .
\end{aligned} \tag{5.7.6}$$

Before the quarks are coupled to the gluons, they do not yet participate in the strong interaction. The gluons are then still strongly interacting among each other, and are confined inside glueballs, but they decouple from the other fields.[9] Without quark-gluon couplings, the quarks are not confined inside hadrons but represent physical states. Such a world may be considered a theorist's paradise, because the physics would be mostly perturbative and thus analytically calculable.

We will now switch on the quark-gluon coupling. While the real world, in which quarks are confined, is much more interesting than the theorist's paradise, it will also turn out to be much more difficult to understand. In particular, since strong non-perturbative effects then dominate at low energies, perturbation theory breaks down, and quantitative results can often be obtained only by means of numerical Monte Carlo calculations. Suppressing color indices, under gauge transformations $\Omega(x) \in SU(N_\mathrm{c})$ the quark fields transform as

$$\begin{aligned}
q_\mathrm{L}'(x) &= \Omega(x) q_\mathrm{L}(x) \ , \ \bar q_\mathrm{L}'(x) = \bar q_\mathrm{L}(x) \Omega(x)^\dagger \ , \\
u_\mathrm{R}'(x) &= \Omega(x) u_\mathrm{R}(x) \ , \ \bar u_\mathrm{R}'(x) = \bar u_\mathrm{R}(x) \Omega(x)^\dagger \ , \\
d_\mathrm{R}'(x) &= \Omega(x) d_\mathrm{R}(x) \ , \ \bar d_\mathrm{R}'(x) = \bar d_\mathrm{R}(x) \Omega(x)^\dagger \ ,
\end{aligned} \tag{5.7.7}$$

[9]In the absence of quark-gluon couplings, only gravity would establish communication between gluons and the rest of the world.

i.e. the quark fields q transform in the fundamental $\{N_\mathrm{c}\}$ representation of $SU(N_\mathrm{c})$, while the anti-quarks $\bar{q}$ transform in the anti-fundamental $\{\overline{N_\mathrm{c}}\}$ representation. Making the color indices c explicit, for the left-handed quark doublet the covariant derivative takes the form

$$D_\mu q_\mathrm{L}^c(x) = \left[\left(\partial_\mu + \mathrm{i}Y_{q_\mathrm{L}} g' B_\mu(x) + \mathrm{i}g W_\mu^a(x)\frac{\sigma^a}{2}\right)\delta_{cc'} + \mathrm{i}g_\mathrm{s} G_\mu^a \frac{\lambda^a_{cc'}}{2}\right] q_\mathrm{L}^{c'}(x) \ . \tag{5.7.8}$$

Here λ^a with $a \in \{1, 2, \dots, N_\mathrm{c}^2 - 1\}$ are the generators of $SU(N_\mathrm{c})$ (the eight Gell-Mann matrices for $N_\mathrm{c} = 3$ displayed in Appendix ???). Using $W_\mu(x) = \mathrm{i}g_\mathrm{s} W_\mu^a(x)\sigma^a/2$ and $G_\mu(x) = \mathrm{i}g_\mathrm{s} G_\mu^a(x)\lambda^a/2$ one can also write

$$D_\mu q_\mathrm{L}(x) = \left[\partial_\mu + \mathrm{i}Y_{q_\mathrm{L}} g' B_\mu(x) + W_\mu(x) + G_\mu(x)\right] q_\mathrm{L}(x) \ . \tag{5.7.9}$$

For the right-handed quark singlets the covariant derivatives are given by

$$\begin{aligned} D_\mu u_\mathrm{R}^c(x) &= \left[\left(\partial_\mu + \mathrm{i}Y_{u_\mathrm{R}} g' B_\mu(x)\right)\delta_{cc'} + \mathrm{i}g_\mathrm{s} G_\mu^a(x)\frac{\lambda^a_{cc'}}{2}\right] u_\mathrm{R}^{c'}(x) \ , \\ D_\mu d_\mathrm{R}^c(x) &= \left[\left(\partial_\mu + \mathrm{i}Y_{d_\mathrm{R}} g' B_\mu(x)\right)\delta_{cc'} + \mathrm{i}g_\mathrm{s} G_\mu^a(x)\frac{\lambda^a_{cc'}}{2}\right] d_\mathrm{R}^{c'}(x) \ , \end{aligned} \tag{5.7.10}$$

or alternatively, suppressing the color indices,

$$\begin{aligned} D_\mu u_\mathrm{R}(x) &= \left[\partial_\mu + \mathrm{i}Y_{u_\mathrm{R}} g' B_\mu(x) + G_\mu(x)\right] u_\mathrm{R}(x) \ , \\ D_\mu d_\mathrm{R}(x) &= \left[\partial_\mu + \mathrm{i}Y_{d_\mathrm{R}} g' B_\mu(x) + G_\mu(x)\right] d_\mathrm{R}(x) \ . \end{aligned} \tag{5.7.11}$$

The Lagrangian describing the propagation of the quarks as well as their interactions with the $U(1)_Y$, $SU(2)_\mathrm{L}$, and $SU(N_\mathrm{c})$ gauge fields then takes the form

$$\begin{aligned} \mathcal{L}(\bar{u}, u, \bar{d}, d, B_\mu, W_\mu, G_\mu) &= \bar{q}_\mathrm{L}\gamma_\mu D_\mu q_\mathrm{L} + \bar{u}_\mathrm{R}\gamma_\mu D_\mu u_\mathrm{R} + \bar{d}_\mathrm{R}\gamma_\mu D_\mu d_\mathrm{R} \\ &= (\bar{u}_\mathrm{L}, \bar{d}_\mathrm{L})\gamma_\mu D_\mu \begin{pmatrix} u_\mathrm{L}(x) \\ d_\mathrm{L}(x) \end{pmatrix} + \bar{u}_\mathrm{R}\gamma_\mu D_\mu u_\mathrm{R} + \bar{d}_\mathrm{R}\gamma_\mu D_\mu d_\mathrm{R} \ . \end{aligned} \tag{5.7.12}$$

5.8 Anomaly Cancellation

In complete analogy to the leptons, the quarks also contribute to the various triangle anomalies. First of all, the quark triangle diagram with external

$U(1)_Y$ bosons attached to all three vertices contributes

$$A_q^{444} = N_\mathrm{c}\left(2Y_{q_\mathrm{L}}^3 - Y_{u_\mathrm{R}}^3 - Y_{d_\mathrm{R}}^3\right) \ . \tag{5.8.1}$$

For the same reasons as in the lepton sector, the triangle diagrams with one or three external $SU(2)_\mathrm{L}$ gauge bosons vanish. The diagram with two external $SU(2)_\mathrm{L}$ and one external $U(1)_Y$ boson, on the other hand, is non-zero and contributes

$$A_q^{ab4} = \delta_{ab} 2N_\mathrm{c} Y_{q_\mathrm{L}} \ , \quad a, b \in \{1, 2, 3\} \ . \tag{5.8.2}$$

In order to cancel the triangle anomalies in the lepton sector we now demand

$$\begin{aligned} &A_l^{ab4} + A_q^{ab4} = 0 \ \Rightarrow \ Y_{q_\mathrm{L}} = \frac{1}{2N_\mathrm{c}} \ , \\ &A_l^{444} + A_q^{444} = 0 \ \Rightarrow \ 2Y_{q_\mathrm{L}}^3 - Y_{u_\mathrm{R}}^3 - Y_{d_\mathrm{R}}^3 = -\frac{3}{4N_\mathrm{c}} \ \Rightarrow \ Y_{u_\mathrm{R}}^3 + Y_{d_\mathrm{R}}^3 = \frac{1}{4N_\mathrm{c}^3} + \frac{3}{4N_\mathrm{c}} \ . \end{aligned} \tag{5.8.3}$$

Since the quarks also couple to the gluons, there are additional triangle anomalies which are absent in the lepton sector. In particular, the range of the indices a, b, c now extends from 1, 2, 3 for $SU(2)_\mathrm{L}$ and 4 for $U(1)_Y$ to $a-4, b-4, c-4 \in \{1, 2, \ldots, N_\mathrm{c}^2-1\}$. Since there is the same number of left- and right-handed color N_c-plets, the pure QCD part of the Standard Model is a non-chiral *vector-like theory*, in which the corresponding pure $SU(N_\mathrm{c})$ anomaly cancels trivially. As a consequence, the triangle diagram with three external gluons vanishes. Triangle diagrams with a single external gluon vanish due to the tracelessness of λ^a, while those with two external gluons and one external $SU(2)_\mathrm{L}$ gauge boson vanish due to the tracelessness of σ^a. The triangle diagram with two external gluons and one external $U(1)_Y$ boson, on the other hand, is proportional to

$$A_q^{ab4} = \delta_{ab} N_\mathrm{c}(2Y_{q_\mathrm{L}} - Y_{u_\mathrm{R}} - Y_{d_\mathrm{R}}) \ , \quad a-4, b-4 \in \{1, 2, \ldots, N_\mathrm{c}\} \ . \tag{5.8.4}$$

The cancellation of this anomaly, which does not receive a contribution from the lepton sector, thus requires

$$Y_{u_\mathrm{R}} + Y_{d_\mathrm{R}} = 2Y_{q_\mathrm{L}} = \frac{1}{N_\mathrm{c}} \ . \tag{5.8.5}$$

Combined with Eq. (5.8.3) this relation implies

$$Y_{q_\mathrm{L}} = \frac{1}{2N_\mathrm{c}} \ , \ Y_{u_\mathrm{R}} = \frac{1}{2}\left(\frac{1}{N_\mathrm{c}} + 1\right) \ , \ Y_{d_\mathrm{R}} = \frac{1}{2}\left(\frac{1}{N_\mathrm{c}} - 1\right) \ , \tag{5.8.6}$$

i.e. anomaly cancellation completely fixes the weak hypercharges of the quarks. Interestingly, the resulting values are related to the generator T^3_R and the baryon number $B = 1/N_\mathrm{c}$ by

$$Y = T^3_\mathrm{R} + \frac{1}{2}B \ . \tag{5.8.7}$$

In the real world with $N_\mathrm{c} = 3$ the baryon number of a quark is $B = 1/3$ and the weak hypercharges are then given by

$$Y_{q_\mathrm{L}} = \frac{1}{6} \ , \ Y_{u_\mathrm{R}} = \frac{2}{3} \ , \ Y_{d_\mathrm{R}} = -\frac{1}{3} \ . \tag{5.8.8}$$

It is often argued that in the Standard Model the number of colors must be exactly $N_\mathrm{c} = 3$ in order to achieve anomaly cancellation. In contrast to this claim, we have just seen that the Standard Model would indeed be *consistent for any odd number N_c. cite Rudas, Abbas, Baer, Wiese.* As we will discuss in Chapter ???, there is sufficient experimental evidence to single out $N_\mathrm{c} = 3$. However, we would like to point out that the reasons for this are more subtle than it is often assumed. In particular, $N_\mathrm{c} = 3$ does *not* follow from the requirement of mathematical consistency (*i.e.* anomaly cancellation) of the Standard Model.

5.9 Electric Charges of Quarks and Baryons

In complete analogy to the lepton sector, one identifies the electric charge of the quarks as

$$Q = T^3_\mathrm{L} + Y = T^3_\mathrm{L} + T^3_\mathrm{R} + \frac{1}{2}B \ . \tag{5.9.1}$$

For the left-handed up and down quark this equation takes the form

$$Q_{u_\mathrm{L}} = T^3_{\mathrm{L}u_\mathrm{L}} + Y_{q_\mathrm{L}} = \frac{1}{2} + \frac{1}{2N_\mathrm{c}} = \frac{1}{2}\left(\frac{1}{N_\mathrm{c}} + 1\right) \ ,$$
$$Q_{d_\mathrm{L}} = T^3_{\mathrm{L}d_\mathrm{L}} + Y_{q_\mathrm{L}} = -\frac{1}{2} + \frac{1}{2N_\mathrm{c}} = \frac{1}{2}\left(\frac{1}{N_\mathrm{c}} - 1\right) \ . \tag{5.9.2}$$

For the right-handed quark fields one finds the same values of the electric charges,

$$
\begin{aligned}
Q_{u_\mathrm{R}} &= T^3_{\mathrm{L}u_\mathrm{R}} + Y_{u_\mathrm{R}} = 0 + \frac{1}{2}\left(\frac{1}{N_\mathrm{c}} + 1\right) \ , \\
Q_{d_\mathrm{R}} &= T^3_{\mathrm{L}d_\mathrm{R}} + Y_{d_\mathrm{R}} = 0 + \frac{1}{2}\left(\frac{1}{N_\mathrm{c}} - 1\right) \ .
\end{aligned} \tag{5.9.3}
$$

In the real world with $N_\mathrm{c} = 3$ the electric charges of the quarks are thus given by

$$
Q_u = \frac{2}{3} \ , \quad Q_d = -\frac{1}{3} \ . \tag{5.9.4}
$$

Since quarks have lepton number $L = 0$ and leptons have baryon number $B = 0$, the electric charges of the fermionic matter fields in the Standard Model are given by

$$
\boxed{Q = T^3_\mathrm{L} + Y = T^3_\mathrm{L} + T^3_\mathrm{R} + \frac{1}{2}(B - L) \ .} \tag{5.9.5}
$$

As we will see in Section 8.9, the difference between baryon and lepton number, $B-L$, generates an exact global symmetry of the Standard Model, while B and L individually are explicitly broken by anomalies. Once we will introduce Majorana mass terms for the neutrinos, also $B - L$ conservation will be explicitly broken.

As we will discuss in Chapter 12, just like gluons, quarks are confined inside hadrons. Hadrons containing N_c more quarks than anti-quarks are known as baryons (with baryon number $B = 1$). The most important baryons in the real world are the proton and the neutron, each containing three quarks, as well as a fluctuating number of quark–anti-quark pairs and gluons. In a constituent quark picture, the proton consists of two up quarks and one down quark, while the neutron contains one up quark and two down quarks. Indeed, the resulting electric charges,

$$
\begin{aligned}
Q_p &= 2Q_u + Q_d = 2\frac{2}{3} - \frac{1}{3} = 1 \ , \\
Q_n &= Q_u + 2Q_d = \frac{2}{3} - 2\frac{1}{3} = 0 \ ,
\end{aligned} \tag{5.9.6}
$$

are the familiar ones of proton and neutron, which are integer multiples of the charge $-e$ of an electron. Despite numerous experimental studies, including Milikan-type experiments, fundamental fractional electric charges have never been observed in Nature.[10] This is a consequence of quark confinement combined with anomaly cancellation.

In a hypothetical, but mathematically fully consistent world with an arbitrary odd number of colors N_c, there would still be protons and neutrons. However, as we will discuss in more detail in Chapter 12, a proton would then contain $(N_\mathrm{c}+1)/2$ up quarks and $(N_\mathrm{c}-1)/2$ down quarks, while a neutron would contain $(N_\mathrm{c}-1)/2$ up quarks and $(N_\mathrm{c}+1)/2$ down quarks. Hence, just as in the real world, we would still obtain

$$\begin{aligned} Q_p &= \frac{N_\mathrm{c}+1}{2} Q_u + \frac{N_\mathrm{c}-1}{2} Q_d = \frac{N_\mathrm{c}+1}{4}\left(\frac{1}{N_\mathrm{c}}+1\right) + \frac{N_\mathrm{c}-1}{4}\left(\frac{1}{N_\mathrm{c}}-1\right) = 1 , \\ Q_n &= \frac{N_\mathrm{c}-1}{2} Q_u + \frac{N_\mathrm{c}+1}{2} Q_d = \frac{N_\mathrm{c}-1}{4}\left(\frac{1}{N_\mathrm{c}}+1\right) + \frac{N_\mathrm{c}+1}{4}\left(\frac{1}{N_\mathrm{c}}-1\right) = 0 . \end{aligned} \tag{5.9.7}$$

Consequently, confinement combined with anomaly cancellation is responsible for charge quantization in integer units even for an arbitrary odd number N_c of colors.[11]

5.10 Anomaly Matching

Gerard 't Hooft has argued that anomaly cancellation should take place even if one considers only the *low-energy limit* of a given theory. Anomalies must therefore be cancelled properly also in a low-energy effective theory for the Standard Model. This *anomaly matching condition* puts non-trivial constraints on the possible dynamics of such effective theories. For example, at low energies quarks are confined inside protons and neutrons, also known as *nucleons*, and so the anomalies should also cancel between leptons and nucleons. To convince ourselves that this is indeed the case, let us reconsider

[10] Fractional charges carried by Laughlin quasi-particles emerge as a collective phenomenon in the condensed matter physics of the fractional quantum Hall effect.

[11] When the number of fermion generations were even, N_c could as well be even. In that case, baryons would be bosons with half-integer electric charges. This would change the physics drastically.

the first generation now expressed in terms of nucleon (rather than quark) degrees of freedom,

$$\begin{pmatrix} \nu_{\rm L} \\ e_{\rm L} \end{pmatrix} , \ e_{\rm R} \ , \quad \begin{pmatrix} p_{\rm L} \\ n_{\rm L} \end{pmatrix} , \ p_{\rm R} \ , \ n_{\rm R} \ . \tag{5.10.1}$$

Indeed, the global gauge anomaly is still cancelled because the left-handed nucleons form one $SU(2)_{\rm L}$ doublet. The weak hypercharge assignments for the nucleons are

$$Y_{N_{\rm L}} = \frac{1}{2} \ , \ Y_{p_{\rm R}} = 1 \ , \ Y_{n_{\rm R}} = 0 \ . \tag{5.10.2}$$

Here the index N refers to nucleons. The electric charges of the left-handed nucleons then result as

$$\begin{aligned} Q_{p_{\rm L}} &= T^3_{{\rm L}p_{\rm L}} + Y_{N_{\rm L}} = \frac{1}{2} + \frac{1}{2} = 1 \ , \\ Q_{n_{\rm L}} &= T^3_{{\rm L}n_{\rm L}} + Y_{N_{\rm L}} = -\frac{1}{2} + \frac{1}{2} = 0 \ . \end{aligned} \tag{5.10.3}$$

Similarly, for the right-handed proton and neutron we obtain the same values

$$\begin{aligned} Q_{p_{\rm R}} &= T^3_{{\rm L}p_{\rm R}} + Y_{p_{\rm R}} = 0 + 1 \ , \\ Q_{n_{\rm R}} &= T^3_{{\rm L}n_{\rm R}} + Y_{n_{\rm R}} = 0 + 0 \ . \end{aligned} \tag{5.10.4}$$

The corresponding contributions to the $SU(2)_{\rm L} \times U(1)_Y$ triangle anomalies in the nucleon sector are then given by

$$\begin{aligned} A_N^{444} &= 2Y^3_{N_{\rm L}} - Y^3_{p_{\rm R}} - Y^3_{n_{\rm R}} = 2\left(\frac{1}{2}\right)^3 - 1^3 - 0^3 = -\frac{3}{4} \ , \\ A_N^{ab4} &= \delta_{ab} 2Y_{N_{\rm L}} = \delta_{ab} \ , \end{aligned} \tag{5.10.5}$$

which again cancels the anomalies A_l^{444} and A_l^{ab4} of the leptons.

In view of our analysis for a general odd number $N_{\rm c}$, this nucleon consideration corresponds exactly to the case $N_{\rm c} = 1$, provided we identify the proton with the up "quark" and the neutron with the down "quark". In this respect, the discussion of a possible generalization is not just academic. In fact, as early as 1949 Jack Steinberger was first to calculate a triangle diagram with nucleons propagating around the loop. It is sometimes stated that he accidentally got the right answer although he neglected the quark content of protons and neutrons, and thus the color factor N_c. Of course, in 1949 Steinberger did not know about quarks or color, but he was still using a consistent low-energy description of our world. Indeed, thanks to anomaly matching, Steinberger's result is the correct answer irrespective of the value of N_c.

5.11 Right-handed Neutrinos

The minimal version of the Standard Model, which we have presented until now, does not contain right-handed neutrino fields. If one insists on perturbative renormalizability, the absence of right-handed neutrino fields implies that neutrinos are massless, which may thus be viewed as a prediction of the Standard Model. However, since the observation of neutrino oscillations in 1998, it is known that (at least some) neutrinos must have mass.[12] One may then conclude that the minimal Standard Model is indeed in conflict with experiment and must thus be extended. One may do this in two alternative ways. First, one may view the Standard Model as an effective theory, formulated only in terms of the relevant low-energy degrees of freedom. The leading terms in the effective Lagrangian are indeed the renormalizable interactions that we have considered until now. However, in an effective theory framework there are additional higher-order corrections to the effective Lagrangian which need not be renormalizable. As we will discuss in more detail in Chapter 9, one can indeed construct non-renormalizable neutrino mass terms by using just the left-handed neutrino fields introduced until now.

An alternative way to proceed, which reflects a drastically different point of view, is to assume that the Standard Model is an integral part of a renormalizable theory with a larger field content that extends to much higher energies beyond the TeV range. This approach is pursued, for example, in the framework of grand unified theories (GUT), which will be discussed in detail in Chapter 18. If one insists on perturbative renormalizability, the incorporation of neutrino mass terms requires the introduction of right-handed

[12] The values of the neutrino masses are presently not known experimentally. Neutrino oscillations only imply non-zero neutrino mass differences.

neutrino fields ν_R and $\bar{\nu}_\mathrm{R}$. As we will now discuss, the Standard Model can be extended by right-handed neutrinos in a straightforward manner.

Right-handed neutrinos are leptons (with lepton number $L = 1$), *i.e.* under global $U(1)_L$ transformations they transform as

$$\nu_\mathrm{R}'(x) = \exp(\mathrm{i}\chi)\nu_\mathrm{R}(x) \ , \quad \bar{\nu}_\mathrm{R}'(x) = \bar{\nu}_\mathrm{R}(x)\exp(-\mathrm{i}\chi) \ . \qquad (5.11.1)$$

Just like right-handed electrons, right-handed neutrinos are both $SU(N_\mathrm{c})$ color and $SU(2)_\mathrm{L}$ singlets, and one has

$$T^3_{\mathrm{L}\nu_\mathrm{R}} = 0 \ , \quad T^3_{\mathrm{R}\nu_\mathrm{R}} = \frac{1}{2} \ . \qquad (5.11.2)$$

Using $Y = T^3_\mathrm{R} - L/2$ and $Q = T^3_\mathrm{L} + Y$, we then obtain

$$Y_{\nu_\mathrm{R}} = T^3_{\mathrm{R}\nu_\mathrm{R}} - \frac{1}{2} = 0 \ , \quad Q_{\nu_\mathrm{R}} = T^3_{\mathrm{L}\nu_\mathrm{R}} + Y_{\nu_\mathrm{R}} = 0 \ . \qquad (5.11.3)$$

This implies that the right-handed neutrino is neutral, not only electrically, but under all gauge interaction in the Standard Model. Consequently, right-handed neutrinos are *"sterile"*, *i.e.* they do not participate in the electromagnetic, weak, or strong interaction. Since right-handed neutrinos do not couple to the gauge fields of the Standard Model, they do not contribute to the gauge anomalies. Hence, these anomalies remain properly cancelled.

As we will see in Chapter 9, left- and right-handed neutrino fields can be combined in a Yukawa coupling term to the Higgs field. When the Higgs field picks up a vacuum expectation value v, this term gives rise to a non-zero neutrino mass proportional to v. We will also see that right-handed neutrino fields alone can be used to form additional Majorana mass terms, which are not tied to the electroweak symmetry breaking scale v. In fact, besides v, the Majorana mass M will appear as a second dimensionful parameter in this extended Standard Model.

5.12 Lepton and Baryon Number Anomalies

As we discussed before, the lepton-gauge field Lagrangian $\mathcal{L}(\bar{\nu}, \nu, \bar{e}, e, W_\mu, B_\mu)$ of Eq. (5.3.13) as well as the quark-gauge field Lagrangian $\mathcal{L}(\bar{u}, u, \bar{d}, d, G_\mu, W_\mu, B_\mu)$

of Eq. (5.7.12) are invariant against global $U(1)_L$ lepton number and $U(1)_B$ baryon number transformations. Hence, at the classical level, lepton and baryon number are conserved quantities. As we discussed in the Preface, usually global symmetries are only approximate, while exact symmetries are local. Would it be possible to gauge the $U(1)_L$ and $U(1)_B$ symmetries in the Standard Model? As we will see, this is not possible, because both symmetries are explicitly broken by anomalies, and are thus indeed only approximate. Still, we will find that the combination $B - L$ is conserved exactly.

Let us imagine that there is a hypothetical $U(1)_L$ gauge boson that couples to lepton number. Such a particle would mediate a *fifth force*, beyond gravity, electromagnetism, as well as the weak and strong interactions. There is no experimental evidence for such a force, and we will now see that gauging $U(1)_L$ is, in fact, impossible in the Standard Model because this symmetry suffers from triangle anomalies. After the introduction of right-handed neutrinos, $U(1)_L$ is a vector-like symmetry, *i.e.* both left- and right-handed leptons carry the same lepton number $L = 1$. As a consequence, the pure $U(1)_L$ triangle anomaly with three external hypothetical $U(1)_L$ gauge boson vanishes trivially. Still, there may be mixed anomalies. First of all, triangle diagrams containing external gluons vanish because leptons do not participate in the strong interaction. Triangle diagrams with two external $U(1)_L$ and one external $SU(2)_{\mathrm{L}}$ gauge boson vanish due the tracelessness of σ^a. The mixed anomaly with two $U(1)_L$ and one $U(1)_Y$ gauge boson is proportional to

$$\begin{aligned} A^{4LL} &= 2\left[\mathrm{Tr}_{\mathrm{L}} L^2 Y - \mathrm{Tr}_{\mathrm{R}} L^2 Y\right] = 2\left[2Y_{l_{\mathrm{L}}} - Y_{\nu_{\mathrm{R}}} - Y_{e_{\mathrm{R}}}\right] \\ &= 2\left[2\left(-\frac{1}{2}\right) - 0 - (-1)\right] = 0 \ , \end{aligned} \tag{5.12.1}$$

and thus vanishes. We still need to consider the triangle diagrams with just one external hypothetical $U(1)_L$ gauge boson. The diagram with two external $U(1)_Y$ gauge bosons contributes

$$\begin{aligned} A^{44L} &= 2\left[\mathrm{Tr}_{\mathrm{L}} Y^2 L - \mathrm{Tr}_{\mathrm{R}} Y^2 L\right] = 2\left[2Y_{l_{\mathrm{L}}}^2 - Y_{\nu_{\mathrm{R}}}^2 - Y_{e_{\mathrm{R}}}^2\right] \\ &= 2\left[2\left(-\frac{1}{2}\right)^2 - 0^2 - (-1)^2\right] = -1 \neq 0 \ , \end{aligned} \tag{5.12.2}$$

and thus leads to an inconsistency when the $U(1)_L$ lepton number symmetry is gauged. The diagram with one external $U(1)_Y$ and one external $SU(2)_{\mathrm{L}}$

gauge boson vanishes due to the tracelessness of σ^a. On the other hand, the triangle diagram with two external $SU(2)_{\rm L}$ gauge bosons (with $a, b \in \{1, 2, 3\}$) contributes

$$\begin{aligned} A^{abL} &= \mathrm{Tr}_{\rm L}\left[(T^aT^b + T^bT^a)L\right] - \mathrm{Tr}_{\rm R}\left[(T^aT^b + T^bT^a)L\right] \\ &= \mathrm{Tr}_{\rm L}\left[\frac{1}{4}(\sigma^a\sigma^b + \sigma^b\sigma^a)L\right] = \delta_{ab}\mathrm{Tr}_{\rm L}L = 2\delta_{ab} \neq 0 \ , \end{aligned} \tag{5.12.3}$$

and thus gives rise to another anomaly.

Let us now investigate potential anomalies in the $U(1)_B$ baryon number symmetry. In that case, only quarks propagate around the triangle diagrams. In complete analogy to the lepton case, one may convince oneself that the only non-vanishing anomalies are

$$\begin{aligned} A^{44B} &= 2\left[\mathrm{Tr}_{\rm L}Y^2B - \mathrm{Tr}_{\rm R}Y^2B\right] = 2N_{\rm c}\left[2Y_{q_{\rm L}}^2 - Y_{u_{\rm R}}^2 - Y_{d_{\rm R}}^2\right]\frac{1}{N_{\rm c}} \\ &= 2\left[2\left(\frac{1}{2N_{\rm c}}\right)^2 - \frac{1}{4}\left(\frac{1}{N_{\rm c}} + 1\right)^2 - \frac{1}{4}\left(\frac{1}{N_{\rm c}} - 1\right)^2\right] = -1 \\ A^{abB} &= \mathrm{Tr}_{\rm L}\left[(T^aT^b + T^bT^a)B\right] - \mathrm{Tr}_{\rm R}\left[(T^aT^b + T^bT^a)B\right] \\ &= \mathrm{Tr}_{\rm L}\left[\frac{1}{4}(\sigma^a\sigma^b + \sigma^b\sigma^a)B\right] = \delta_{ab}\mathrm{Tr}_{\rm L}B = 2N_{\rm c}\delta_{ab}\frac{1}{N_{\rm c}} = 2\delta_{ab} \neq 0 \ . \end{aligned} \tag{5.12.4}$$

Remarkably, for any number of colors,

$$A^{44B} = A^{44L} \ , \ A^{abB} = A^{abL} \ , \ a, b \in \{1, 2, 3\} \ , \tag{5.12.5}$$

such that the anomalies cancel in the combination $B-L$. Hence, although B and L are individually broken at the quantum level, the difference between baryon and lepton number is an exactly conserved quantum number in the gauge interactions of the Standard Model. This raises the question why the corresponding $U(1)_{B-L}$ symmetry is not gauged. Indeed, there are GUT extensions of the Standard Model with an $SO(10)$ gauge group which contains the $U(1)_{B-L}$ subgroup as a local symmetry. Alternatively, when $U(1)_{B-L}$ remains a global symmetry, Majorana mass terms involving the right-handed neutrino field $\nu_{\rm R}$ explicitly break L even at the classical level, and thus turn $U(1)_{B-L}$ into an approximate symmetry. Similarly, if one views the Standard Model as a low-energy effective theory, $U(1)_{B-L}$ is an *accidental* global symmetry which will be violated by non-renormalizable higher-order corrections to the Lagrangian.

5.13 An Anomaly-Free Technicolor Model

This section addresses physics beyond the Standard Model and can be skipped at a first reading.

Why is the electroweak scale $v = 246\,\mathrm{GeV}$ so much smaller than the Planck scale $M_{\mathrm{Planck}} \approx 10^{19}\mathrm{GeV}$? This is the hierarchy problem that we have discussed in Section 5.6. A possible solution of this problem is based on the idea of *"techni-color"* — a hypothetical gauge interaction even stronger than the strong force — which confines new fundamental fermions — the so-called *"techni-quarks"* — to form the Higgs particle as a composite object. This is analogous to the binding of electrons that form Cooper pairs in a superconductor. In that case, the condensation of Cooper pairs leads to the spontaneous breaking of $U(1)_{\mathrm{em}}$. Similarly, in techni-color models the condensation of techni-quark-techni-anti-quark pairs leads to the spontaneous breaking of $SU(2)_{\mathrm{L}} \times U(1)_Y$ down to $U(1)_{\mathrm{em}}$. Thanks to the property of *asymptotic freedom*, which techni-color models share with QCD, one can explain the large hierarchy between v and M_{Planck} in a natural manner, *i.e.* without fine-tuning any parameters. In fact, techni-color models mimic the dynamics of QCD at the electroweak scale. Since we will discuss the QCD dynamics only in Chapter 11 and 12, we will postpone the discussion of the techni-color dynamics until Chapter 14. However, in this section we already introduce the basic ingredients of a minimal techni-color extension of the Standard Model, and we show that the extended model is still anomaly-free. It should be mentioned that, at present, there is no experimental evidence supporting the idea of techni-color models. Instead, in these models there are severe problems due to flavor-changing neutral currents. Hence, it remains to seen whether techni-color is the right way to go beyond the Standard Model.

Let us construct a concrete techni-color model, in particular, to show explicitly that such constructions are at all possible. In addition to the Standard Model fermions, we want to add a techni-up and a techni-down quark U and D whose left-handed components form an $SU(2)_{\mathrm{L}}$ doublet and whose right-handed components are $SU(2)_{\mathrm{L}}$ singlets. The techni-color gauge group is chosen to be $SU(N_{\mathrm{t}})$ and both the left- and the right-handed techni-quarks transform in the fundamental representation of $SU(N_{\mathrm{t}})$. All the Standard Model fermions are assumed to be techni-color singlets. We will not introduce any techni-leptons. Hence, anomaly cancellation works differently than in the Standard Model. For simplicity, we choose the techni-quarks to be $SU(N_c)$ color singlets. However, they are still confined by techni-color interactions. Let us denote the $U(1)_Y$ quantum numbers of the techni-quarks by Y_{Q_L}, Y_{U_R}, and Y_{D_R}. These parameters will be determined by anomaly cancellation conditions.

The gauge group of our techni-color model is given by $SU(N_t)\times SU(N_c)\times SU(2)_L \times U(1)_Y$. Let us now demand anomaly cancellation. Since, like $SU(N_c)$ color, the techni-color gauge group $SU(N_t)$ is a vector-like symmetry, the triangle diagram with three external techni-gauge bosons automatically vanishes. Triangle diagrams with external techni-gauge bosons and external gluons only also vanish. Triangle diagrams with only one external techni-gauge boson vanish because the generators of $SU(N_t)$ are traceless. The triangle diagram with two techni-gauge bosons and one $SU(2)_L$ boson vanishes due to the tracelessness of σ^a. The triangle diagram with two external techni-gauge bosons and one external $U(1)_Y$ boson vanishes only if

$$2Y_{Q_L} = Y_{U_R} + Y_{D_R} \ . \tag{5.13.1}$$

The techni-quarks also contribute to the anomalies of the Standard Model gauge symmetries. For example, the triangle diagram with two $SU(2)_L$ bosons and one $U(1)_Y$ boson still vanishes only if

$$Y_{Q_L} = 0 \ , \tag{5.13.2}$$

while the diagram with three external $U(1)_Y$ bosons vanishes only if

$$2Y_{Q_L}^3 = Y_{U_R}^3 + Y_{D_R}^3 \ . \tag{5.13.3}$$

Anomaly cancellation hence implies

$$Y_{Q_L} = 0 \ , \ Y_{U_R} + Y_{D_R} = 0 \ . \tag{5.13.4}$$

We still want to be able to couple our new theory to gravity, which is possible only if we cancel the gravitational anomaly. This again requires

$$2Y_{Q_L} = Y_{U_R} + Y_{D_R} \ , \tag{5.13.5}$$

which is hence already satisfied.

In order to reproduce the physics of the Standard Model, we must maintain $U(1)_{\rm em}$ as an unbroken gauge symmetry. This requires the electric charges of the left- and right-handed techni-quarks to be equal.[13] Since we have $Q = T^3_{\rm L} + Y$, we obtain

$$Q_{U_{\rm L}} = \frac{1}{2}\ , \ Q_{D_{\rm L}} = -\frac{1}{2}\ , \ Q_{U_{\rm R}} = Y_{U_{\rm R}}\ , \ Q_{D_{\rm R}} = Y_{D_{\rm R}}\ . \tag{5.13.6}$$

Hence, in order to have equal charges for left- and right-handed techni-quarks we demand

$$Y_{U_{\rm R}} = -Y_{D_{\rm R}} = \frac{1}{2}\ . \tag{5.13.7}$$

In order to also cancel Witten's global gauge anomaly, the total number of $SU(2)_{\rm L}$ doublets must be even and hence $N_{\rm t}$ must also be even. The naive simplest choice $N_{\rm t} = 2$ is not analogous to QCD. Due to the pseudo-real nature of $SU(2)$, techni-quarks and techni-anti-quarks would then be indistinguishable and the chiral symmetry would be $Sp(4)$ instead of $SU(2)_{\rm L} \times SU(2)_{\rm R}$. The actual simplest choice therefore is $N_{\rm t} = 4$. The gauge symmetry of the Standard Model extended by our simple version of techni-color then is $SU(4)_{\rm t} \times SU(3)_{\rm c} \times SU(2)_{\rm L} \times U(1)_Y$.

[13] As we will discuss in Chapter 14, if the left- and right-handed techni-quarks have the same electric charges, the breaking of the techni-chiral $SU(2)_{\rm L} \times SU(2)_{\rm R}$ symmetry leaves $U(1)_{\rm em}$ intact. Otherwise, the techni-chiral condensate $\langle \bar U_{\rm L} U_{\rm R} + \bar U_{\rm L} U_{\rm R} + \bar D_{\rm L} D_{\rm R} + \bar D_{\rm L} D_{\rm R}\rangle$ would carry an electric charge and would turn the vacuum into a superconductor.

Chapter 6

Fermion Masses

At this point, we have introduced all fields in the Standard Model with one generation of fermions. Gauge invariance and anomaly cancellation have led to severe limitations on the terms that can enter the Lagrangian. Altogether, until now, we have introduced five adjustable fundamental parameters: only one dimensionful parameter — the vacuum expectation value v of the Higgs field — as well as the dimensionless Higgs self-coupling λ and the three dimensionless gauge couplings g, g' (or alternatively e and θ_{W}), and g_{s}. In addition, we have made several choices for the fermion representations. For example, in the way we have presented the Standard Model, one may consider the number of colors N_{c} another (integer-valued) parameter to be determined by experiment. In any case, the number of parameters is still moderate at this stage.

Usually, it is emphasized that the Standard Model describes the electromagnetic, weak, and strong interactions, and that there are thus four fundamental forces, including gravity. In this chapter, we will see that the Standard Model also contains so-called *Yukawa interactions* between the Higgs field and the fermions, whose strengths are controlled by a large number of additional adjustable parameters. When the Higgs field picks up its vacuum expectation value v, the Yukawa interactions lead to fermion masses as well as to mixing between different fermion generations. Since they are not mediated by gauge fields, the Yukawa interactions are not very restricted and thus lead to a proliferation of adjustable parameters in the Standard Model. Even with only one generation of fermions, we will now have the Dirac masses of the up and down quarks, and of the electron and the neutrino, as well as an additional Majorana mass for the neutrino. In the next chapter, we will add two more fermion generations which will increase the number of parameters much further. While it is possible to determine the values of the Standard Model parameters by comparison with experiment, their theoretical understanding is a very challenging unsolved problem beyond the Standard Model.

One may expect that the large hadron collider (LHC) at CERN will shed light on the Higgs phenomenon and thus on the dynamical mechanism responsible for electroweak symmetry breaking. It is possible that an extended version of the Standard Model will replace the fundamental Higgs field and the many parameters associated with its Yukawa interactions by a more fundamental and more constrained dynamics, perhaps driven by yet unknown gauge forces. However, other ideas beyond the Standard Model, *e.g.* those relying on supersymmetry — a symmetry that pairs fermions with bosons — could still increase the number of adjustable parameters even further.

6.1 Electron and Down Quark Masses

So far we have not introduced any mass terms for the fermions. An electron mass term would have the form $m_e[\bar{e}_R e_L + \bar{e}_L e_R]$. As we mentioned before, since left- and right-handed fermions transform differently under $SU(2)_\mathrm{L}$ and $U(1)_Y$ gauge transformations, this term violates the chiral gauge symmetry and is thus forbidden. We remember that we encountered a similar situation before for the weak gauge bosons: we know experimentally that they are massive, but explicit mass terms for them are forbidden by gauge invariance. The way out was the Higgs mechanism. By picking up a vacuum expectation value, the Higgs field Φ gave mass to the gauge bosons via spontaneous symmetry breaking. Similarly, Φ can also give mass to fermions. Let us write down a Yukawa interaction term with the *Yukawa* *coupling* f_e[1]

$$\begin{aligned}\mathcal{L}(\bar{\nu},\nu,\bar{e},e,\Phi) &= f_e\ (\bar{\nu}_\mathrm{L},\bar{e}_\mathrm{L})\begin{pmatrix}\Phi^+\\ \Phi^0\end{pmatrix} e_\mathrm{R} + f_e^*\ \bar{e}_\mathrm{R}(\Phi^{+*},\Phi^{0*})\begin{pmatrix}\nu_\mathrm{L}\\ e_\mathrm{L}\end{pmatrix}\\ &= f_e\ \bar{l}_\mathrm{L}\Phi e_\mathrm{R} + f_e^*\ \bar{e}_\mathrm{R}\Phi^\dagger l_\mathrm{L}\ . \end{aligned} \tag{6.1.1}$$

The second term is multiplied by f_e^*, in order to ensure Hermiticity of the corresponding Hamiltonian.[2] The above Lagrangian is $SU(2)_\mathrm{L}$ gauge invariant because both the left-handed leptons and the Higgs are $SU(2)_\mathrm{L}$ doublets, while the right-handed electron is an $SU(2)_\mathrm{L}$ singlet. Moreover, the Lagrangian is also $U(1)_Y$ invariant. To see this, we sum up the hypercharges of the fields in the first term in Eq. (6.1.1),

$$-Y_{l_\mathrm{L}} + Y_\Phi + Y_{e_\mathrm{R}} = \frac{1}{2} + \frac{1}{2} - 1 = 0\ . \tag{6.1.2}$$

Since the hypercharges add up to zero, the corresponding $U(1)_Y$ gauge transformations $\exp(\mathrm{i}Y\varphi(x))$ cancel each other, and the term is thus $U(1)_Y$ gauge invariant. In the second term, the signs of all hypercharges are flipped, and hence its total hypercharge vanishes as well.)

Since charge conjugation as well as parity turn left- into right-handed neutrinos, and since there are no right-handed neutrino fields in the Lagrangian of Eq. (6.1.1), it is clear that it explicitly breaks P and C. Let us now perform a combined CP transformation in the corresponding action, *i.e.*

$$\begin{aligned}&S[{}^{\mathrm{CP}}\bar{l}_\mathrm{L},{}^{\mathrm{CP}}l_\mathrm{L},{}^{\mathrm{CP}}\bar{e}_\mathrm{R},{}^{\mathrm{CP}}e_\mathrm{R},{}^{\mathrm{CP}}\Phi]\\ &= \int d^4x\ \big[-f_e\ l_\mathrm{L}(-\vec{x},x_4)^\mathsf{T} P^\mathsf{T} C^{-1}\Phi(-\vec{x},x_4)^* C P \bar{e}_\mathrm{R}(-\vec{x},x_4)^\mathsf{T}\\ &-f_e^*\ e_\mathrm{R}(-\vec{x},x_4)^\mathsf{T} P^\mathsf{T} C^{-1}\Phi(-\vec{x},x_4)^\mathsf{T} C P \bar{l}_\mathrm{L}(-\vec{x},x_4)^\mathsf{T}\big]\\ &= \int d^4x\ \big[f_e\ \bar{e}_\mathrm{R}(-\vec{x},x_4)\Phi(-\vec{x},x_4)^\dagger \bar{l}_\mathrm{L}(-\vec{x},x_4)\\ &+f_e^*\ \bar{l}_\mathrm{L}(-\vec{x},x_4)\Phi(-\vec{x},x_4)e_\mathrm{R}(-\vec{x},x_4)\big]\ . \end{aligned} \tag{6.1.3}$$

[1]Generally, a Yukawa interaction term has the structure $\bar{\psi}\phi\psi$, where ϕ is a scalar field and $\bar{\psi}$ and ψ are fermion fields.

[2]Note that choosing a Yukawa coupling like f_e to be complex does not yield any problem for the convergence of the path integral. This is in contrast to the scalar self-coupling λ, which must be non-negative. The integrals $\int D\bar{l}_\mathrm{L} Dl_\mathrm{L} D\bar{e}_\mathrm{R} De_\mathrm{R}$ converge in any case because they are Grassmannian.

Hence, it seems that the action is CP-invariant only if the Yukawa coupling f_e is real. However, as we will now discuss, f_e can always be made real by a field redefinition. Let us assume that $f_e = |f_e| \exp(i\theta)$. One can then redefine

$$e'_\mathrm{R}(x) = e_\mathrm{R}(x) \exp(i\theta) \ , \ \bar{e}'_\mathrm{R}(x) = \bar{e}_\mathrm{R}(x) \exp(-i\theta) \ , \tag{6.1.4}$$

which absorbs the complex phase $\exp(i\theta)$ into the right-handed electron field. Expressed in terms of the redefined fields, the Lagrangian then contains the real-valued Yukawa coupling $|f_e|$. It is important to note that the field redefinition leaves the gauge-fermion terms of the Lagrangian invariant. As we will discuss in Chapter ???, such field redefinitions may have subtle effects on the fermionic measure. In any case, from now on we may assume that f_e is real.

Inserting again the vacuum configuration of the Higgs field that we selected before, we obtain

$$\begin{aligned} \mathcal{L}(\bar{\nu}, \nu, \bar{e}, e, \Phi) &= f_e \left[(\bar{\nu}_\mathrm{L}, \bar{e}_\mathrm{L}) \begin{pmatrix} 0 \\ v \end{pmatrix} e_\mathrm{R} + \bar{e}_\mathrm{R}(0, v) \begin{pmatrix} \nu_\mathrm{L} \\ e_\mathrm{L} \end{pmatrix} \right] \\ &= f_e v \left[\bar{e}_\mathrm{L} e_\mathrm{R} + \bar{e}_\mathrm{R} e_\mathrm{L} \right] . \end{aligned} \tag{6.1.5}$$

Indeed, we have arrived at mass term for the electron with the mass given by

$$\boxed{m_e = f_e v} \ , \tag{6.1.6}$$

while the neutrino remains massless. Via the Yukawa coupling f_e, we have just introduced another free parameter into the theory which determines the electron mass. The Standard Model itself does not make any predictions about this parameter. If we want to understand the value of the electron mass, we need to go beyond the Standard Model. In fact, at present nobody understands why the electron has its particular mass of 0.511 MeV. As we continue to add mass terms, the number of adjustable parameters in the Standard Model will increase rapidly.

We see that the Standard Model contains more than just electroweak and strong interactions. Every Yukawa coupling parameterises a *fundamental force* that is not often emphasised on the same level as the gauge forces. There is reason to believe that the Yukawa couplings are not as fundamental as the gauge interactions. For example, in a future theory beyond the Standard Model the Yukawa couplings may ultimately be replaced by some gauge force of a new kind. In this way, we would perhaps gain predictive power and finally understand the value of the electron mass. This underscores that the true origin of mass is not at all well understood. The often celebrated Higgs mechanism leaves many fundamental questions unanswered.

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

$$-Y_{q_\mathrm{L}} + Y_\Phi + Y_{d_\mathrm{R}} = \frac{1}{2N_\mathrm{c}} + \frac{1}{2} + \frac{1}{2}\left(\frac{1}{N_\mathrm{c}} - 1\right) = 0 \ , \tag{6.1.7}$$

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

$$\mathcal{L}(\bar{u}, u, \bar{d}, d, \Phi) = f_d \left[(\bar{u}_\mathrm{L}, \bar{d}_\mathrm{L}) \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} d_\mathrm{R} + \bar{d}_\mathrm{R}(\Phi^{+*}, \Phi^{0*}) \begin{pmatrix} u_\mathrm{L} \\ d_\mathrm{L} \end{pmatrix} \right] . \tag{6.1.8}$$

to the Standard Model Lagrangian. On the other hand, we cannot give mass to the up quark in the same way, just as we did not obtain a massive neutrino.[3]

6.2 Up Quark Mass

We could easily construct a mass term for the up quark if we had another Higgs field

$$\widetilde{\Phi}(x) = \begin{pmatrix} \widetilde{\Phi}^0(x) \\ \widetilde{\Phi}^-(x) \end{pmatrix} , \tag{6.2.1}$$

which would be an $SU(2)_\mathrm{L}$ doublet that takes a vacuum value

$$\widetilde{\Phi}(x) = \begin{pmatrix} \widetilde{v} \\ 0 \end{pmatrix} . \tag{6.2.2}$$

[3]For some time, it was not clear whether the up quark might be massless, which is by now excluded experimentally. However, even if we were ready to accept $m_u = 0$ this wouldn't really help. In the next chapter, we will add two generations of heavier fermions, and the charm and top quarks — which take the position of the up quark in the second and third generation — clearly have a non-zero mass.

Then we could just add another Yukawa term

$$\begin{aligned}\mathcal{L}(\bar{u}, u, \bar{d}, d, \widetilde{\Phi}) &= f_u \left[(\bar{u}_{\rm L}, \bar{d}_{\rm L}) \begin{pmatrix} \widetilde{\Phi}^0 \\ \widetilde{\Phi}^- \end{pmatrix} u_{\rm R} + \bar{u}_{\rm R} (\Phi^{0*}, \Phi^{-*}) \begin{pmatrix} u_{\rm L} \\ d_{\rm L} \end{pmatrix} \right] \\ &= f_u \left[\bar{q}_{\rm L} \tilde{\Phi} u_{\rm R} + \bar{u}_{\rm R} \widetilde{\Phi} q_{\rm L} \right] . \end{aligned} \tag{6.2.3}$$

To render this term gauge invariant, the weak hypercharge of the field $\widetilde{\Phi}$ must obey

$$-Y_{q_{\rm L}} + Y_{\widetilde{\Phi}} + Y_{u_{\rm R}} = -\frac{1}{2N_{\rm c}} + Y_{\widetilde{\Phi}} + \frac{1}{2} \left(\frac{1}{N_{\rm c}} - 1 \right) = 0 \ \Rightarrow \ Y_{\widetilde{\Phi}} = -\frac{1}{2} . \tag{6.2.4}$$

At this point, we could just add a new Higgs field $\widetilde{\Phi}$ with the desired features. In fact, as we have discuused in Section 5.9, this is exactly what Peccei and Quinn have proposed in order to solve the strong CP problem, which we will address in Chapter 17.

However, here we restrict ourselves to the Standard Model, which does not proceed in this manner: in fact, there is more economic way to proceed by "recycling" the Higgs field introduced previously. It may come as a surprise that a field $\widetilde{\Phi}$ with the desired properties can be constructed directly from the known Higgs field Φ as

$$\widetilde{\Phi}(x) = \begin{pmatrix} \widetilde{\Phi}^0(x) \\ \widetilde{\Phi}^-(x) \end{pmatrix} = \begin{pmatrix} -\Phi^{0*}(x) \\ \Phi^{+*}(x) \end{pmatrix} . \tag{6.2.5}$$

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

$$\mathbf{\Phi}(x) = \begin{pmatrix} \Phi^{0*}(x) & \Phi^+(x) \\ -\Phi^{+*}(x) & \Phi^0(x) \end{pmatrix} . \tag{6.2.6}$$

As we have seen in Section 5.2, under $SU(2)_{\rm L}$ gauge transformations $L(x)$ the matrix field transforms as $\mathbf{\Phi}'(x) = L(x)\mathbf{\Phi}(x)$. Since the field $\widetilde{\Phi}$ is nothing but the first column vector of the matrix $\mathbf{\Phi}$, it is clear that it transforms indeed as an $SU(2)_{\rm L}$ doublet. Using the matrix field $\mathbf{\Phi}$, the quark Yukawa coupling terms can be written as

$$\mathcal{L}(\bar{u}, u, \bar{d}, d, \mathbf{\Phi}) = (\bar{u}_{\rm L}, \bar{d}_{\rm L}) \mathbf{\Phi} \mathcal{F} \begin{pmatrix} u_{\rm R} \\ d_{\rm R} \end{pmatrix} + (\bar{u}_{\rm R}, \bar{d}_{\rm R}) \mathcal{F}^\dagger \mathbf{\Phi}^\dagger \begin{pmatrix} u_{\rm L} \\ d_{\rm L} \end{pmatrix} , \tag{6.2.7}$$

where the Yukawa couplings are contained in the diagonal matrix

$$\mathcal{F} = \begin{pmatrix} f_u & 0 \\ 0 & f_d \end{pmatrix} . \tag{6.2.8}$$

The above construction implies $\widetilde{v} = v$ and hence the up quark mass is given by $m_u = f_u v$. Inserting the vacuum value of the Higgs field, the quark mass matrix then results as

$$\mathcal{M} = \mathbf{\Phi}\mathcal{F} = \begin{pmatrix} v & 0 \\ 0 & v \end{pmatrix} \begin{pmatrix} f_u & 0 \\ 0 & f_d \end{pmatrix} = \begin{pmatrix} m_u & 0 \\ 0 & m_d \end{pmatrix} . \tag{6.2.9}$$

6.3 Massive Neutrinos

As we will discuss in more detail in Chapter 10, in 1998 oscillations between different neutrino species were observed by the ... collaboration. This implies that (at least some) neutrinos must be massive. Since the Standard Model does not contain right-handed neutrino fields, one cannot even write down a neutrino mass term, at least as long as one restricts oneself to renormalizable interactions. As we have discussed in Section 5.7, already due to its triviality, the Standard Model is at best a low-energy effective theory, which cannot be valid at arbitrarily high energy scales. If one accepts that the Standard Model is a low-energy effective theory, there is, however, no good reason to exclude non-renormalizable interaction. Instead, those should be added as higher-order corrections to the leading renormalizable Standard Model interactions.

Once we have introduced a right-handed neutrino field, we can give mass to the neutrino in the same way as we just gave mass to the up quark. It has been argued that adding neutrino masses is already "beyond the Standard Model". While this is clearly a matter of semantics, we do not adapt this point of view. First, it is an addition to the former version of the Standard Model, which does not involve a conceptual extension. Second, in some sense the Standard Model without right-handed neutrinos has always looked unnatural, because with massless neutrinos it has an exact global symmetry. As we have claimed already in Chapter 1, exact symmetries should be locally realised, or alternatively, global symmetries should be only approximate.

There is an internal quantum number called *lepton number* L which tends to be conserved in Nature. More precisely, it is the difference $B - L$ which would be conserved exactly, without the existence of right-handed neutrinos. B is the *baryon number*, and the assignments of L and B are simple:

$$L = \begin{cases} 1 & \text{leptons} \\ -1 & \text{anti} - \text{leptons} \\ 0 & \text{all other particles} \end{cases} \qquad B = \begin{cases} 1/3 & \text{quarks} \\ -1/3 & \text{anti} - \text{quarks} \\ 0 & \text{all other particles} \end{cases} \tag{6.3.1}$$

With right-handed neutrinos present, one can construct a *Majorana mass term*, which violates explicitly the conservation of L, and — since no quarks are involved in that term — also of the difference $B - L$.

6.4 The Majorana mass term

Let us start with an addition to the discussion of the Dirac equation in Section 1.7. If a fermion spinor obeys $(i\not\partial - e\not A - m)\Psi = 0$ (where A_μ may represent any gauge field), the spinor for the corresponding anti-fermion fulfils $(i\not\partial + e\not A - m)\Psi_c = 0$. For a symmetric representation of γ^0 we relate the spinors as

$$\Psi_c = C\gamma^0\Psi^* = C\bar{\Psi}^T \ , \tag{6.4.1}$$

where the matrix C has to be chosen such that

$$\gamma^\mu C\gamma^0 = -C\gamma^0\gamma^{\mu *} \ . \tag{6.4.2}$$

With this connection between the two spinors, the above Dirac equations for Ψ and Ψ_c are in fact equivalent. In both common representations, named after Pauli-Dirac and Weyl, the matrix

$$C\gamma^0 = \begin{pmatrix} & & & 1 \\ & & -1 & \\ & -1 & & \\ 1 & & & \end{pmatrix} \tag{6.4.3}$$

is a solution to the condition (6.4.2).

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

$$\nu_M = \nu_R + C\bar{\nu}_R^T \ . \tag{6.4.4}$$

The charge conjugate of a Majorana neutrino is identical with itself,

$$C\bar{\nu}_M^T = C\bar{\nu}_R^T + \nu_R = \nu_M \ , \tag{6.4.5}$$

and hence it is its own anti-particle. This is possible only for a neutral gauge singlet, i.e. a particle which is neutral with respect to all gauge fields. The Majorana condition (6.4.5) also implies that the corresponding spinor has only two degrees of freedom. [4] A Majorana mass term takes the form

$$\mathcal{L}_M(\nu_R) = M\bar{\nu}_M\nu_M \tag{6.4.6}$$

and changes the lepton number L by two. It is automatically gauge invariant. This mass term does not require the inclusion of the Higgs field. Consequently, the Majorana mass M is not tied to the electroweak scale v. In fact, M is the *second dimensional parameter* that enters the Standard Model once we introduce right-handed neutrino fields. [5]

A renormalisable Majorana mass term cannot be constructed from the left-handed neutrino field because it would not be gauge invariant. However, as described above, we can construct Dirac mass terms that couple left- and right-handed neutrino fields through a Yukawa coupling to the Higgs field. Altogether we can write a neutrino mass matrix of the form

$$(\bar{\nu}_L, C\nu_R^T)\begin{pmatrix} 0 & f_\nu v \\ f_\nu v & M \end{pmatrix}\begin{pmatrix} C\bar{\nu}_L^T \\ \nu_R \end{pmatrix} \ . \tag{6.4.7}$$

For $M \gg f_\nu v$ the eigenvalues of the mass matrix are

$$m_1 \simeq M \ , \quad m_2 \simeq \frac{f_\nu^2 v^2}{M} \ , \tag{6.4.8}$$

i.e. there is a large mass m_1 and a much smaller mass m_2. [6] When we discuss Grand Unified Theories (GUT), we will see that the assumption $f_\nu v \ll M \sim$

[4]This is in contrast to the unconstrained Dirac spinor, which has 4 degrees of freedom, describing a spin-1/2 particle and its independent anti-particle.

[5]Note that all other free parameters that we introduced in the Standard Model up to now, such as the Yukawa couplings, the Higgs self-coupling λ and the weak mixing angle θ_w are in fact dimensionless. The vacuum Higgs value v — or equivalently the Higgs mass — contributed the dimension for all the particle masses that we found.

[6]Actually the sign of m_2 comes out negative, but this negative sign can be absorbed by a phase transformation of the spinor field.

10^{10} GeV ... 10^{15} GeV is in fact reasonable. Then m_2 naturally describes light neutrinos. In GUT theories this is called the "seesaw mechanism". Seesaw (in German "Schaukel") describes the process that one heavy mass (m_1) arranges for another mass (m_2) to become very light. The latter then agrees with observations, while the former is so heavy that it escapes observations.

Chapter 7

Three Generations of Quarks and Leptons

Once we will add the two remaining generations, an interesting additional effect emerges — the explicit *breaking of CP invariance*, i.e. the combined charge conjugation and parity transformation.

7.1 The CKM Quark Mixing Matrix

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

$$\left(\begin{array}{c} \nu_{eL} \\ e_L \end{array}\right) , \ e_R \ ; \quad \left(\begin{array}{c} u'_L \\ d'_L \end{array}\right) , \ u'_R \ , \ d'_R \ . \tag{7.1.1}$$

Here we have modified the notation in two respects:

- The neutrino that we dealt with so far is now denoted as the *electron-neutrino* ν_e, so it can be distinguished from the further neutrinos that we are about to add.
- We now write the weak interaction eigenstates for the up and down quarks as u' and d' (so far we called them simply u and d). Once we

add the other generations, u' and d' mix with the other quarks to form the mass eigenstates u and d.

In the second generation, we have the *muon* μ (as a heavier copy of the electron) and its neutrino ν_μ, as well as *charm and strange quarks* (as the heavier copies of up and down quarks). The lepton and quark multiplets of the second generation then take the form

$$\begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix} , \ \mu_R \ ; \quad \begin{pmatrix} c'_L \\ s'_L \end{pmatrix} , \ c'_R , \ s'_R \ . \tag{7.1.2}$$

The charge assignments (Q, Y and T_3) are exactly the same as in the first generation. The heavy fermions tend to decay into the light ones. The heavier they are, the faster this happens, and it is more difficult to generate such particles at all. Based on the concept of one generation, "strange" effects were sometimes observed and related to the s quark, which was then completed to a generation by the subsequent discovery of the c quark.

Later on, yet an other generation was revealed step by step. In the third generation we have the *tauon* τ, its neutrino ν_τ, as well as *top and bottom* (or truth and beauty) quarks

$$\begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} , \ \tau_R \ ; \quad \begin{pmatrix} t'_L \\ b'_L \end{pmatrix} , \ t'_R \ , \ b'_R \ . \tag{7.1.3}$$

As a last ingredient, the top quark was found experimentally in the Tevatron proton–anti-proton collider at Fermilab (near Chicago) in 1995. Its existence had been expected long before on theoretical grounds. The Standard Model only works if generations are complete, and the b quark had been observed already in 1977. However, the Standard Model could not predict the top mass, which was found around 180 GeV.

Let us be more precise now about the lepton number, which we introduced before in eq. (6.3.1). Actually, each lepton carries a generation specific lepton number, we call them L_e, L_μ and L_τ. For (anti-)fermions this number is 1 (-1) in the corresponding generation, and zero otherwise. Usually, even the generation specific lepton number is conserved. A typical example is the decay

$$\mu \to e + \bar{\nu}_e + \nu_\mu \ , \tag{7.1.4}$$

which sets in after a muon life time of $2.2 \cdot 10^{-6}$ sec. The tauon is still significantly heavier than the muon ($m_\tau \simeq 1.8$ GeV vs. $m_\mu \simeq 106$ MeV) hence its life-time is much shorter (about $3 \cdot 10^{-13}$ sec). It can decay either into $e + \bar{\nu}_e + \nu_\tau$, or into $\mu + \bar{\nu}_\mu + \nu_\tau$ (where the muon will soon decay again), or even into hadrons built from the first quark generation.

The conservation of the lepton number in each generation also restricts the possible Z decays, in addition to charge conservation. Therefore, the leptonic decay of the Z-boson can only lead to a lepton and its own anti-lepton. The decay width of Z allows us to sum up the leptonic decay channels, and thus to identify the number of generations. The result — found in particular in LEP experiments at CERN — implies that there are no further leptons beyond these three generations.
A conceivable objection is, however, a lepton generation which is so heavy that the Z-boson cannot decay into any of its members. However, given the sequence of masses found so far, this scenario seems unlikely; it would require a new neutrino with mass $m_\nu > m_Z/2 \simeq 45.6$ GeV.

Up, charm, and top quarks are indistinguishable from the point of view of the electroweak and strong interactions. Therefore their mass eigenstates can mix to build the states appearing in the gauge interaction terms. The down, strange and bottom quarks can do the same. On the other hand, a mixing between up and strange quarks, for example, is forbidden because they sit in different positions of $SU(2)_L$ doublets.
After spontaneous symmetry breaking, the most general quark mass term — expressed in the gauge eigenstates — takes the form

$$(\bar{d}'_L,\ \bar{s}'_L,\ \bar{b}'_L) M^D \begin{pmatrix} d'_R \\ s'_R \\ b'_R \end{pmatrix} + (\bar{u}'_L,\ \bar{c}'_L,\ \bar{t}'_L) M^U \begin{pmatrix} u'_R \\ c'_R \\ t'_R \end{pmatrix} . \qquad (7.1.5)$$

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

$$U_L^{D\dagger} M^D U_R^D = \mathrm{diag}(m_d, m_s, m_b)\ ,\ U_L^{U\dagger} M^U U_R^U = \mathrm{diag}(m_u, m_c, m_t)\ , \quad (7.1.6)$$

where the four matrices $U^D_{L,R}\,, U^U_{L,R}$ are all unitary, and physics tells us that quark masses $m_u, m_d, \dots, m_t$ are real and positive. These bi-unitary transformations relate the weak interaction eigenstates to the mass eigenstates

$$\begin{pmatrix} d'_L \\ s'_L \\ b'_L \end{pmatrix} = U_L^D \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} , \quad \begin{pmatrix} d'_R \\ s'_R \\ b'_R \end{pmatrix} = U_R^D \begin{pmatrix} d_R \\ s_R \\ b_R \end{pmatrix} ,$$
$$\begin{pmatrix} u'_L \\ c'_L \\ t'_L \end{pmatrix} = U_L^U \begin{pmatrix} u_L \\ c_L \\ t_L \end{pmatrix} , \quad \begin{pmatrix} u'_R \\ c'_R \\ t'_R \end{pmatrix} = U_R^U \begin{pmatrix} u_R \\ c_R \\ t_R \end{pmatrix} . \quad (7.1.7)$$

Let us now express the *weak interaction currents* in the basis of mass eigenstates. The *neutral* quark current contains terms such as

$$\bar{u}'_L \gamma_\mu u'_L + \bar{c}'_L \gamma_\mu c'_L + \bar{t}'_L \gamma_\mu t'_L \qquad \text{or} \qquad \bar{d}'_R \gamma_\mu d'_R + \bar{s}'_R \gamma_\mu s'_R + \bar{b}'_R \gamma_\mu b'_R \, .$$

When we rotate these terms into the basis of mass eigenstates, we can simply drop the primes, because $U_L^{U\dagger} U_L^U = \mathbb{1}$, $U_R^{D\dagger} U_R^D = \mathbb{1}$. Hence, the neutral current interactions do not lead to changes among different quark flavours: the Standard Model is *free of flavour-changing neutral currents.* This is a characteristic of the Standard Model, which does not hold for a number of different approaches to describe particle physics. Hence this property has been verified to a high precision in numerous experiments.
The *charged* currents, on the other hand, take the form

$$\begin{aligned} j_\mu^+ &= \bar{u}'_L \gamma_\mu d'_L + \bar{c}'_L \gamma_\mu s'_L + \bar{t}'_L \gamma_\mu b'_L = (\bar{u}_L, \bar{c}_L, \bar{t}_L) \gamma_\mu U_L^{U\dagger} U_L^D \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} \\ &= (\bar{u}_L, \bar{c}_L, \bar{t}_L) \gamma_\mu V \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} , \; j_\mu^- = (\bar{d}_L, \bar{s}_L, \bar{b}_L) \gamma_\mu V^\dagger \begin{pmatrix} u_L \\ c_L \\ t_L \end{pmatrix} . \end{aligned} \quad (7.1.8)$$

Here we have introduced the *Cabbibo-Kobayashi-Maskawa (CKM) quark mixing matrix*

$$V = U_L^{U\dagger} U_L^D \in U(3) \, . \quad (7.1.9)$$

This matrix describes the extent of flavour-changing in the charged current interactions of the Standard Model.

Let us count the number of physical parameters in the mixing matrix V for the case of N generations. We proceed in three steps:

- Since V is unitary, one would naively expect N^2 parameters.

- However, one can change the matrices U_L^U and U_L^D by multiplying them with diagonal unitary matrices — we call them D^U and D^D — from the right,

$$U_L'^U = U_L^U D^U \ , \quad U_L'^D = U_L^D D^D \ . \tag{7.1.10}$$

This still leaves the resulting mass matrices diagonal, and it turns the matrix V into

$$V' = U_L'^{U\dagger} U_L'^D = D^{U\dagger} V D^D \ . \tag{7.1.11}$$

Our requirement was the diagonalisation of the mass matrices; now we see that this can be achieved in different ways. The corresponding ambiguity should be subtracted from the set of physical parameters. To be more explicit: the matrices D^U and D^D have $2N$ parameters (the complex phases on their diagonals), which should not be counted as physical parameters in V. These phases are fixed by the condition that the quark masses have to be real positive.

- However, an overall phase factor common to both, D^U and D^D, would not affect V' at all. This means that in the preceding step we wanted to subtract one parameter, which does not actually exist. The correct number of parameters to be subtracted is therefore $2N - 1$.

Hence the proper counting of physical parameters in the CKM matrix is

$$N^2 - (2N - 1) = (N - 1)^2 \ . \tag{7.1.12}$$

With a single generation there is no mixing and hence no free parameter.

With two generations there is one physical parameter. This situation was assumed for some time, and the allowed quark mixing was described by the so-called *Cabbibo angle* θ_C.
It is instructive to take a closer look at the case of $N = 2$ generations: a general unitary 2×2 matrix can be written as

$$V = \begin{pmatrix} Ae^{i\varphi} & Be^{i\varphi} \\ -B^* & A^* \end{pmatrix} = \begin{pmatrix} |A|e^{i\alpha}e^{i\varphi} & |B|e^{i\beta}e^{i\varphi} \\ -|B|e^{-i\beta} & |A|e^{-i\alpha} \end{pmatrix} , \tag{7.1.13}$$

where A and B are arbitrary complex numbers with phases α and β, and $|A|^2 + |B|^2 = 1$. (φ is the phase of Det V, which generalises our former representation of $SU(2)$ matrices.) We can now use the freedom to introduce D^U and D^D in order to change any given V to

$$V' = D^{U\dagger} V D^D = \begin{pmatrix} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{pmatrix} . \tag{7.1.14}$$

Choosing

$$D^U = \mathrm{diag}(e^{i(\varphi+\beta)}, e^{-i\alpha}) , \quad D^D = \mathrm{diag}(e^{i(\beta-\alpha)}, 1) , \tag{7.1.15}$$

one indeed turns the general $U(2)$ matrix V (which seems to have four parameters) into the special form $V' \in SO(2)$, which only depends on the Cabbibo angle. Experiments led to a Cabbibo angle of $\theta_c \simeq 13^0$. Since it is non-zero, flavour-changing weak decays do occur, but their transition rate is slowed down by the relatively modest mixing angle.

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

$$(N-1)^2 - \frac{N(N-1)}{2} = \frac{(N-1)(N-2)}{2} > 0 . \tag{7.1.16}$$

For example, for the physical case of $N = 3$ generations, the CKM matrix contains $(N-1)(N-2)/2 = 1$ complex phase, in addition to $N(N-1)/2 = 3$ Cabbibo-type Euler angles. In the quark sector alone the Yukawa couplings give rise to ten free parameters of the Standard Model — the six quark masses, three mixing angles, and one complex phase. Experimentalists are working hard on the identification of these parameters and there are constraints for them based on a variety of processes.
For instance, one considers the interaction between two leptonic currents trough a gauge boson $W^\pm$. Such a scattering amplitude is proportional to the product of two CKM matrix elements — one for each flavour changing transition in the two currents. If we invert the directions of the scattering and replace the particles involved by their anti-particles (and vice versa), we obtain the complex conjugate of the above product of matrix elements, c.f. eq. (7.1.8). In total this means that a CP transformation changes V to $V^\dagger$. Therefore, the complex phase in the CKM matrix implies a *violation of CP invariance* (if this phase is non-zero).
The breaking of CP invariance was in fact observed already in 1966 at CERN

in the decay of neutral kaons. For a long period, this remained the only process where CP violation could be detected. Recently it was reported that this has also been achieved in decay of neutral mesons which include the b quark.

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

$$W = U_L^{\nu\dagger} U_L^E \ . \tag{7.1.17}$$

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

Chapter 8

The Structure of the Strong Interactions

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

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

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

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

8.1 Quantum Chromodynamics

We introduce the gluons via an algebra-valued gauge potential

$$G_\mu(x) = iG^a_\mu(x)\lambda_a, \ a \in \{1, 2, ..., 8\}. \tag{8.1.1}$$

The gluon field strength is

$$G_{\mu\nu}(x) = \partial_\mu G_\nu(x) - \partial_\nu G_\mu(x) + g_s[G_\mu(x), G_\nu(x)]. \tag{8.1.2}$$

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

$$G'_\mu(x) = g(x)(G_\mu(x) + \frac{1}{g_s}\partial_\mu)g^+(x). \tag{8.1.3}$$

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

$$G'_{\mu\nu}(x) = g(x)G_{\mu\nu}(x)g^+(x). \tag{8.1.4}$$

The QCD Lagrange function takes the gauge invariant form

$$\begin{aligned}\mathcal{L}_{QCD}(\bar{\Psi}, \Psi, G_\mu) &= \sum_f \bar{\Psi}_f(x)(i\gamma^\mu(\partial_\mu + g_s G_\mu(x)) - m_f)\Psi_f(x) \\ &- \frac{1}{4}\mathrm{Tr}G^{\mu\nu(x)}G_{\mu\nu(x)}.\end{aligned} \tag{8.1.5}$$

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

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

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

$$Z = \int \mathcal{D}\bar{\Psi}\mathcal{D}\Psi\mathcal{D}G \exp(-i \int d^4x \ \mathcal{L}_{QCD}(\bar{\Psi}, \Psi, G_\mu)) \tag{8.1.6}$$

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

$$\Psi(x) = Z_\Psi(\varepsilon)^{1/2}\Psi^R(x),\ G_\mu(x) = Z_G(\varepsilon)^{1/2}G_\mu^R(x), \tag{8.1.7}$$

and also the coupling constant is renormalized via

$$g_s = \frac{Z(\varepsilon)}{Z_\Psi(\varepsilon)Z_G(\varepsilon)^{1/2}}g_s^R. \tag{8.1.8}$$

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

$$\Gamma^R_{n_\Psi,n_G}(k_i,p_j) = \lim_{\varepsilon\to 0} Z_\Psi(\varepsilon)^{n_\Psi/2}Z_G(\varepsilon)^{n_G/2}\Gamma_{n_\Psi,n_G}(k_i,p_j,\varepsilon). \tag{8.1.9}$$

Demanding convergence of the renormalized vertex function fixes the divergent part of the Z-factors. To fix the finite part as well one must specify renormalization conditions. In QCD this can be done using the vertex functions $\Gamma_{0,2}$, $\Gamma_{2,0}$ and $\Gamma_{2,1}$, i.e. the inverse gluon and quark propagators and the quark-gluon vertex. As opposed to QED, where mass and charge of the electron are directly observable, in QCD one chooses an arbitrary scale $\mathcal{M}$ to formulate the renormalization conditions

$$\begin{aligned}
&\Gamma^R_{0,2}(p,-p)^{\mu\nu}_{ab}|_{p^2=-\mathcal{M}^2} = i(-g_{\mu\nu}p^2 + p^\mu p^\nu)\delta_{ab},\\
&\Gamma^R_{2,0}(k,k)|_{k^2=-\mathcal{M}^2} = i\gamma^\mu k_\mu,\\
&\Gamma^R_{2,1}(k,k,k)^\mu_a|_{k^2=-\mathcal{M}^2} = -ig_s^R\frac{\lambda_a}{2}\gamma^\mu.
\end{aligned} \tag{8.1.10}$$

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

$$g_s^R = g_s^R(g,\varepsilon,\mathcal{M}). \tag{8.1.11}$$

This relation defines the β-function

$$\beta(g_s^R) = \lim_{\varepsilon\to 0}\mathcal{M}\frac{\partial}{\partial\mathcal{M}}g_s^R(g,\varepsilon,\mathcal{M}). \tag{8.1.12}$$

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

$$\beta(g_s^R) = -\frac{(g_s^R)^3}{16\pi^2}(11 - \frac{2}{3}N_f). \tag{8.1.13}$$

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

$$11 - \frac{2}{3}N_f > 0 \ \Rightarrow \ N_f \leq 16, \tag{8.1.14}$$

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

$$\begin{aligned}
&\beta(g_s^R) = \mathcal{M}\frac{\partial}{\partial \mathcal{M}}g_s^R = -\frac{(g_s^R)^3}{16\pi^2}(11 - \frac{2}{3}N_f) \ \Rightarrow \\
&\frac{\partial g_s^R}{\partial \mathcal{M}}/(g_s^R)^3 = \frac{1}{2}\frac{\partial (g_s^R)^2}{\partial \mathcal{M}}/(g_s^R)^4 = -\frac{11 - \frac{2}{3}N_f}{16\pi^2}\frac{1}{\mathcal{M}} \ \Rightarrow \\
&\frac{\partial (g_s^R)^2}{(g_s^R)^4} = -\frac{33 - 2N_f}{24\pi^2}\frac{\partial \mathcal{M}}{\mathcal{M}} \ \Rightarrow \ \frac{1}{(g_s^R)^2} = \frac{33 - 2N_f}{24\pi^2}\log\frac{\mathcal{M}}{\Lambda_{QCD}}.
\end{aligned} \tag{8.1.15}$$

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

$$\alpha_s^R = \frac{(g_s^R)^2}{4\pi}, \tag{8.1.16}$$

we obtain

$$\alpha_s^R(\mathcal{M}) = \frac{6\pi}{33 - 2N_f}\frac{1}{\log(\mathcal{M}/\Lambda_{QCD})}. \tag{8.1.17}$$

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

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

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

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

Let us consider the quark part of the QCD Lagrange density

$$\mathcal{L}(\bar{\Psi}, \Psi, G_\mu) = \bar{\Psi}(x)(i\gamma^\mu(\partial_\mu + g_s G_\mu(x)) - \mathcal{M})\Psi(x). \qquad (8.2.1)$$

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

$$\Psi_R(x) = \frac{1}{2}(1+\gamma_5)\Psi(x), \ \Psi_L(x) = \frac{1}{2}(1-\gamma_5)\Psi(x), \ \Psi(x) = \Psi_R(x) + \Psi_L(x). \qquad (8.2.2)$$

Here we have used

$$\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3, \ \{\gamma^\mu, \gamma^\nu\} = 2g_{\mu\nu}, \ \{\gamma^\mu, \gamma_5\} = 0. \qquad (8.2.3)$$

Next we consider the adjoint spinors

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

$$\begin{aligned}\bar{\Psi}_L(x) &= \Psi_L(x)^+\gamma^0 = \Psi(x)^+\frac{1}{2}(1-\gamma_5^+)\gamma^0 = \Psi(x)^+\gamma^0\frac{1}{2}(1+\gamma_5) \\ &= \bar{\Psi}(x)\frac{1}{2}(1+\gamma_5). \end{aligned} \tag{8.2.4}$$

Here we used

$$\gamma^0\gamma_5^+\gamma^0 = -\gamma_5. \tag{8.2.5}$$

Inserting the decomposed spinors in the Lagrange density we obtain

$$\mathcal{L}(\bar{\Psi}, \Psi, G_\mu) = (\bar{\Psi}_R(x) + \bar{\Psi}_L(x))(i\gamma^\mu(\partial_\mu + g_s G_\mu(x)) - \mathcal{M})(\Psi_R(x) + \Psi_L(x)). \tag{8.2.6}$$

First, we investigate the γ^μ term

$$\begin{aligned}&\bar{\Psi}_R(x)i\gamma^\mu(\partial_\mu + g_s G_\mu(x))\Psi_L(x) \\ &= \bar{\Psi}(x)\frac{1}{2}(1-\gamma_5)i\gamma^\mu(\partial_\mu + g_s G_\mu(x))\frac{1}{2}(1-\gamma_5)\Psi(x) \\ &= \bar{\Psi}(x)\frac{1}{4}(1-\gamma_5)(1+\gamma_5)i\gamma^\mu(\partial_\mu + g_s G_\mu(x))\Psi(x) = 0.\end{aligned} \tag{8.2.7}$$

On the other hand, for the mass term one finds

$$\begin{aligned}\bar{\Psi}_R(x)\mathcal{M}\Psi_R(x) &= \bar{\Psi}(x)\frac{1}{2}(1-\gamma_5)\mathcal{M}\frac{1}{2}(1+\gamma_5)\Psi(x) \\ &= \bar{\Psi}(x)\frac{1}{4}(1-\gamma_5)(1+\gamma_5)\mathcal{M}\Psi(x) = 0.\end{aligned} \tag{8.2.8}$$

Hence, we can write

$$\begin{aligned}\mathcal{L}(\bar{\Psi}, \Psi, G_\mu) &= \bar{\Psi}_R(x)i\gamma^\mu(\partial_\mu + g_s G_\mu(x))\Psi_R(x) \\ &+ \bar{\Psi}_L(x)i\gamma^\mu(\partial_\mu + g_s G_\mu(x))\Psi_L(x) \\ &- \bar{\Psi}_R(x)\mathcal{M}\Psi_L(x) - \bar{\Psi}_L(x)\mathcal{M}\Psi_R(x).\end{aligned} \tag{8.2.9}$$

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

$$\begin{aligned}&\Psi'_R(x) = R\Psi_R(x),\ \bar{\Psi}'(x) = \bar{\Psi}_R(x)R^+,\ R \in U(N_f)_R, \\ &\Psi'_L(x) = L\Psi_L(x),\ \bar{\Psi}'(x) = \bar{\Psi}_L(x)L^+,\ L \in U(N_f)_L.\end{aligned} \tag{8.2.10}$$

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

$$SU(N_f)_L \otimes SU(N_f)_R \otimes U(1)_{L=R} = SU(N_f)_L \otimes SU(N_f)_R \otimes U(1)_B. \quad (8.2.11)$$

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 $\mathcal{M} = m\mathbf{1}$, one would have

$$\bar{\Psi}'_R(x)\mathcal{M}\Psi'_L(x) = \bar{\Psi}_R(x)R^+ m\mathbf{1}L\Psi_L(x) = \bar{\Psi}_R(x)R^+ L\mathcal{M}\Psi_L(x). \quad (8.2.12)$$

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

$$SU(N_f)_{L=R} \otimes U(1)_{L=R} = SU(N_f)_F \otimes U(1)_B, \quad (8.2.13)$$

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

$$\otimes_f U(1)_f = U(1)_u \otimes U(1)_d \otimes U(1)_s. \quad (8.2.14)$$

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.

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

$$\begin{aligned} J_\mu^{La}(x) &= \bar{\Psi}_L(x)\gamma_\mu \frac{\sigma^a}{2}\Psi_L(x), \\ J_\mu^{Ra}(x) &= \bar{\Psi}_R(x)\gamma_\mu \frac{\sigma^a}{2}\Psi_R(x), \end{aligned} \quad (8.2.15)$$

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

$$V_\mu^a(x) = J_\mu^{La}(x) + J_\mu^{Ra}(x)$$

$$
\begin{aligned}
&= \bar{\Psi}(x)\frac{1}{2}(1+\gamma_5)\gamma_\mu\frac{\sigma^a}{2}\frac{1}{2}(1-\gamma_5)\Psi(x) \\
&+ \bar{\Psi}(x)\frac{1}{2}(1-\gamma_5)\gamma_\mu\frac{\sigma^a}{2}\frac{1}{2}(1+\gamma_5)\Psi(x) \\
&= \bar{\Psi}(x)\frac{1}{2}(1+\gamma_5)\gamma_\mu\frac{\sigma^a}{2}\Psi(x) + \bar{\Psi}(x)\frac{1}{2}(1-\gamma_5)\gamma_\mu\frac{\sigma^a}{2}\Psi(x) \\
&= \bar{\Psi}(x)\gamma_\mu\frac{\sigma^a}{2}\Psi(x), \\
A^a_\mu(x) &= J^{La}_\mu(x) - J^{Ra}_\mu(x) = \bar{\Psi}(x)\gamma_5\gamma_\mu\frac{\sigma^a}{2}\Psi(x). \qquad (8.2.16)
\end{aligned}
$$

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

$$\langle\Phi|J^{La}_\mu(x)J^{Rb}_\nu(y)|\Phi\rangle = \langle\Phi|J^{Ra}_\mu(x)J^{Lb}_\nu(y)|\Phi\rangle = 0, \qquad (8.2.17)$$

and hence

$$\langle\Phi|V^a_\mu(x)V^b_\nu(y)|\Phi\rangle = \langle\Phi|A^a_\mu(x)A^b_\nu(y)|\Phi\rangle. \qquad (8.2.18)$$

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

Let us now consider the states

$$
\begin{aligned}
Q^a_V|0\rangle &= \int d^3x V^a_0(\vec{x},0)|0\rangle, \\
Q^a_A|0\rangle &= \int d^3x A^a_0(\vec{x},0)|0\rangle, \qquad (8.2.19)
\end{aligned}
$$

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

$$Q^a_V|\Phi\rangle = Q^a_A|\Phi\rangle = 0. \qquad (8.2.20)$$

The real QCD vacuum is not chirally invariant because

$$Q_A^a|0\rangle \neq 0. \tag{8.2.21}$$

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

$$[H_{QCD}, Q_A^a] = 0. \tag{8.2.22}$$

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

$$H_{QCD} Q_A^a|0\rangle = Q_A^a H_{QCD}|0\rangle = 0 \tag{8.2.23}$$

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

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

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

$$\langle\bar{\Psi}\Psi\rangle = \langle 0|\bar{\Psi}(x)\Psi(x)|0\rangle = \langle 0|\bar{\Psi}_R(x)\Psi_L(x) + \bar{\Psi}_L(x)\Psi_R(x)|0\rangle. \tag{8.2.24}$$

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

Part III

SPECIAL TOPICS

Chapter 9

Phenomenology of the Strong Interactions

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

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

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

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

9.1 Isospin Symmetry

Proton and neutron have almost the same masses

$$M_p = 0.938 \text{ GeV}, \ M_n = 0.940 \text{ GeV}. \tag{9.1.1}$$

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

$$M_{\pi^+} = M_{\pi^-} = 0.140 \text{ GeV}, \ M_{\pi^0} = 0.135 \text{ GeV}. \tag{9.1.2}$$

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

$$M_{\Delta^{++}} \approx M_{\Delta^+} \approx M_{\Delta^0} \approx M_{\Delta^-} \approx 1.232 \text{ GeV} \tag{9.1.3}$$

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

$$M_{\rho^+} \approx M_{\rho^0} \approx M_{\rho^-} \approx 0.768 \text{ GeV}. \tag{9.1.4}$$

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 9.1: *The isospin classification of hadrons.*

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

$$S_z = -S, -S + 1, ..., S - 1, S. \tag{9.1.5}$$

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

$$I_3 = -I, -I + 1, ..., I - 1, I. \tag{9.1.6}$$

A representation with isospin I contains $2I + 1$ states and is denoted by $\{2I+1\}$. We can classify the hadrons by their isospin. This is done in table 9.1. For the baryons (nucleon and Δ) isospin projection and electric charge are related by $Q = I_3 + \frac{1}{2}$, and for the mesons (π and ρ) $Q = I_3$.

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

$$\{2\} \otimes \{3\} = \{2\} \oplus \{4\} \tag{9.1.7}$$

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

9.2 Nucleon and Δ-Isobar in the Constituent Quark Model

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

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

$$\{2\} \otimes \{2\} \otimes \{2\} = (\{1\} \oplus \{3\}) \otimes \{2\} = \{2\} \oplus \{2\} \oplus \{4\}, \tag{9.2.1}$$

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

$$Q = I_3 + \frac{1}{2} = \sum_{q=1}^{3} (I_{3q} + \frac{1}{6}) = \sum_{q=1}^{3} Q_q, \tag{9.2.2}$$

and hence we obtain for the charges of the quarks

$$Q_q = I_{3q} + \frac{1}{6},\ Q_u = \frac{1}{2} + \frac{1}{6} = \frac{2}{3},\ Q_d = -\frac{1}{2} + \frac{1}{6} = -\frac{1}{3}. \tag{9.2.3}$$

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

$$\boxed{1}\boxed{2}\boxed{3}_{\,3/2} = uuu \equiv \Delta^{++},$$

$$
\begin{aligned}
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,1/2} &= \frac{1}{\sqrt{3}}(uud + udu + duu) \equiv \Delta^+, \\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,-1/2} &= \frac{1}{\sqrt{3}}(udd + dud + ddu) \equiv \Delta^0, \\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,-3/2} &= ddd \equiv \Delta^-.
\end{aligned}
\tag{9.2.4}
$$

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

We write the general coupling of the three quarks as

$$
\begin{array}{|c|}\hline 1\\\hline\end{array} \otimes \begin{array}{|c|}\hline 2\\\hline\end{array} \otimes \begin{array}{|c|}\hline 3\\\hline\end{array} = \begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array} \oplus \begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1}\end{array} \oplus \begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1}\end{array} \oplus \begin{array}{|c|}\hline 1\\\hline 2\\\hline 3\\\hline\end{array}\,. \tag{9.2.5}
$$

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

$$
\{2\} \otimes \{2\} \otimes \{2\} = \{4\} \oplus \{2\} \oplus \{2\} \oplus \{0\}. \tag{9.2.6}
$$

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

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

$$
\begin{aligned}
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,3/2} &= \uparrow\uparrow\uparrow, \\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,1/2} &= \frac{1}{\sqrt{3}}(\uparrow\uparrow\downarrow \;+\; \uparrow\downarrow\uparrow \;+\; \downarrow\uparrow\uparrow), \\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,-1/2} &= \frac{1}{\sqrt{3}}(\uparrow\downarrow\downarrow \;+\; \downarrow\uparrow\downarrow \;+\; \downarrow\downarrow\uparrow), \\
\begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array}_{\,-3/2} &= \downarrow\downarrow\downarrow\,.
\end{aligned}
\tag{9.2.7}
$$

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

$$|\Delta I_3 S_z\rangle = \begin{array}{|c|c|c|}\hline 1 & 2 & 3 \\ \hline \end{array}_{\,I_3} \begin{array}{|c|c|c|}\hline 1 & 2 & 3 \\ \hline \end{array}_{\,S_z}. \tag{9.2.8}$$

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

$$\begin{aligned} |\Delta \frac{1}{2}\frac{1}{2}\rangle &= \frac{1}{3}(u\uparrow u\uparrow d\downarrow + u\uparrow u\downarrow d\uparrow + u\downarrow u\uparrow d\uparrow \\ &+ u\uparrow d\uparrow u\downarrow + u\uparrow d\downarrow u\uparrow + u\downarrow d\uparrow u\uparrow \\ &+ d\uparrow u\uparrow u\downarrow + d\uparrow u\downarrow u\uparrow + d\downarrow u\uparrow u\uparrow). \end{aligned} \tag{9.2.9}$$

One sees explicitly that this state is totally symmetric.

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

$$\begin{aligned} \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array}_{\,1/2} &= \frac{1}{\sqrt{6}}(2uud - udu - duu), \\ \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array}_{\,-1/2} &= \frac{1}{\sqrt{6}}(udd + dud - 2ddu), \\ \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array}_{\,1/2} &= \frac{1}{\sqrt{2}}(udu - duu), \\ \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array}_{\,-1/2} &= \frac{1}{\sqrt{2}}(udd - dud). \end{aligned} \tag{9.2.10}$$

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

$$\ydiagram{2,1}_{\,I_3} \times \ydiagram{2,1}_{\,S_z} = \ydiagram{3}_{\,I_3 S_z} \oplus \ydiagram{2,1}_{\,I_3 S_z} \oplus \ydiagram{1,1,1}_{\,I_3 S_z} \tag{9.2.11}$$

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

$$|N I_3 S_z\rangle = \frac{1}{\sqrt{2}}\left(\begin{ytableau} 1 & 2 \\ 3 \end{ytableau}_{\,I_3} \begin{ytableau} 1 & 2 \\ 3 \end{ytableau}_{\,S_z} + \begin{ytableau} 1 & 3 \\ 2 \end{ytableau}_{\,I_3} \begin{ytableau} 1 & 3 \\ 2 \end{ytableau}_{\,S_z} \right). \tag{9.2.12}$$

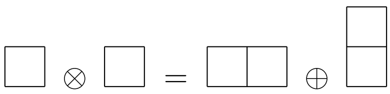
In our construction we have implicitly assumed that the orbital angular momentum of the constituent quarks inside a hadron vanishes. Then the orbital state is completely symmetric in the coordinates of the quarks. The orbital part of the baryon wave function therefore is described by the Young tableau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

$$\ydiagram{1,1,1} = \frac{1}{\sqrt{6}}(rgb - rbg + gbr - grb + brg - bgr). \tag{9.2.13}$$

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

9.3 Anti-Quarks and Mesons

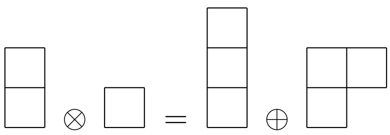
We have seen that the baryons (nucleon and Δ) consist of three constituent quarks (isospin doublets, spin doublets, color triplets). Now we want to construct the mesons (pion and ρ) in a similar manner. Since these particles have spin 0 and 1 respectively, they must contain an even number of constituent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



$$\{3\} \otimes \{3\} = \{6\} \oplus \{\bar{3}\}, \tag{9.3.1}$$

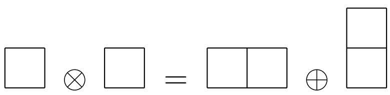
does not contain a singlet as desired by the confinement hypothesis.

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



$$\{\bar{3}\} \otimes \{3\} = \{1\} \oplus \{8\}, \tag{9.3.2}$$

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



$$\{2\} \otimes \{2\} = \{3\} \oplus \{1\}, \tag{9.3.3}$$

and we obtain

$$
\begin{aligned}
\square\!\square_{\,1} &= u\bar{d}, \\
\square\!\square_{\,0} &= \frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d}), \\
\square\!\square_{\,-1} &= d\bar{u}, \\
\begin{array}{c}\square\\[-4pt]\square\end{array}_{0} &= \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d}).
\end{aligned}
\tag{9.3.4}
$$

We proceed analogously for the spin and we obtain

$$
\begin{aligned}
|\pi I_3 S_z\rangle &= \square\!\square_{\,I_3} \begin{array}{c}\square\\[-4pt]\square\end{array}_{S_z}, \\
|\rho I_3 S_z\rangle &= \square\!\square_{\,I_3} \square\!\square_{\,S_z}.
\end{aligned}
\tag{9.3.5}
$$

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

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

$$
\begin{aligned}
|\omega I_3 S_z\rangle &= \begin{array}{c}\square\\[-4pt]\square\end{array}_{I_3} \square\!\square_{\,S_z}, \\
|\eta' I_3 S_z\rangle &= \begin{array}{c}\square\\[-4pt]\square\end{array}_{I_3} \begin{array}{c}\square\\[-4pt]\square\end{array}_{S_z}.
\end{aligned}
\tag{9.3.6}
$$

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

9.4 Strange Hadrons

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

$$M_\pi = 0.138\text{GeV}, M_K = 0.496\text{GeV}, M_\eta = 0.549\text{GeV}, M_{\eta'} = 0.958\text{GeV}, \tag{9.4.1}$$

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

$$M_\rho = 0.770\text{GeV}, M_\omega = 0.783\text{GeV}, M_{K^*} = 0.892\text{GeV}, M_\varphi = 1.020\text{GeV}. \tag{9.4.2}$$

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

$$\{\bar{2}\} \otimes \{2\} = \{1\} \oplus \{3\}. \tag{9.4.3}$$

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

$$\{\bar{3}\} \otimes \{3\} = \{1\} \oplus \{8\}. \tag{9.4.4}$$

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

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

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix},$$
$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{9.4.5}$$

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

$$I_1 = \frac{1}{2}\lambda_1, \ I_2 = \frac{1}{2}\lambda_2, \ I_3 = \frac{1}{2}\lambda_3. \tag{9.4.6}$$

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

$$Y = \frac{1}{\sqrt{3}}\lambda_8, \tag{9.4.7}$$

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

$$\begin{aligned} &I^2 u = \frac{1}{2}(\frac{1}{2}+1)u = \frac{3}{4}u, \ I_3 u = \frac{1}{2}u, \ Yu = \frac{1}{3}u, \\ &I^2 d = \frac{1}{2}(\frac{1}{2}+1)d = \frac{3}{4}d, \ I_3 d = -\frac{1}{2}d, \ Yd = \frac{1}{3}d, \\ &I^2 s = 0, \ I_3 s = 0, \ Ys = -\frac{2}{3}s. \end{aligned} \tag{9.4.8}$$

The electric charge is now given by

$$Q = I_3 + \frac{1}{2}Y, \tag{9.4.9}$$

such that

$$Q_u = \frac{2}{3}, \ Q_d = -\frac{1}{3}, \ Q_s = -\frac{1}{3}, \tag{9.4.10}$$

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

$$\{3\} \otimes \{3\} \otimes \{3\} = \{10\} \oplus 2\{8\} \oplus \{1\} \tag{9.4.11}$$

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

$$\{2\} \otimes \{2\} \otimes \{2\} = \{4\} \oplus 2\{2\} \oplus \{0\}. \tag{9.4.12}$$

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

The baryon masses for the spin 1/2 baryons are

$$M_N = 0.939\text{GeV}, \ M_\Lambda = 1.116\text{GeV}, \ M_\Sigma = 1.193\text{GeV}, \ M_\Xi = 1.318\text{GeV}, \tag{9.4.13}$$

while the spin 3/2 baryon masses are

$$M_\Delta = 1.232\text{GeV},\ M_{\Sigma^*} = 1.385\text{GeV},\ M_{\Xi^*} = 1.530\text{GeV},\ M_\Omega = 1.672\text{GeV}. \tag{9.4.14}$$

Proton and neutron are part of an octet: $\ydiagram{2,1}$ is $\{2\}$ in $SU(2)_I$ and $\{8\}$ in $SU(3)_F$. The Δ-isobar is part of a decouplet: $\ydiagram{3}$ is $\{4\}$ in $SU(2)_I$ and $\{10\}$ in $SU(3)_F$. One does not find an $SU(3)_F$ singlet $\ydiagram{1,1,1}$. This is because a spatially symmetric color singlet wave function is totally antisymmetric. To obtain a totally antisymmetric wave function also the spin part should transform as $\ydiagram{1,1,1}$. Of course, in $SU(2)_S$ this is impossible.

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

$$M_{\Sigma^*} - M_\Delta = M_{\Xi^*} - M_{\Sigma^*} = M_\Omega - M_{\Xi^*} = M_s - M_q. \tag{9.4.15}$$

In fact one finds experimentally

$$M_{\Sigma^*} - M_\Delta = 0.153\text{GeV},\ M_{\Xi^*} - M_{\Sigma^*} = 0.145\text{GeV},\ M_\Omega - M_{\Xi^*} = 0.142\text{GeV}. \tag{9.4.16}$$

9.5 The Gellman-Okubo Baryon Mass Formula

We have seen that the constituent quark model leads to a successful classification of hadron states in terms of flavor symmetry. The results about the hadron dynamics are, however, of more qualitative nature, and the assumption that a hadron is essentially a collection of a few constituent quarks is certainly too naive. The fundamental theory of the strong interactions is

QCD. Here we want to use very basic QCD physics together with group theory to describe patterns in the hadron spectrum. The interaction between quarks and gluons is flavor independent, and therefore $SU(3)_F$ symmetric. Also the gluon self-interaction is flavor symmetric because the gluons are flavor singlets. A violation of flavor symmetry results only from the quark mass matrix

$$\mathcal{M} = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_d & 0 \\ 0 & 0 & m_s \end{pmatrix}. \tag{9.5.1}$$

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

$$\begin{aligned} \mathcal{M} &= \frac{2m_q + m_s}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{m_q - m_s}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \\ &= \frac{2m_q + m_s}{3} 1 + \frac{m_q - m_s}{\sqrt{3}} \lambda_8. \end{aligned} \tag{9.5.2}$$

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

$$H_{QCD} = H_1 + H_8. \tag{9.5.3}$$

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 $|B_1 Y I I_3\rangle$

$$H_1 |B_1 Y I I_3\rangle = M_{B_1} |B_1 Y I I_3\rangle. \tag{9.5.4}$$

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

$$M_B = M_{B_1} + \langle B_1 Y I I_3 | H_8 | B_1 Y I I_3\rangle. \tag{9.5.5}$$

A diagonalization in the space of degenerate states is not necessary, since H_8 transforms as the λ_8 component of an octet, and can therefore not change

Y, I and I_3. Next we will compute the required matrix elements using group theory. Starting with the baryon decouplet, we obtain a nonzero value only if $\{8\}$ and $\{10\}$ can couple to $\{10\}$. Indeed the decouplet appears in the reduction. Using the Wigner-Eckart theorem we obtain

$$\langle B_1 YII_3|H_8|B_1 YII_3\rangle = \langle B_1||H_8||B_1\rangle\langle\{10\}YII_3|\{8\}000\{10\}YII_3\rangle, \tag{9.5.6}$$

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

$$\langle\{10\}YII_3|\{8\}000\{10\}YII_3\rangle = \frac{Y}{\sqrt{8}}. \tag{9.5.7}$$

Then we obtain for the baryon masses in the decouplet

$$M_B = M_{B_1} + \langle B_1||H_8||B_1\rangle\frac{Y}{\sqrt{8}}, \tag{9.5.8}$$

and hence

$$M_{\Sigma^*} - M_\Delta = M_{\Xi^*} - M_{\Sigma^*} = M_\Omega - M_{\Xi^*} = -\frac{1}{\sqrt{8}}\langle B_1||H_8||B_1\rangle. \tag{9.5.9}$$

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

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

$$\{8\}\otimes\{8\} = \{27\}\oplus\{10\}\oplus\{\bar{10}\}\oplus 2\{8\}\oplus\{1\}. \tag{9.5.10}$$

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

$$\begin{aligned}\langle B_1 YII_3|H_8|B_1 YII_3\rangle &= \langle B_1||H_8||B_1\rangle_s\langle\{8\}YII_3|\{8\}000\{8\}YII_3\rangle_s \\ &+ \langle B_1||H_8||B_1\rangle_a\langle\{8\}YII_3|\{8\}000\{8\}YII_3\rangle_a.\end{aligned} \tag{9.5.11}$$

The Clebsch-Gordon coefficients are given by

$$\begin{aligned}\langle\{8\}YII_3|\{8\}000\{8\}YII_3\rangle_s &= \frac{1}{\sqrt{5}}(I(I+1) - \frac{1}{4}Y^2 - 1), \\ \langle\{8\}YII_3|\{8\}000\{8\}YII_3\rangle_a &= \sqrt{\frac{3}{4}}Y,\end{aligned} \tag{9.5.12}$$

and we obtain for the baryon octet

$$M_B = M_{B_1} + \langle B_1||H_8||B_1\rangle_s \frac{1}{\sqrt{5}}(I(I+1) - \frac{1}{4}Y^2 - 1) + \langle B_1||H_8||B_1\rangle_a \sqrt{\frac{3}{4}}Y. \tag{9.5.13}$$

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

$$\begin{aligned} 2M_N + 2M_\Xi &= 4M_{B_1} + \langle B_1||H_8||B_1\rangle_s \frac{4}{\sqrt{5}}(\frac{3}{4} - \frac{1}{4} - 1), \\ M_\Sigma + 3M_\Lambda &= 4M_{B_1} + \langle B_1||H_8||B_1\rangle_s \frac{1}{\sqrt{5}}((2-1) + 3(-1)), \\ 2M_N + 2M_\Xi &= M_\Sigma + 3M_\Lambda. \end{aligned} \tag{9.5.14}$$

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

9.6 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

$$H_1|M_1YII_3\rangle = M_{M_1}|M_1YII_3\rangle. \tag{9.6.1}$$

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

$$M_M = M_{M_1} + \langle M_1YII_3|H_8|M_1YII_3\rangle. \tag{9.6.2}$$

In complete analogy to the baryon octet we obtain

$$M_M = M_{M_1} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(I(I+1) - \frac{1}{4}Y^2 - 1) + \langle M_1||H_8||M_1\rangle_a \sqrt{\frac{3}{4}}Y. \tag{9.6.3}$$

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

$$\begin{aligned} M_{K^+} &= M_{M_1} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(\frac{3}{4} - \frac{1}{4} - 1) + \langle M_1||H_8||M_1\rangle_a \sqrt{\frac{3}{4}}, \\ M_{K^-} &= M_{M_1} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(\frac{3}{4} - \frac{1}{4} - 1) - \langle M_1||H_8||M_1\rangle_a \sqrt{\frac{3}{4}}. \end{aligned} \tag{9.6.4}$$

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

$$\langle M_1||H_8||M_1\rangle_a = 0. \tag{9.6.5}$$

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

$$\begin{aligned} \langle\omega_1|H_8|\omega_1\rangle &= 0, \\ \langle\omega_8|H_8|\omega_8\rangle &= \langle M_1||H_8||M_1\rangle_s \langle\{8\}000|\{8\}000\{8\}000\rangle_s \\ &= \langle M_1||H_8||M_1\rangle_s (-\frac{1}{\sqrt{5}}). \end{aligned} \tag{9.6.6}$$

The actual meson masses are the eigenvalues of the matrix

$$\mathcal{M} = \begin{pmatrix} M_{\omega_1} & \langle\omega_1|H_8|\omega_8\rangle \\ \langle\omega_8|H_8|\omega_1\rangle & M_{\omega_8} - \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}} \end{pmatrix}. \tag{9.6.7}$$

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

$$\begin{aligned} |\varphi\rangle &= \cos\theta|\omega_1\rangle - \sin\theta|\omega_8\rangle, \\ |\omega\rangle &= \sin\theta|\omega_1\rangle + \cos\theta|\omega_8\rangle. \end{aligned} \tag{9.6.8}$$

Here θ is the meson mixing angle. One obtains

$$\begin{aligned} M_\varphi + M_\omega &= M_{\omega_1} + M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s, \\ M_\varphi M_\omega &= M_{\omega_1}(M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s) - |\langle\omega_1|H_8|\omega_8\rangle|^2. \end{aligned} \tag{9.6.9}$$

Also we have

$$\begin{aligned} M_\rho &= M_{\omega_8} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(2-1), \\ M_{K^*} &= M_{\omega_8} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(\frac{3}{4} - \frac{1}{4} - 1), \end{aligned} \tag{9.6.10}$$

and hence

$$\begin{aligned} \frac{4}{3}M_{K^*} - \frac{1}{3}M_\rho &= M_{\omega_8} + \langle M_1||H_8||M_1\rangle_s \frac{1}{\sqrt{5}}(\frac{4}{3}(-\frac{1}{2}) - \frac{1}{3}) \\ &= M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s, \end{aligned} \tag{9.6.11}$$

such that

$$\begin{aligned} M_{\omega_1} &= M_\varphi + M_\omega - \frac{4}{3}M_{K^*} + \frac{1}{3}M_\rho = 0.870\text{GeV}, \\ |\langle \omega_1|H_8|\omega_8\rangle|^2 &= M_{\omega_1}(\frac{4}{3}M_{K^*} - \frac{1}{3}M_\rho) - M_\varphi M_\omega = (0.113\text{GeV})^2. \end{aligned} \tag{9.6.12}$$

The mixing angle is now determined from

$$\begin{aligned} & M_{\omega_1}\cos\theta - \langle \omega_1|H_8|\omega_8\rangle \sin\theta = M_\varphi \cos\theta, \\ & \langle \omega_8|H_8|\omega_1\rangle \cos\theta - (M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s)\sin\theta = -M_\varphi \sin\theta. \end{aligned} \tag{9.6.13}$$

and we obtain

$$\begin{aligned} & (M_{\omega_1} + M_{\omega_8} - \frac{1}{\sqrt{5}}\langle M_1||H_8||M_1\rangle_s)\sin\theta\cos\theta - \langle \omega_1|H_8|\omega_8\rangle = \\ & 2M_\varphi \sin\theta\cos\theta, \end{aligned} \tag{9.6.14}$$

and hence

$$\frac{1}{2}\sin(2\theta) = \pm\frac{\sqrt{(M_\varphi + M_\omega - \frac{4}{3}M_{K^*} + \frac{1}{3}M_\rho)(\frac{4}{3}M_{K^*} - \frac{1}{3}M_\rho) - M_\varphi M_\omega}}{M_\varphi - M_\omega}. \tag{9.6.15}$$

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

$$\begin{aligned} |\varphi\rangle &\approx s\bar{s} \;\text{ or }\; \frac{1}{3}(2u\bar{u} + 2d\bar{d} - s\bar{s}), \\ |\omega\rangle &\approx \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d}) \;\text{ or }\; -\frac{1}{\sqrt{18}}(u\bar{u} + d\bar{d} + 4s\bar{s}). \end{aligned} \tag{9.6.16}$$

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

$$|\varphi\rangle \approx s\bar{s}, \quad |\omega\rangle \approx \frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d}). \tag{9.6.17}$$

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

Chapter 10

Topology of Gauge Fields

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

10.1 The Anomaly

Let us consider the baryon number current in the Standard Model

$$J_\mu(x) = \sum_{\mathrm{f}} \bar{\Psi}^{\mathrm{f}}(x)\gamma_\mu \Psi^{\mathrm{f}}(x), \tag{10.1.1}$$

where $\Psi^{\mathrm{f}}(x)$ is the quark field for flavor $\mathrm{f} = u, d, s, \ldots$. The Lagrangian of the Standard Model is invariant under global $U(1)_{\mathrm{B}}$ baryon number transformations. The corresponding Noether current $J_\mu(x)$ is hence conserved at the classical level

$$\partial_\mu J_\mu(x) = 0. \tag{10.1.2}$$

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

$$\partial_\mu J_\mu(x) = N_{\rm g} P(x). \tag{10.1.3}$$

Here $N_{\rm g}$ is the number of generations ($N_{\rm g} = 3$ in the Standard Model), and

$$P(x) = -\frac{1}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}\left[W_{\mu\nu}(x)W_{\rho\sigma}(x)\right] \tag{10.1.4}$$

is the *Chern-Pontryagin density.* Here $W_{\mu\nu}$ is the field strength tensor of the $SU(2)_{\rm L}$ gauge field.

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

$$J^5_\mu(x) = \sum_{\rm f} \bar\Psi^{\rm f}(x)\gamma_5\gamma_\mu\Psi^{\rm f}(x). \tag{10.1.5}$$

The Lagrangian of QCD with massless quarks is invariant under global $U(1)_{\rm A}$ transformations, and hence also $J^5_\mu(x)$ is conserved at the classical level

$$\partial_\mu J^5_\mu(x) = 0. \tag{10.1.6}$$

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

$$\partial^\mu J^5_\mu(x) = 2N_{\rm f} P(x). \tag{10.1.7}$$

Now $N_{\rm f}$ is the number of quark flavors, and

$$P(x) = -\frac{1}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}\left[G_{\mu\nu}(x)G_{\rho\sigma}(x)\right] \tag{10.1.8}$$

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

In the following we consider the topology of a general non-Abelian vector potential $G_\mu(x)$. The anomaly equation can be derived in perturbation theory and it follows from a triangle diagram. The Chern-Pontryagin density can be written as a total divergence

$$P(x) = \partial_\mu \Omega^{(0)}_\mu(x), \tag{10.1.9}$$

where $\Omega_\mu^{(0)}(x)$ is the *Chern-Simons density* or 0-cochain, which is given by

$$\Omega_\mu^{(0)}(x) = -\frac{1}{8\pi^2}\epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}\left[G_\nu(x)(\partial_\rho G_\sigma(x) + \frac{2}{3}G_\rho(x)G_\sigma(x)\right]. \qquad (10.1.10)$$

It is a good exercise to convince oneself that this satisfies eq.(10.1.9). We can now formally construct a conserved current

$$\tilde{J}_\mu^5(x) = J_\mu^5(x) - 2N_\mathrm{f}\Omega_\mu^{(0)}(x), \qquad (10.1.11)$$

because

$$\partial_\mu \tilde{J}_\mu(x) = \partial_\mu J_\mu(x) - 2N_\mathrm{f}P(x) = 0. \qquad (10.1.12)$$

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

10.2 Topological Charge

In this section, we define the topological charge of a Euclidean non-Abelian field configuration. We like to point out, that the concept of an intervalued topological charge does not carry over to Minkowski space-time. In general, field configurations in Euclidean space-time do not represent physical processes in real time. The topological charge is defined as

$$\begin{aligned} Q &= -\frac{1}{32\pi^2}\int d^4x\ \epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}\left[G_{\mu\nu}(x)G_{\rho\sigma}(x)\right] = \int d^4x\ P(x) \\ &= \int d^4x\ \partial_\mu\Omega_\mu^{(0)}(x) = \int_{S^3} d^3\sigma_\mu\ \Omega_\mu^{(0)}(x)\ . \qquad (10.2.1) \end{aligned}$$

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

$$G_\mu(x) = g(x)\partial_\mu g(x)^\dagger\ . \qquad (10.2.2)$$

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

$$\begin{aligned}
Q &= -\frac{1}{8\pi^2}\int_{S^3} d^3\sigma_\mu\ \epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}\left[(g\partial_\nu g^\dagger)(\partial_\rho(g\partial_\sigma g^\dagger)\right.\\
&+ \left.\frac{2}{3}(g\partial_\rho g^\dagger)(g\partial_\sigma g^\dagger))\right]\\
&= -\frac{1}{8\pi^2}\int_{S^3} d^3\sigma_\mu\ \epsilon_{\mu\nu\rho\sigma}\\
&\times\ \mathrm{Tr}\left[-(g\partial_\nu g^\dagger)(g\partial_\rho g^\dagger)(g\partial_\sigma g^\dagger)\right.\\
&+ \left.\frac{2}{3}(g\partial_\nu g^\dagger)(g\partial_\rho g^\dagger)(g\partial_\sigma g^\dagger)\right]\\
&= \frac{1}{24\pi^2}\int_{S^3} d^3\sigma_\mu\ \epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}\left[(g\partial_\nu g^\dagger)(g\partial_\rho g^\dagger)(g\partial_\sigma g^\dagger)\right].
\end{aligned} \tag{10.2.3}$$

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

$$g : S^3 \to SU(N). \tag{10.2.4}$$

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

$$\Pi_3[SU(N)] = \mathbf{Z}. \tag{10.2.5}$$

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

is exactly that integer. For this purpose we decompose

$$g = VW \ , \qquad W = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & \tilde{g}_{11} & \tilde{g}_{12} & \dots & \tilde{g}_{1N} \\ 0 & \tilde{g}_{21} & \tilde{g}_{22} & \dots & \tilde{g}_{1N} \\ . & . & . & & . \\ . & . & . & & . \\ 0 & \tilde{g}_{N1} & \tilde{g}_{N2} & \dots & \tilde{g}_{NN} \end{pmatrix} \ , \tag{10.2.6}$$

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

$$V = \begin{pmatrix} g_{11} & -g_{21}^* & -\frac{g_{31}^*(1+g_{11})}{1+g_{11}^*} & \dots & -\frac{g_{N1}^*(1+g_{11})}{1+g_{11}^*} \\ g_{21} & \frac{1+g_{11}^*-|g_{21}|^2}{1+g_{11}} & -\frac{g_{31}^* g_{21}}{1+g_{11}^*} & \dots & -\frac{g_{N1}^* g_{21}}{1+g_{11}^*} \\ g_{31} & -\frac{g_{21}^* g_{31}}{1+g_{11}} & \frac{1+g_{11}^*-|g_{31}|^2}{1+g_{11}^*} & \dots & -\frac{g_{N1}^* g_{31}}{1+g_{11}^*} \\ . & . & . & & . \\ . & . & . & & . \\ g_{N1} & -\frac{g_{21}^* g_{N1}}{1+g_{11}} & -\frac{g_{31}^* g_{N1}}{1+g_{11}^*} & \dots & \frac{1+g_{11}^*-|g_{N1}|^2}{1+g_{11}^*} \end{pmatrix} \in SU(N) \ . \tag{10.2.7}$$

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

$$\begin{aligned} &\epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}\left[(VW)\partial_\nu(VW)^\dagger(VW)\partial_\rho(VW)^\dagger(VW)\partial_\sigma(VW)^\dagger\right] = \\ &\epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}\left[(V\partial_\nu V^\dagger)(V\partial_\rho V^\dagger)(V\partial_\sigma V^\dagger)\right] \\ &+\epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}\left[(W\partial_\nu W^\dagger)(W\partial_\rho W^\dagger)(W\partial_\sigma W^\dagger)\right] \\ &+3\partial_\nu\epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}\left[((V\partial_\rho V^\dagger)(W\partial_\sigma W^\dagger)\right] . \end{aligned} \tag{10.2.8}$$

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

$$\begin{aligned} Q &= \frac{1}{24\pi^2}\int_{S^3} d^3\sigma_\mu \epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(g\partial_\nu g^\dagger)(g\partial_\rho g^\dagger)(g\partial_\sigma g^\dagger)] \\ &= \frac{1}{24\pi^2}\int_{S^3} d^3\sigma_\mu \epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[(V\partial_\nu V^\dagger)(V\partial_\rho V^\dagger)(V\partial_\sigma V^\dagger) \\ &+ (W\partial_\nu W^\dagger)(W\partial_\rho W^\dagger)(W\partial_\sigma W^\dagger)]. \end{aligned} \tag{10.2.9}$$

The ∂_ν term of the formula eq.(10.2.8) drops out using Gauss' law together with the fact that S^3 has no boundary. It follows that the topological charge of a product of two gauge transformations V and W is the sum of the topological charges of V and W. Since V only depends on $g_{11}, g_{21}, \dots, g_{N1}$, it can be viewed as a map of S^3 into the sphere S^{2N-1}

$$V : S^3 \to S^{2N-1}. \tag{10.2.10}$$

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

$$\Pi_3[S^{2N-1}] = \{0\}. \tag{10.2.11}$$

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

$$\Pi_1[S^2] = \{0\}. \tag{10.2.12}$$

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

$$\Pi_m[S^n] = \{0\}, \tag{10.2.13}$$

for $m < n$, whereas

$$\Pi_n[S^n] = \mathbf{Z}. \tag{10.2.14}$$

On the other hand, $\Pi_m[S^n]$ with $m > n$ is not necessarily trivial, for example

$$\Pi_4[S^3] = \mathbf{Z}(2). \tag{10.2.15}$$

Make a table for homotopy groups?

Since the map V of eq.(10.2.10) is topologically trivial, its contribution to the topological charge vanishes. The remaining W term reduces to the $SU(N-1)$ contribution

$$Q = \frac{1}{24\pi^2} \int_{S^3} d^3\sigma_\mu \; \epsilon_{\mu\nu\rho\sigma} \mathrm{Tr} \left[(\tilde{g}\partial_\nu \tilde{g}^\dagger)(\tilde{g}\partial_\rho \tilde{g}^\dagger)(\tilde{g}\partial_\sigma \tilde{g}^\dagger) \right]. \tag{10.2.16}$$

The separation of the V contribution works only if the decomposition of g into V and $\tilde{g}$ is non-singular. In fact, the expression for V is singular for

$g_{11} = -1$. This corresponds to a $((N-1)^2-1)$-dimensional subspace of the (N^2-1)-dimensional $SU(N)$ group space. The map g itself covers a 3-d subspace of $SU(N)$. Hence it is of zero measure to hit a singularity. Since we have now reduced the $SU(N)$ topological charge to the $SU(N-1)$ case, we can go down all the way to $SU(2)$. It remains to be shown that the $SU(2)$ expression is actually an integer. First of all

$$\tilde{g} : S^3 \to SU(2) = S^3, \qquad (10.2.17)$$

and indeed

$$\Pi_3[SU(2)] = \Pi_3[S^3] = \mathbf{Z}. \qquad (10.2.18)$$

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

$$g = \exp(\mathrm{i}\varphi) : S^1 \to U(1) = S^1. \qquad (10.2.19)$$

The relevant homotopy group is

$$\Pi_1[U(1)] = \Pi_1[S^1] = \mathbf{Z}. \qquad (10.2.20)$$

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

$$\begin{aligned} Q &= -\frac{1}{2\pi\mathrm{i}} \int_{S^1} d\sigma_\mu \; \epsilon_{\mu\nu}(g(x)\partial_\nu g(x)^\dagger) = \frac{1}{2\pi} \int_{S^1} d\sigma_\mu \; \epsilon_{\mu\nu}\partial_\nu\varphi(x) \\ &= \frac{1}{2\pi}(\varphi(2\pi) - \varphi(0)). \end{aligned} \qquad (10.2.21)$$

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

$$\begin{aligned} &\tilde{g}(x) = \exp(\mathrm{i}\vec{\alpha}(x) \cdot \vec{\sigma}) = \cos\alpha(x) + \mathrm{i}\sin\alpha(x)\vec{e}_\alpha(x) \cdot \vec{\sigma}, \\ &\vec{e}_\alpha(x) = (\sin\theta(x)\sin\varphi(x), \sin\theta(x)\cos\varphi(x), \cos\theta(x)). \end{aligned} \qquad (10.2.22)$$

It is a good exercise to convince oneself that

$$\begin{aligned}&\epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}\left[(\tilde{g}(x)\partial_\nu\tilde{g}(x)^\dagger)(\tilde{g}(x)\partial_\rho\tilde{g}(x)^\dagger)(\tilde{g}(x)\partial_\sigma\tilde{g}(x)^\dagger)\right]\\&= 12\sin^2\alpha(x)\sin\theta(x)\epsilon_{\mu\nu\rho\sigma}\partial_\nu\alpha(x)\partial_\rho\theta(x)\partial_\sigma\varphi(x).\end{aligned} \tag{10.2.23}$$

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

$$Q = \frac{1}{2\pi^2}\int_{S^3} d^3\sigma_\mu \ \sin^2\alpha(x)\sin\theta(x)\epsilon_{\mu\nu\rho\sigma}\partial_\nu\alpha(x)\partial_\rho\theta(x)\partial_\sigma\varphi(x) = \frac{1}{2\pi^2}\int_{S^3} d\tilde{g} \ . \tag{10.2.24}$$

The volume of the 3-sphere is given by $2\pi^2$. When the map $\tilde{g}$ covers the sphere Q times, the integral gives Q times the volume of S^3. This finally explains why the prefactor $1/32\pi^2$ was introduced in the original expression of eq.(10.2.1) for the topological charge.

10.3 Topology of a Gauge Field on a Compact Manifold

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

$$Q = \int_M d^4x \ P(x) \ . \tag{10.3.1}$$

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

$$P(x) = \partial_\mu \Omega^{(0)}_\mu(x) \ , \tag{10.3.2}$$

and using Gauss' law we obtain

$$Q = \int_M d^4x \ \partial_\mu\Omega^{(0)}_\mu(x) = \int_{\partial M} d^3\sigma_\mu \ \Omega^{(0)}_\mu(x) = 0 \ . \tag{10.3.3}$$

Here we have used that M has no boundary, *i.e.* ∂M is an empty set. A gauge field whose Chern-Pontryagin density can globally be written as a total divergence is indeed topologically trivial on a compact manifold. The important observation is that eq.(10.3.2) may be valid only locally. In other words, gauge singularities may prevent us from using Gauss' law as we did above. In general, it will be impossible to work in a gauge that makes the gauge field non-singular 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 non-trivial gauge transformations, which form a fibre bundle of transition functions. A topologically non-trivial gauge field will contain singularities at some points $x_i \in M$. We cover the manifold M by closed sets c_i such that $x_i \in c_i \backslash \partial c_i$, *i.e.* each singularity lies in the interior of a set c_i. Also $M = \cup_i c_i$ with $c_i \cap c_j = \partial c_i \cap \partial c_j$.

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

$$G^i_\mu(x) = g_i(x)(G_\mu(x) + \partial_\mu)g_i^\dagger(x). \tag{10.3.4}$$

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

$$\begin{aligned} Q &= \sum_i \int_{c_i} d^4x \ P(x) = \sum_i \int_{\partial c_i} d^3\sigma_\mu \ \Omega^{(0)}_\mu(i) \\ &= \frac{1}{2}\sum_{ij} \int_{c_i \cap c_j} d^3\sigma_\mu \ [\Omega^{(0)}_\mu(i) - \Omega^{(0)}_\mu(j)]. \end{aligned} \tag{10.3.5}$$

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

$$G^i_\mu(x) = v_{ij}(x)(G^j_\mu(x) + \partial_\mu)v_{ij}(x)^\dagger, \tag{10.3.6}$$

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

$$v_{ij}(x) = g_i(x)g_j(x)^\dagger. \tag{10.3.7}$$

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

of three regions

$$v_{ik}(x) = v_{ij}(x) v_{jk}(x). \tag{10.3.8}$$

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

$$\Delta\Omega_\mu^{(0)}(i,j) = \Omega_\mu^{(0)}(i) - \Omega_\mu^{(0)}(j). \tag{10.3.9}$$

It is straight forward to show that

$$\begin{aligned}\Delta\Omega_\mu^{(0)}(i,j) = &-\frac{1}{24\pi^2}\epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[v_{ij}(x)\partial_\nu v_{ij}(x)^\dagger v_{ij}(x)\partial_\rho v_{ij}(x)^\dagger v_{ij}(x)\partial_\sigma v_{ij}(x)^\dagger] \\ &-\frac{1}{8\pi^2}\epsilon_{\mu\nu\rho\sigma}\partial_\nu \mathrm{Tr}[\partial_\rho v_{ij}(x)^\dagger v_{ij}(x) G_\sigma^i(x)].\end{aligned} \tag{10.3.10}$$

The above equation for the topological charge then takes the form

$$\begin{aligned} Q &= -\frac{1}{48\pi^2}\sum_{ij}\int_{c_i\cap c_j} d^3\sigma_\mu \ \epsilon_{\mu\nu\rho\sigma} \\ &\times \ \mathrm{Tr}[v_{ij}(x)\partial_\nu v_{ij}(x)^\dagger v_{ij}(x)\partial_\rho v_{ij}(x)^\dagger v_{ij}(x)\partial_\sigma v_{ij}(x)^\dagger] \\ &- \ \frac{1}{16\pi^2}\sum_{ij}\int_{\partial(c_i\cap c_j)} d^2\sigma_{\mu\nu} \ \epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[\partial_\rho v_{ij}(x)^\dagger v_{ij}(x) G_\sigma^i(x)].\end{aligned} \tag{10.3.11}$$

Using the cocycle condition this can be rewritten as

$$\begin{aligned} Q &= -\frac{1}{48\pi^2}\sum_{ij}\int_{c_i\cap c_j} d^3\sigma_\mu \ \epsilon_{\mu\nu\rho\sigma} \\ &\times \ \mathrm{Tr}[v_{ij}(x)\partial_\nu v_{ij}(x)^\dagger v_{ij}(x)\partial_\rho v_{ij}(x)^\dagger v_{ij}(x)\partial_\sigma v_{ij}(x)^\dagger] \\ &-\frac{1}{48\pi^2}\sum_{ijk}\int_{c_i\cap c_j\cap c_k} d^2\sigma_{\mu\nu} \ \epsilon_{\mu\nu\rho\sigma}\mathrm{Tr}[v_{ij}(x)\partial_\rho v_{ij}(x)^\dagger v_{jk}(x)\partial_\rho v_{jk}(x)^\dagger].\end{aligned} \tag{10.3.12}$$

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

In the appropriate mathematical language the gauge transformations g_i form sections of the fibre bundle. Using formula (10.2.8) together with eq.(10.3.7) one can show that the topological charge is expressed in terms of the section in the following way

$$\begin{aligned} Q &= \sum_i Q_i \\ &= \frac{1}{24\pi^2} \sum_i \int_{\partial c_i} d^3\sigma_\mu \epsilon_{\mu\nu\rho\sigma} \mathrm{Tr}[g_i(x)\partial_\nu g_i(x)^\dagger g_i(x)\partial_\rho g_i(x)^\dagger g_i(x)\partial_\sigma g_i(x)^\dagger]. \end{aligned} \tag{10.3.13}$$

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

$$g_i : \partial c_i \to SU(3), \tag{10.3.14}$$

and hence

$$Q_i \in \Pi_3[SU(3)] = \mathbf{Z}. \tag{10.3.15}$$

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

$$G_\mu(x)' = g(x)(G_\mu(x) + \partial_\mu)g(x)^\dagger. \tag{10.3.16}$$

Deriving the gauge transformation properties of the section and using formula (10.2.8) this is again straightforward.

10.4 The Instanton in $SU(2)$

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

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

$$G_\mu(x) = g(x)\partial_\mu g(x)^\dagger. \tag{10.4.1}$$

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

$$g(x) = \frac{x_0 + i\vec{x}\cdot\vec{\sigma}}{|x|}, \ |x| = \sqrt{x_0^2 + |\vec{x}|^2}. \tag{10.4.2}$$

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

$$G_\mu(x) = f(|x|)g(x)\partial_\mu g(x)^\dagger, \tag{10.4.3}$$

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

$$S[G_\mu] = \int d^4x \ \frac{1}{2g^2}\mathrm{Tr}[G_{\mu\nu}(x)G_{\mu\nu}(x)]. \tag{10.4.4}$$

Let us consider the following integral

$$\begin{aligned}
&\int d^4x \ \mathrm{Tr}[(G_{\mu\nu}(x) \pm \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}G_{\rho\sigma}(x))(G_{\mu\nu}(x) \pm \frac{1}{2}\epsilon_{\mu\nu\kappa\lambda}G_{\kappa\lambda}(x)) = \\
&\int d^4x \ \mathrm{Tr}[G_{\mu\nu}(x)G_{\mu\nu}(x) \pm \epsilon_{\mu\nu\rho\sigma}G_{\mu\nu}(x)G_{\rho\sigma}(x) + G_{\mu\nu}(x)G_{\mu\nu}(x)] = \\
&4g_s^2 S[G_\mu] \pm 32\pi^2 Q[G_\mu].
\end{aligned} \tag{10.4.5}$$

We have integrated a square. Hence it is obvious that

$$S[G_\mu] \pm \frac{8\pi^2}{g^2} Q[G_\mu] \geq 0 \;\Rightarrow\; S[G_\mu] \geq \frac{8\pi^2}{g^2} |Q[G_\mu]|, \tag{10.4.6}$$

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

$$S[G_\mu] = \frac{8\pi^2}{g^2} |Q[G_\mu]|. \tag{10.4.7}$$

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

$$G_{\mu\nu}(x) = \pm \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} G_{\rho\sigma}(x). \tag{10.4.8}$$

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

$$f(|x|) = \frac{|x|^2}{|x|^2 + \rho^2} \tag{10.4.9}$$

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

$$G_\mu(x) = \frac{|x|^2}{|x|^2 + \rho^2} g(x) \partial_\mu g(x)^\dagger. \tag{10.4.10}$$

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

10.5 θ-Vacua

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

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

$$G_i(x) = g(x)\partial_i g(x)^\dagger. \tag{10.5.1}$$

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

$$n \in \Pi_3[SU(3)] = \mathbf{Z}, \tag{10.5.2}$$

which is given by

$$n = \frac{1}{24\pi^2}\int_{S^3} d^3x \ \epsilon_{ijk}\mathrm{Tr}[g(x)\partial_i g(x)^\dagger g(x)\partial_j g(x)^\dagger g(x)\partial_k g(x)^\dagger]. \tag{10.5.3}$$

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

$$Q = n - m. \tag{10.5.4}$$

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

$$\langle n|U(\infty,-\infty)|m\rangle = \int \mathcal{D}G_\mu^{(n-m)} \exp(-S[G_\mu]). \tag{10.5.5}$$

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

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

$$T\Psi[G_i] = \Psi[g(G_i + \partial_i)g^\dagger]. \tag{10.5.6}$$

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

$$T|n\rangle = |n+1\rangle, \tag{10.5.7}$$

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

$$T|\theta\rangle = \exp(i\theta)|\theta\rangle. \tag{10.5.8}$$

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

$$|\theta\rangle = \sum_n c_n|n\rangle. \tag{10.5.9}$$

Using

$$\begin{aligned} T|\theta\rangle &= \sum_n c_n T|n\rangle = \sum_n c_n|n+1\rangle \\ &= \sum_n c_{n-1}|n\rangle = \exp(i\theta)\sum_n c_n|n\rangle, \end{aligned} \tag{10.5.10}$$

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

$$|\theta\rangle = \sum_n \exp(-in\theta)|n\rangle. \tag{10.5.11}$$

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

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

$$\begin{aligned}
&\langle\theta|U(\infty,-\infty)|\theta'\rangle = \sum_{m,n} \exp(in\theta)\exp(-im\theta')\langle n|U(\infty,-\infty)|m\rangle \\
&= \sum_{n,Q=n-m} \exp(in\theta - i(n-Q)\theta') \int \mathcal{D}G_\mu^{(Q)} \exp(-S[G_\mu]) \\
&= \delta(\theta-\theta') \sum_Q \int \mathcal{D}G_\mu^{(Q)} \exp(-S[G_\mu]) \exp(i\theta Q[G_\mu]) \\
&= \int \mathcal{D}G_\mu \exp(-S_\theta[G_\mu]).
\end{aligned} \tag{10.5.12}$$

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

$$S_\theta[G_\mu] = S[G_\mu] - i\theta Q[G_\mu]. \tag{10.5.13}$$

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

$$\gamma_\mu(G_\mu(x) + \partial_\mu)\Psi(x) = 0. \tag{10.5.14}$$

These eigenvectors have a definite handedness, *i.e.*

$$\frac{1}{2}(1 \pm \gamma_5)\Psi(x) = \Psi(x), \tag{10.5.15}$$

because

$$\gamma_5\gamma_\mu(G_\mu(x) + \partial_\mu)\Psi(x) = -\gamma_\mu(G_\mu(x) + \partial_\mu)\gamma_5\Psi(x) = 0. \qquad (10.5.16)$$

The Atiyah-Singer index theorem states that

$$Q = n_L - n_R, \qquad (10.5.17)$$

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

10.6 The $U(1)$-Problem

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

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

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

Bell-Jackiw anomaly

$$\partial_\mu J^5_\mu(x) = 2N_f P(x), \tag{10.6.1}$$

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

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

$$g^2 = g_s^2 N_c, \tag{10.6.2}$$

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

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

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

$$S[G_\mu] \geq \frac{8\pi^2}{g_s^2}|Q[G_\mu]| = \frac{8\pi^2 N_c}{g^2}|Q[G_\mu]|, \tag{10.6.3}$$

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

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

$$\chi_t = \int d^4x \ {}_{pg}\langle 0|P(0)P(x)|0\rangle_{pg} = \frac{\langle Q^2\rangle}{V} \tag{10.6.4}$$

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 non-trivial field configurations. Therefore in full QCD (with massless quarks)

$$\int d^4x \ \langle 0|P(0)P(x)|0\rangle = 0, \qquad (10.6.5)$$

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

$$\chi_t - \sum_m \frac{\langle 0|P|m\rangle\langle m|P|0\rangle}{M_m^2} = 0, \qquad (10.6.6)$$

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

$$\chi_t = \frac{|\langle 0|P|\eta'\rangle|^2}{M_{\eta'}^2}. \qquad (10.6.7)$$

Using the anomaly equation one obtains

$$\langle 0|P|\eta'\rangle = \frac{1}{2N_f}\langle 0|\partial_\mu A_\mu|\eta'\rangle = \frac{1}{\sqrt{2N_f}} M_{\eta'}^2 f_{\eta'}. \qquad (10.6.8)$$

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

$$\chi_t = \frac{f_\pi^2 M_{\eta'}^2}{2N_f}. \qquad (10.6.9)$$

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 Nambu-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-Nambu-Goldstone boson mass squared is hence proportional to $1/N_c$. So far we have assumed that all quarks are massless. When a non-zero s quark mass is taken into account, the formula changes to

$$\chi_t = \frac{1}{6} f_\pi^2 (M_{\eta'}^2 + M_\eta^2 - 2M_K^2) = (0.180\text{GeV})^4. \tag{10.6.10}$$

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

10.7 Baryon Number Violation in the Standard Model

The classical Lagrange density of the standard model does not contain baryon number violating interactions. However, this does not imply that the standard model conserves baryon number after quantization. Indeed, due to the chiral couplings of the fermions, the baryon number current has an anomaly in the standard model. Although the Lagrange density has a global $U(1)$ baryon number symmetry, this symmetry is explicitly broken in the quantum theory. The same is true for lepton number. The difference, $B - L$, on the other hand, remains conserved. The existence of baryon number violating processes at the electroweak scale may change the baryon asymmetry that has been generated at the GUT scale.

Let us consider the vacuum structure of a non-Abelian gauge theory (like the $SU(2)$ sector of the standard model). A classical vacuum solution is

$$\Phi(\vec{x}) = \begin{pmatrix} \Phi_+(\vec{x}) \\ \Phi_0(\vec{x}) \end{pmatrix} = \begin{pmatrix} 0 \\ v \end{pmatrix}, \ A_i(\vec{x}) = 0. \tag{10.7.1}$$

Of course, gauge transformations of this solution are also vacua. However, states that are related by a gauge transformation are physically equivalent, and one should not consider the other solutions as additional vacua. Still, there is a subtlety, because there are gauge transformations with different topological properties. First of all, there are the so-called small gauge transformations, which can be continuously deformed into the identity, and one should indeed not distinguish between states related by small gauge transformations. However, there are also large gauge transformations — those that can not be deformed into a trivial gauge transformation — and they indeed give rise to additional vacuum states. The gauge transformations

$$g : \mathbf{R}^3 \to SU(2) \tag{10.7.2}$$

can be viewed as maps from coordinate space into the group space. When one identifies points at spatial infinity $\mathbf{R}^3$ is compactified to S^3. On the other hand, the group space of $SU(2)$ is also S^3. Hence, the gauge transformations are maps

$$g : S^3 \to S^3. \tag{10.7.3}$$

Such maps are known to fall into topologically distinct classes characterized by a winding number

$$n[g] \in \Pi_3[SU(2)] = \mathbf{Z} \tag{10.7.4}$$

from the third homotopy group of the gauge group. In this case, maps with any integer winding number are possible. Denoting a map with winding number n by g_n we can thus construct a set of topologically inequivalent vacuum states

$$\Phi^{(n)}(\vec{x}) = g_n(\vec{x}) \begin{pmatrix} 0 \\ v \end{pmatrix}, \; A_i^{(n)}(\vec{x}) = g_n(\vec{x}) \partial_i g_n(\vec{x})^\dagger. \tag{10.7.5}$$

Topologically distinct vacua are separated by energy barriers, and thus there is a periodic potential in the space of field configurations.

Classically, the system is in one of the degenerate vacuum states. Quantum mechanically, however, the system can tunnel from one vacuum to another. It turns out that a transition from the vacuum (m) to the vacuum (n) is accompanied by a baryon number violating process of strength $\Delta B = N_g(n-m)$, where N_g is the number of generations of quarks and leptons. Also the lepton number changes by $\Delta L = N_g(n-m)$, such that $B - L$ is conserved. The tunnel amplitude — and hence the rate of baryon number violating processes — is controlled by the barrier height between adjacent classical vacua. The unstable field configuration at the to of the barrier is known as a sphaleron (meaning ready to decay). In the standard model the height of the barrier (the sphaleron energy) is given by $4\pi v/g$ and the resulting tunneling rate is

$$\exp(-\frac{8\pi^2}{g^2}) \approx \exp(-200), \tag{10.7.6}$$

which is totally negligible. Hence, for some time people assumed that baryon number violation in the standard model is only of academic interest. However, it was overlooked that in the early Universe one need not tunnel through the barrier — one can simply step over it classically due to large thermal fluctuations. Then one must assume that in the TeV range baryon number violating processes are un-suppressed in the standard model. This means that any pre-existing baryon asymmetry — carefully created at the GUT scale — will be washed out, because baryon number violating processes are again in thermal equilibrium. Since the electroweak phase transition is of second or of weakly first order, it is unlikely (but not excluded) that a sufficient baryon asymmetry is re-generated at the electroweak scale.

However, we should not forget that $B - L$ is conserved in the standard model. This means that this mode is not thermalized. When baryon and lepton asymmetries ΔB_i and ΔL_i have been initially generated at the GUT scale, equilibrium sphaleron processes will imply that finally

$$\Delta(B_f + L_f) = 0, \tag{10.7.7}$$

but still

$$\Delta(B_f - L_f) = \Delta(B_i - L_i) = 0. \tag{10.7.8}$$

Hence, the present baryon and lepton asymmetries then are

$$\Delta B_f = -\Delta L_f = \frac{1}{2}\Delta(B_i - L_i). \tag{10.7.9}$$

This again leads to a problem, because also the minimal $SU(5)$ model conserves $B - L$. An asymmetry $\Delta(B_i - L_i)$ must hence be due to processes in the even earlier Universe. Then we would know as much as before. Fortunately, there is a way out. Other GUTs like $SO(10)$ and E_6 are not ruled out via proton decay and indeed do not conserve $B - L$. The reason for $B - L$ violation in these models is related to the existence of massive neutrinos. The so-called "see-saw" mechanism gives rise to one heavy neutrino of mass 10^{14} GeV and one light neutrino of mass in the eV range, that is identified with the neutrinos that we observe. Hence, we can explain the baryon asymmetry using GUTs only if the neutrinos are massive. Otherwise, we must assume that it was generated at times before 10^{-34} sec after the Big Bang, or we must find a way to go sufficiently out of thermal equilibrium around the electroweak phase transition and generate the baryon asymmetry via sphaleron processes.

Chapter 11

The Strong CP-Problem

We have seen that non-Abelian $SU(N)$ gauge fields have nontrivial topological structure. In particular, classical vacuum (pure gauge) field configurations are characterized by an integer winding number from the homotopy group $\Pi_3[SU(N)] = \mathbf{Z}$. Instantons are examples of Euclidean field configurations with topological charge Q that describe tunneling between topologically distinct classical vacua. Due to tunneling, the quantum vacuum is a linear superposition of classical vacua characterized by a vacuum angle $\theta \in [-\pi, \pi]$. In the Euclidean action the vacuum angle manifests itself as an additional term $i\theta Q$. For $\theta \neq 0, \pi$ this term explicitly breaks the CP symmetry. As a consequence, the neutron would have an electric dipole moment proportional to θ, while without CP violation the dipole moment vanishes. Indeed, the observed electric dipole moment of the neutron is indistinguishable from zero. This puts a stringent bound on the vacuum angle $|\theta| < 10^{-9}$. The question arises why in Nature $\theta = 0$ to such a high accuracy. This is the strong CP-problem.

Within QCD itself, one could "solve" the strong CP-problem simply by demanding CP symmetry. In the standard model, however, the Yukawa couplings already lead to CP violation which is indeed observed in the neutral kaon system. This effect is rather subtle and requires the presence of at least three generations. If there were CP violation in the strong interactions, it would give rise to much more drastic effects. Naively, one might hope to solve the strong CP-problem by the assumption that gluon fields with $Q \neq 0$ are very much suppressed. However, this probably does not work. First of all, the quantitative solution of the $U(1)$-problem relies on the fact that gluon fields with $Q \neq 0$ appear frequently in the pure gauge theory. Of course, this need not necessarily be the case in full QCD with quarks. Indeed, if the up quark would be massless, the Atiyah-Singer index theorem would imply that fermionic zero-modes of the Dirac operator completely eliminate gluon fields with $Q \neq 0$ from the path integral. In that case, the θ-vacuum term would vanish and all θ-vacua would be physically equivalent and thus CP conserving. It is a controversial issue if the up quark might indeed be massless, but most experts of chiral perturbation theory believe that this possibility is excluded. In any case, if the up quark would indeed be massless, and we would solve the CP problem in that way, we would immediately face the m_u-problem: why is the up quark massless?

We have seen already that the Chern-Pontryagin topological charge density is intimately connected with the divergence of the flavor-singlet axial current. This implies that the vacuum angle can be rotated using an axial $U(1)$ transformation. In this way, one can indeed get rid of any hypothetical θ'-angle in the electroweak $SU(2)_L$ gauge field. The strong $SU(3)_c$ θ-vacuum angle, on the other hand, cannot be rotated away in this fashion, because it just gets transformed into a complex phase of the determinant of the quark mass matrix. Still, such a transformation can be quite useful, for example, because we can then investigate the θ-vacuum dynamics using chiral Lagrangians. For example, for unequal up and down quark masses, one finds a phase transition at $\theta = \pi$ at which CP is spontaneously broken. Hence, despite the fact that $\theta = \pi$ does not break CP explicitly, the CP symmetry is now broken dynamically. This means that θ cannot be π in Nature and must indeed be zero.

The chiral Lagrangian method also allows us to study θ-vacuum effects in the large N_c limit. In this limit, the axial $U(1)$ anomaly vanishes and the η'-meson becomes a massless Goldstone boson. In fact, the η'-meson couples to the complex phase of the quark mass matrix — and hence to θ — and can indeed be used to rotate θ away. Hence, there is no strong CP-problem at $N_c = \infty$. Of course, we know that in our world $N_c = 3$ (although some of the textbook arguments for this (anomaly cancellation, π^0 decay) are incorrect), and we indeed face the strong CP-problem.

A very appealing solution of the strong CP-problem was suggested by

Peccei and Quinn. They suggested an extension of the standard model with two Higgs doublets. This situation also naturally arises in supersymmetric extensions of the standard model. As a consequence of the presence of the second Higgs field, there is an extra $U(1)_{PQ}$ — a so-called Peccei-Quinn symmetry — which allows one to rotate θ away even at finite N_c. When $SU(2)_L \otimes U(1)_Y$ breaks down to $U(1)_{em}$, the Peccei-Quinn $U(1)_{PQ}$ symmetry also gets spontaneously broken. It was first pointed out by Weinberg and Wilczek that this leads to a new pseudo-Goldstone boson — the axion. Unfortunately, so far nobody has ever detected an axion despite numerous experimental efforts and it is still unclear if this is indeed the correct solution of the strong CP problem. Although the original Peccei-Quinn model was soon ruled out by experiments, the symmetry breaking scale of the model can be shifted to higher energy scales making the axion more or less invisible.

Axions are very interesting players in the Universe. They couple only weakly to ordinary matter, but they still have interesting effects. First of all, they are massive and could provide enough energy to close the Universe. If it exists, the axion can also shorten the life-time of stars. Stars live so long, because they cannot get rid of their energy by radiation very fast. For example, a photon that is generated in a nuclear reaction in the center of the sun spends 10^7 years before it reaches the sun's surface, simply because its electromagnetic cross section with the charged matter in the sun is large. An axion, on the other hand, interacts weakly and can thus get out much father. Like neutrinos, axions can therefore act as a super coolant for stars. The observed life-time of stars can thus be used to put astrophysical limits on axion parameters like the axion mass. Axions can be generated in the early Universe in multiple ways. First, they can simply be thermally produced. Then they can be generated by a disoriented $U(1)_{PQ}$ condensate. This mechanism is similar to the recently discussed pion production via a disoriented chiral condensate in a heavy ion collision generating a quark-gluon plasma. Also, the spontaneous breakdown of a $U(1)$ symmetry is accompanied by the generation of cosmic strings. Indeed, if the axion exists, axionic cosmic strings should exist as well. A network of such fluctuating strings could radiate energy by emitting the corresponding Goldstone bosons, namely axions.

11.1 Rotating θ into the Mass Matrix

Let us assume that there is a θ-vacuum term $i\theta Q$ in the Euclidean action of QCD. We have seen already that such a term is intimately connected with the flavor-singlet axial $U(1)$ symmetry. Indeed, due to the axial anomaly, the fermionic measure is not invariant under axial $U(1)$ transformations. Let us discuss this in the theory with N_f quark flavors. Under an axial $U(1)$ transformation

$$q_L' = \exp(-i\theta/2N_f)q_L, \ q_R' = \exp(i\theta/2N_f)q_R, \tag{11.1.1}$$

the fermion determinant in the background of a gluon field with topological charge Q is not invariant. In fact, it changes by $\exp(i\theta Q)$. Hence, the above axial transformation can be used to cancel any pre-existing θ-vacuum term in the QCD action. Of course, the transformation must be applied consistently everywhere. It cancels out in the quark-gluon gauge interactions which are chirally invariant, but not in mass terms. In fact, the mass matrix $\mathcal{M} = \mathrm{diag}(m_u, m_d, ..., m_{N_f})$ now turns into

$$\mathcal{M}' = \mathrm{diag}(m_u \exp(i\theta/N_f), m_d \exp(i\theta/N_f), ..., m_{N_f} \exp(i\theta/N_f)), \tag{11.1.2}$$

i.e. θ turns into the complex phase of the determinant of the quark mass matrix. If one of the quarks is massless, the determinant vanishes and its phase becomes physically irrelevant. Interestingly, strong CP violation manifests itself by a complex phase in the quark mass matrix, while the CP violation due to the Yukawa couplings leads to the complex phase in the Cabibbo-Kobayashi-Maskawa quark mixing matrix.

The strong interaction θ-angle cannot be completely rotated away, because both left- and right-handed quarks are coupled to the gluons. A potential electroweak interaction θ'-angle, on the other hand, can simply be rotated away, because only the left-handed fermions couple to the $SU(2)_L$ gauge field. For example, in order to remove a θ'-angle, one just performs a left-handed $U(1)$ transformation

$$\left(\begin{array}{c} \nu_{eL}' \\ e_L' \end{array} \right) = \exp(-i\theta') \left(\begin{array}{c} \nu_{eL} \\ e_L \end{array} \right). \tag{11.1.3}$$

The change of the fermion measure under the transformation cancels against the θ'-term, the gauge interactions remain unchanged, but the mass terms are again affected. However, we can now simply rotate the right-handed fields as well

$$\nu'_{eR} = \exp(-i\theta')\nu_{eR}, \ e'_R = \exp(-i\theta')e_R, \qquad (11.1.4)$$

which then leaves the mass term invariant. Unlike in the QCD case, this does not regenerate the θ'-term because the right-handed fermions do not couple to the $SU(2)_L$ gauge field.

11.2 The θ-Angle in Chiral Perturbation Theory

Let us now discuss how the vacuum angle affects the low-energy QCD dynamics. Since we know how the quark mass matrix enters the chiral Lagrangian, and since θ is just the complex phase in that matrix, it is clear how to include θ in chiral perturbation theory. To lowest order the chiral perturbation theory action then takes the form

$$S[U] = \int d^4x \ \{\frac{F_\pi^2}{4}\mathrm{Tr}[\partial^\mu U^\dagger \partial_\mu U] + \frac{1}{2N_f}\langle\bar{\Psi}\Psi\rangle \mathrm{Tr}[\mathcal{M}'U^\dagger + U\mathcal{M}'^\dagger]\}, \ (11.2.1)$$

where $\mathcal{M}'$ is the θ-dependent quark mass matrix of eq.(11.1.2). The above action is not 2π-periodic in θ. Instead, it is only $2\pi N_f$-periodic. Still, it is easy to show that the resulting path integral is indeed 2π-periodic. The situation in QCD itself is similar. While the contribution $i\theta Q$ to the action itself is not periodic in θ, it enters the path integral through the 2π-periodic Boltzmann factor $\exp(i\theta Q)$. Hence, the path integral is periodic while the action itself is not. Let us check that a non-zero vacuum angle indeed breaks CP. On the level of the chiral Lagrangian charge conjugation corresponds to ${}^C U = U^T$, while parity corresponds to ${}^P U(\vec{x}, t) = U(-\vec{x}, t)^\dagger$. The action from above breaks P while it leaves C invariant, and hence it indeed violates CP.

Let us now examine the effect of θ on the vacuum of the pion theory in the $N_f = 2$ case. Then the mass matrix takes the form

$$\mathcal{M}' = \mathrm{diag}(m_u \exp(i\theta/2), m_d \exp(i\theta/2)). \qquad (11.2.2)$$

In order to find the vacuum configuration, we must minimize the potential energy, and hence we must maximize

$$\mathrm{Tr}[\mathcal{M}'U^\dagger + U\mathcal{M}'^\dagger] = m_u \cos(\frac{\theta}{2} + \varphi) + m_d \cos(\frac{\theta}{2} - \varphi). \tag{11.2.3}$$

Here we have parametrized $U = \mathrm{diag}(\exp(i\varphi), \exp(-i\varphi))$. The minimum energy configuration has

$$\tan\varphi = \frac{m_d - m_u}{m_u + m_d} \tan\frac{\theta}{2}. \tag{11.2.4}$$

As expected, for $\theta = 0$ one obtains $\varphi = 0$ and hence $U = \mathbf{1}$. It is interesting that the θ-angle affects the pion vacuum only for nondegenerate quark masses. At $\theta = \pm\pi$ the pion vacuum configuration has $\varphi = \pm\pi/2$, i.e. $U = \pm\mathrm{diag}(i, -i)$. These two vacua are both not CP invariant. Instead, they are CP images of one another. This indicates that, at $\theta = \pm\pi$, the CP symmetry is spontaneously broken. Hence, despite the fact that for $\theta = \pm\pi$ there is no explicit CP violation, the symmetry is still not intact. This means that in Nature we have $\theta = 0$, not $\theta = \pi$.

11.3 The θ-Angle at Large N_c

We have seen that the $U(1)$ problem can be understood quantitatively in the limit of many colors N_c. At $N_c = \infty$ the anomalous axial $U(1)$ symmetry is restored and the η'-meson becomes a Goldstone boson. At large but finite N_c the η'-meson is a pseudo-Goldstone boson with a mass

$$M_{\eta'}^2 = \frac{N_f \chi_t}{F_\pi^2}, \tag{11.3.1}$$

proportional to $1/N_c$ (note that F_π^2 is of order N_c). Here $\chi_t = \langle Q^2 \rangle / V$ is the topological susceptibility of the pure gauge theory which is of order one in the large N_c limit.

Since for large N_c the η'-meson becomes light, it must be included in the low-energy chiral Lagrangian. Since the axial $U(1)$ symmetry is restored at $N_c = \infty$ and is then spontaneously broken, the chiral symmetry is now $U(N_f)_L \otimes U(N_f)_R$ broken to $U(N_f)_{L=R}$. Consequently, the Goldstone bosons now live in the coset space $U(N_f)_L \otimes U(N_f)_R/U(N_f)_{L=R} = U(N_f)$. Hence, now there are N_f^2 Goldstone bosons. The additional η' Goldstone boson is described by the complex phase of the determinant of a unitary matrix $\tilde{U}$, which would have determinant one if the η'-meson were heavy. For large N_c the chiral perturbation theory action takes the form

$$\begin{aligned} S[U] &= \int d^4x \, \{ \frac{F_\pi^2}{4} \mathrm{Tr}[\partial^\mu \tilde{U}^\dagger \partial_\mu \tilde{U}] + \frac{1}{2N_f} \langle \bar{\Psi}\Psi \rangle \mathrm{Tr}[\mathcal{M}' \tilde{U}^\dagger + \tilde{U} \mathcal{M}'^\dagger] \\ &+ N_f \chi_t (i \log \det \tilde{U})^2 \}. \end{aligned} \tag{11.3.2}$$

If there is a θ-angle in the quark mass matrix, this angle can now be absorbed into the η'-meson field, i.e. in the complex phase of the determinant of the Goldstone boson field $\tilde{U}$. Then the action turns into

$$\begin{aligned} S[U] &= \int d^4x \, \{ \frac{F_\pi^2}{4} \mathrm{Tr}[\partial^\mu \tilde{U}^\dagger \partial_\mu \tilde{U}] + \frac{1}{2N_f} \langle \bar{\Psi}\Psi \rangle \mathrm{Tr}[\mathcal{M} \tilde{U}^\dagger + \tilde{U} \mathcal{M}^\dagger] \\ &+ N_f \chi_t (i \log \det \tilde{U} - \theta)^2 \}. \end{aligned} \tag{11.3.3}$$

In the large N_c limit the last term which is of order one can be neglected compared to the other terms which are of order N_c. Hence, at $N_c = \infty$, the vacuum angle drops out of the theory, and all θ-vacua become physically equivalent. Hence, for infinitely many colors there is no CP problem. Essentially, the restored $U(1)$ symmetry then allows us to rotate θ away, despite the fact that it is still explicitly broken by the quark masses.

11.4 The Peccei-Quinn Symmetry

At finite N_c, the axial $U(1)$ symmetry is inevitably broken by the anomaly. Hence, we will not be able to rotate θ away using that symmetry. The idea of Peccei and Quinn was to introduce another $U(1)_{PQ}$ symmetry — now known as a Peccei-Quinn symmetry — that will allow us to get rid of θ despite the fact that the axial $U(1)$ symmetry is explicitly anomalously broken. We will discuss the Peccei-Quinn symmetry in the context of the single generation standard model. The generalization to more generations is straightforward. Of course, it should be noted that with less than three generations, there is no CP violating phase in the quark mixing matrix and θ would be the only source of CP violation. Let us first remind ourselves how the up and down quarks get their masses in the standard model. As we have seen earlier, the mass of the down quark $m_d = f_d v$ is due to the Yukawa coupling

$$\mathcal{L}(u_L, d_L, d_R, \Phi) = f_d[\bar{d}_R(\Phi_+^* \Phi_0^*) \left(\begin{array}{c} u_L \\ d_L \end{array} \right) + (\bar{u}_L \bar{d}_L) \left(\begin{array}{c} \Phi_+ \\ \Phi_0 \end{array} \right) d_R], \quad (11.4.1)$$

while the mass of the up quark $m_u = f_u v$ is due to the term

$$\mathcal{L}(u_L, d_L, d_R, \Phi) = f_d[\bar{u}_R(\Phi_0'^* \Phi_-'^*) \left(\begin{array}{c} u_L \\ d_L \end{array} \right) + (\bar{u}_L \bar{d}_L) \left(\begin{array}{c} \Phi_0' \\ \Phi_-' \end{array} \right) u_R]. \quad (11.4.2)$$

In the standard model the Higgs field Φ' is constructed out of the Higgs field Φ as

$$\Phi' = \left(\begin{array}{c} \Phi_0' \\ \Phi_-' \end{array} \right) = \left(\begin{array}{c} \Phi_0^* \\ -\Phi_+^* \end{array} \right). \quad (11.4.3)$$

When we write the Higgs field as a matrix

$$\Phi = \left(\begin{array}{cc} \Phi_0^* & \Phi_+ \\ -\Phi_+^* & \Phi_0 \end{array} \right), \quad (11.4.4)$$

both Yukawa couplings can be combined into one expression

$$\mathcal{L}(u_L, d_L, u_R, d_R, \Phi) = (\bar{u}_R \bar{d}_R) \mathcal{F}^\dagger \Phi^\dagger \left(\begin{array}{c} u_L \\ d_L \end{array} \right) + (\bar{u}_L \bar{d}_L) \Phi \mathcal{F} \left(\begin{array}{c} u_R \\ d_R \end{array} \right), \quad (11.4.5)$$

where $\mathcal{F} = \mathrm{diag}(f_u, f_d)$ is the diagonal matrix of Yukawa couplings. When a θ-term is present in the QCD Lagrangian, it can be rotated into the matrix of Yukawa couplings by the transformation

$$\left(\begin{array}{c} u_L' \\ d_L' \end{array} \right) = \exp(-i\theta/4) \left(\begin{array}{c} u_L \\ d_L \end{array} \right), \ \left(\begin{array}{c} u_R' \\ d_R' \end{array} \right) = \exp(-i\theta/4) \left(\begin{array}{c} u_R \\ d_R \end{array} \right). \quad (11.4.6)$$

This turns the matrix of Yukawa couplings into

$$\mathcal{F}' = \mathrm{diag}(f_u \exp(i\theta/2), f_d \exp(i\theta/2)). \quad (11.4.7)$$

Since the Higgs field matrix Φ is proportional to an $SU(2)$ matrix, the complex phase $\exp(i\theta)$ of the matrix of Yukawa couplings cannot be absorbed into it, and hence θ cannot be rotated away. Here we have assumed that f_u and f_d are real. Otherwise, the effective vacuum angle would still be the complex phase of the determinant of $\mathcal{F}'$.

It is instructive to include the Yukawa couplings in the chiral perturbation theory action

$$S[U,\Phi] = \int d^4x \ \{\frac{F_\pi^2}{4}\mathrm{Tr}[\partial^\mu U^\dagger \partial_\mu U] + \frac{1}{2N_f}\langle\bar{\Psi}\Psi\rangle\mathrm{Tr}[\Phi\mathcal{F}'U^\dagger + U\mathcal{F}'^\dagger\Phi^\dagger]\}. \tag{11.4.8}$$

Again, the complex phase in $\mathcal{F}'$ cannot be absorbed into the Higgs field matrix Φ because it is proportional to an $SU(2)$ matrix. The Goldstone boson matrix U is also an $SU(2)$ matrix, and hence θ cannot be rotated away. As we have seen, θ can actually be rotated away if the Goldstone boson matrix is in $U(2)$ and contains the η'-meson field as a complex phase of its determinant. This, however, is the case only at large N_c.

The basic idea of Peccei and Quinn can be boiled down to extending the standard model Higgs field to a matrix proportional to $U(2)$ — not just to $SU(2)$. The extra $U(1)_{PQ}$ Peccei-Quinn symmetry then allows us to rotate θ away. The actual proposal of Peccei and Quinn does a bit more. It introduces two completely independent Higgs doublets Φ and Φ' which can be combined to form a $GL(2,\mathbb{C})$ matrix. Working with $GL(2,\mathbb{C})$ rather than with $U(2)$ matrices ensures that the Higgs sector is described by a perturbatively renormalizable linear σ-model, instead of a perturbatively nonrenormalizable nonlinear σ-model. Still, this is not too relevant since, as we have discussed earlier, both the linear and the nonlinear σ-model are trivial in the continuum limit, and physically equivalent. For simplicity, we will not follow Peccei and Quinn all the way and introduce two Higgs fields. Instead we will just extend the standard Higgs field to a matrix $\tilde{\Phi}$ proportional to a $U(2)$ matrix. This means that we introduce just one additional degree of freedom, while Peccei and Quinn introduced four. The complex phase in $\mathcal{F}'$ can then be absorbed in a redefinition of $\tilde{\Phi}$ and one obtains

$$S[U,\tilde{\Phi}] = \int d^4x \ \{\frac{F_\pi^2}{4}\mathrm{Tr}[\partial^\mu U^\dagger \partial_\mu U] + \frac{1}{2N_f}\langle\bar{\Psi}\Psi\rangle\mathrm{Tr}[\tilde{\Phi}\mathcal{F}U^\dagger + U\mathcal{F}^\dagger\tilde{\Phi}^\dagger]\}. \tag{11.4.9}$$

Since this expression now contains the original real Yukawa coupling matrix $\mathcal{F}$, all signs of the vacuum angle have completely disappeared from the theory. Instead the complex phase $\exp(ia/v)$ of $\tilde{\Phi}$ now plays the role of θ. In particular, the axion field $a(x)/v$ behaves like a space-time dependent θ-vacuum angle.

11.5 $U(1)_{PQ}$ Breaking and the Axion

The scalar potential $V(\tilde{\Phi})$ in the extension of the standard model is invariant against $SU(2)_L \otimes SU(2)_R \otimes U(1)_{PQ}$ transformations. In the vacuum the Higgs field takes the value $\tilde{\Phi} = \mathrm{diag}(v, v)$, which breaks this symmetry down spontaneously to $SU(2)_{L=R}$. Hence, there are $3 + 3 + 1 - 3 = 4$ massless Goldstone bosons. As usual, we then gauge $SU(2)_L$ as well as the $U(1)_Y$ subgroup of $SU(2)_R$, which amounts to a partial explicit breaking of $SU(2)_R$. The unbroken subgroup of $SU(2)_{L=R}$ then is just $U(1)_{em}$. Via the Higgs mechanism, three of the four Goldstone bosons are eaten by the gauge bosons and become the longitudinal components of Z_0 and $W^\pm$. Since $U(1)_{PQ}$ remains a global symmetry, the fourth Goldstone boson does not get eaten. This Goldstone boson is the axion.

Let us construct a low-energy effective theory that contains all Goldstone bosons of the extended standard model, namely the pions and the axion. This is easy to do, because we have already included $\tilde{\Phi}$ in the chiral Lagrangian. After spontaneous symmetry breaking at the electroweak scale v, we can write

$$\tilde{\Phi} = v\mathrm{diag}(\exp(ia/v), \exp(ia/v)), \tag{11.5.1}$$

where a parametrizes the axion field. Similarly, we can write

$$U = \mathrm{diag}(\exp(i\pi^0/F_\pi), \exp(-i\pi^0/F_\pi)). \tag{11.5.2}$$

Of course, the field U also contains the charged pions. At this point, we are interested in axion-pion mixing. Since the axion is electrically neutral, it cannot mix with the charged pions and we thus ignore them. Let us first search for the vacuum of the axion-pion system. Minimizing the energy implies maximizing

$$\mathrm{Tr}[\tilde{\Phi}\mathcal{F}U^\dagger + U\mathcal{F}^\dagger\tilde{\Phi}^\dagger] = m_u \cos(a/v + \pi^0/F_\pi) + m_d \cos(a/v - \pi^0/F_\pi). \tag{11.5.3}$$

Obviously, this expression is maximized for $a = \pi^0 = 0$. Next we expand around this vacuum to second order in the fields. The resulting mass squared matrix takes the form

$$M^2 = \frac{\langle\bar{\Psi}\Psi\rangle}{4} \left(\begin{array}{cc} (m_u + m_d)/F_\pi^2 & (m_u - m_d)/F_\pi v \\ (m_u - m_d)/F_\pi v & (m_u + m_d)/v^2 \end{array} \right). \tag{11.5.4}$$

In the limit $v \to \infty$ this matrix turns into

$$M^2 = \frac{\langle \bar{\Psi}\Psi \rangle}{4} \left(\begin{array}{cc} (m_u + m_d)/F_\pi^2 & 0 \\ 0 & 0 \end{array} \right), \tag{11.5.5}$$

from which we read off the familiar mass squared of the pion

$$M_\pi^2 = \frac{\langle \bar{\Psi}\Psi \rangle (m_u + m_d)}{4F_\pi^2}. \tag{11.5.6}$$

In this limit the axion remains massless and there is no axion-pion mixing. Next, we keep v finite, but we still use $v \gg F_\pi$. Then there is a small amount of mixing between the axion and the pion, but the pion mass is to leading order unaffected. The determinant of the mass squared matrix is given by

$$\frac{\langle \bar{\Psi}\Psi \rangle^2}{16} [\frac{(m_u + m_d)^2}{F_\pi^2 v^2} - \frac{(m_u - m_d)^2}{F_\pi^2 v^2}] = \frac{4 m_u m_d}{F_\pi^2 v^2} = M_\pi^2 M_a^2. \tag{11.5.7}$$

Hence, the axion mass squared is given by

$$M_a^2 = \frac{\langle \bar{\Psi}\Psi \rangle m_u m_d}{(m_u + m_d) v^2}. \tag{11.5.8}$$

It vanishes in the chiral limit, and even if just one of the quark masses is zero. The ratio of the axion and pion mass squares is

$$\frac{M_a^2}{M_\pi^2} = \frac{4 m_u m_d F_\pi^2}{(m_u + m_d)^2 v^2}. \tag{11.5.9}$$

Hence, for $m_u = m_d$ we have

$$\frac{M_a}{M_\pi} = \frac{F_\pi}{v} \approx \frac{250\text{GeV}}{0.1\text{GeV}} = 2500 \ \Rightarrow \ M_a \approx \frac{0.14\text{GeV}}{2500} \approx 50\text{keV}. \tag{11.5.10}$$

This is an unusually light particle that should have observable effects. Indeed, there have been several experimental searches for this "standard" axion, but they did not find anything. The only exception was an experiment performed in Aachen (Germany). The signature they saw was called the "Aachion" but it was not confirmed by other experiments, and the standard axion with a mass around 50keV has actually been ruled out. Still, by pushing the $U(1)_{PQ}$ breaking scale far above the electroweak scale one can make the axion more weakly coupled and thus make it invisible to all experiments performed so far. Presently, there are still searches going on that attempt to detect the "invisible" axion. Before they are successful, we cannot be sure that Peccei and Quinn's elegant solution of the strong CP problem is actually correct.

Invisible axions are light and interact only weakly, because the axion coupling constants are proportional to the mass. Still, unless they are too light — and thus too weakly interacting — axions can cool stars very efficiently. Their low interaction cross section allows them to carry away energy more easily than the more strongly coupled photon. A sufficiently interacting axion could shorten the life-time of stars by a substantial amount. From the observed life-time one can hence infer an upper limit on the axion mass. In this way invisible axions heavier than 1 eV have been ruled out. This implies that the Peccei-Quinn symmetry breaking scale must be above 10^7 GeV. If they exist, axions would also affect the cooling of a neutron star that forms after a supernova explosion. There would be less energy taken away by neutrinos. The observed neutrino burst of the supernova SN 1987A would have consisted of fewer neutrinos if axions had also cooled the neutron star. This astrophysical observation excludes axions of masses between 10^{-3} and 0.02 MeV — a range that cannot be investigated in the laboratory.

There are various mechanisms in the early Universe that can lead to the generation of axions. The simplest is via thermal excitation. One can estimate that thermally generated axions must be rather heavy in order to contribute substantially to the energy density of the Universe. In fact, thermal axions cannot close the Universe, because the required mass is already ruled out by the astrophysical limits. Another interesting mechanism for axion production relies on the fact that the axion potential is not completely flat but has a unique minimum. At high temperatures the small axion mass is irrelevant, the potential is practically flat and corresponds to a family of degenerate minima related to one another by $U(1)_{PQ}$ symmetry transformations. Hence, there are several degenerate vacua labeled by different values of the axion field (hence with different values of θ) and all values of θ are equally probable. Then different regions of the hot early Universe must have been in different θ-vacua. When the temperature decreases, $a = 0$ is singled out as the unique minimum. In order to minimize its energy, the scalar field then "rolls" down to this minimum, and oscillates about it. The oscillations are damped by axion emission, and finally $a = 0$

is reached everywhere in the Universe. The axions produced in this way would form a Bose condensate that could close the Universe for an axion mass in the 10^{-5} eV range. This makes the axion an attractive candidate for dark matter in the Universe. This axion production mechanism via a disoriented Peccei-Quinn condensate is very similar to the pion-production mechanism that has been discussed via disorienting the chiral condensate in a heavy ion collision. At temperatures high above the QCD scale, $U(1)_{PQ}$ is almost an exact global symmetry, which gets spontaneously broken at some high scale. This necessarily leads to the generation of a network of cosmic strings. Such a string network can lower its energy by radiating axions. Once the axion mass becomes important, the string solutions become unstable, and the string network disappears, again leading to axion emission. This production mechanism may also lead to enough axions to close the Universe.

Appendix A

Units, Scales, and Hierarchies in Particle Physics

Physical units represent man-made conventions influenced by the historical development of physics. Interestingly, there are also *natural* units which express physical quantities in terms of fundamental constants of Nature: Newton's gravitational constant G, the velocity of light c, and Planck's action quantum h. In this appendix we consider the units commonly used in particle physics, and we discuss energy scales and mass hierarchies.

A.1 Man-Made versus Natural Units

The most basic physical quantities — length, time, and mass — are measured in units of meters (m), seconds (sec), and kilograms (kg). Obviously, these are man-made units appropriate for the use at our human scales. For example, the length of a step is roughly one meter, the duration of a heart beat is about one second, and one kilogram is a reasonable fraction of our body weight, *e.g.* the weight of a loaf of bread.

Time is measured by counting periodic phenomena. An individual cesium (Cs) atom is an extremely accurate clock. In fact, 1 second is defined as 9192631770 periods of a particular microwave transition of the Cs atom.

While the meter was originally defined by the length of the meter stick kept in the “Bureau International des Poids et Measure” in Paris, one now defines the meter through the speed of light c and the second as

$$1\,\mathrm{m} = 3.333564097 \times 10^{-7} c \text{ sec} \ . \tag{A.1.1}$$

In other words, the measurement of distance is reduced to the measurement of time by invoking a natural constant. Together with the meter stick, a certain amount of platinum-iridium alloy was deposited in Paris more than hundred years ago. The corresponding mass was defined to be one kilogram.

Expressed in those man-made units, Nature's most fundamental constants are the speed of light

$$c = 2.99792458 \times 10^{8}\,\mathrm{m\ sec}^{-1} \ , \tag{A.1.2}$$

Planck's action quantum (divided by 2π)

$$\hbar = 1.05457163(5) \times 10^{-34}\,\mathrm{kg\ m^2 sec^{-1}} \ , \tag{A.1.3}$$

and Newton's gravitational constant

$$G = 6.6743(1) \times 10^{-11}\,\mathrm{m^3\,kg\ sec^{-2}} \ . \tag{A.1.4}$$

Appropriately combining these fundamental constants, Nature provides us with her own natural units (also known as Planck units): the Planck length

$$l_{\text{Planck}} = \sqrt{\frac{G\hbar}{c^3}} = 1.6160 \times 10^{-35}\,\mathrm{m} \ , \tag{A.1.5}$$

and the Planck time

$$t_{\text{Planck}} = \sqrt{\frac{G\hbar}{c^5}} = 5.3904 \times 10^{-44}\,\mathrm{sec} \ , \tag{A.1.6}$$

which represent the shortest distances and times relevant in physics. Today we are very far from exploring such short length- and time-scales experimentally. It is even expected that our classical concepts of space and time may break down at the Planck scale. One might speculate that, at the Planck scale, space and time are not resoluble any more, and that l_{Planck}

and $t_{\rm Planck}$ may represent the shortest elementary quantized units of space and time. We can also define the Planck mass

$$M_{\rm Planck} = \sqrt{\frac{\hbar c}{G}} = 2.1768 \times 10^{-8}\,\mathrm{kg}\ . \tag{A.1.7}$$

Planck units are not very practical in our everyday life. For example, a step has a length of about $10^{35}\,l_{\rm Planck}$, a heart beat lasts roughly $10^{44}\,t_{\rm Planck}$, and the mass of our body is about $10^{10}\,M_{\rm Planck}$. Still, $l_{\rm Planck}$, $t_{\rm Planck}$, and $M_{\rm Planck}$ are the most fundamental basic units that Nature provides us with. It is interesting to ask why we exist at scales so far removed from the Planck scale. For example, why does a kilogram correspond to about $10^{8}\,M_{\rm Planck}$? In some sense, this is a "historical" question. The amount of platinum-iridium alloy deposited in Paris a long time ago, which defines the kilogram, obviously is an arbitrarily chosen man-made unit. Why was it chosen in this particular manner? If we assume that the kilogram was chosen because it is a reasonable fraction of our body weight, we may rephrase the question as a biological one: Why do intelligent beings weigh about $10^{10}\,M_{\rm Planck}$? If biology could explain the number of cells in our body and, with some help from chemistry, could also explain the number of atoms necessary to form a cell, we can reduce the question to a physics problem. Since atoms get their mass from protons and neutrons (which have about the same mass), we are led to ask: Why is the proton mass

$$M_p = 1.67266 \times 10^{-27}\,\mathrm{kg} = 7.6840 \times 10^{-20}\,M_{\rm Planck} \tag{A.1.8}$$

so light compared to the Planck mass? This hierarchy puzzle, which is discussed in Chapter ???, has been understood at least qualitatively using the property of asymptotic freedom of Quantum Chromodynamics — the quantum field theory of quarks and gluons whose interaction energy explains the mass of the proton. As discussed in Chapter ???, eq. (A.1.8) also explains why gravity is an extremely weak force.

Since, the ratio $M_p/M_{\rm Planck} \approx 10^{-19}$ is so tiny, it is unpractical to use $M_{\rm Planck}$ as a basic unit of mass in particle physics. Instead it is common to use one electron Volt, the energy that an electron (of charge $-e$) picks up when it is accelerated by a potential difference of one Volt,

$$1\,\mathrm{eV} = 1.6022 \times 10^{-19}\,\mathrm{kg\ m^2\,sec^{-2}}\ , \tag{A.1.9}$$

as a basic energy unit. Obviously, the Volt, and therefore also the eV, is again a man-made unit — as arbitrarily chosen as, for example, 1 kg. The rest energy of a proton is then given by

$$M_p c^2 = 0.93827203(8)\,\mathrm{GeV}\ . \tag{A.1.10}$$

In particle physics it is common practice to put $\hbar = c = 1$. Then masses and momenta are measured in energy units, and lengths and times are measured in units of inverse energy. In particular, one has

$$\hbar c = 3.1616 \times 10^{-26}\,\mathrm{kg\ m^3\,sec^{-2}}\ , \tag{A.1.11}$$

such that $\hbar c = 1$ implies

$$1\,\mathrm{fm} = 10^{-15}\,\mathrm{m} = (0.1973\,\mathrm{GeV})^{-1} \tag{A.1.12}$$

for the scale of nuclear radii.

The strength of the electromagnetic interaction is determined by the quantized charge unit e (the electric charge of a proton). In natural units it gives rise to the experimentally determined fine-structure constant

$$\alpha = \frac{e^2}{\hbar c} = \frac{1}{137.03599968(1)}\ . \tag{A.1.13}$$

The strength of electromagnetism is determined by this pure number which is completely independent of any man-made conventions. An interesting question (that *e.g.* Wolfgang Pauli was faszinated by) is why α takes this particular value. At the moment, we have no clue how to answer this question. Some physicists like to use the anthropic principle: if α would be different, atomic physics and thus chemistry would work differently, and life as we know it would be impossible. Obviously, we can only exist in a Universe with a value of α that is hospitable to life. According to the anthropic principle, our existence may "explain" the value of α. The authors prefer not to subscribe to this way of thinking. In particular, the anthropic principle should only be used as a last resort, when all other explanations fail (which may still turn out to be the case for α). Let us be more optimistic and hope that some extension of the Standard Model will eventually explain the measured value of α.

A.2 Energy Scales and Particle Masses

Table A.1 lists the charges, masses, and life times of some important particles. The photon is, as far as we can tell, exactly massless, in agreement with the unbroken gauge symmetry $U(1)_{\mathrm{em}}$. Then it cannot possibly decay into anything lighter and is therefore stable.[1] Just as the photon mediates the electromagnetic interaction, the heavy gauge bosons W^+, W^-, and Z mediate the weak interaction. Unlike the photon, the electroweak gauge bosons are unstable against the decay into other particles and live only for about 10^{-25} sec. Due to their large mass, there is a large phase space for the decay into light particles which causes the short life times of the W- and Z-bosons. The inverse of their mass determines the very short range 10^{-17} m of the weak interaction.

Ordinary matter consists of protons and neutrons forming atomic nuclei which are surrounded by a cloud of electrons. While the life time of an isolated neutron is finite (about 13.5 minutes), because it decays into proton, electron, and anti-neutrino, a neutron bound inside a stable atomic nucleus cannot decay. Despite numerous experimental searches, protons have never been observed to decay. Still, as discussed in Chapter ???, the Standard Model does predict proton decay, however, at such a tiny rate that its experimental confirmation is practically impossible. Grand Unified Theories (GUTs) predict proton decay at a larger and perhaps detectable rate. Such theories may eventually explain the baryon asymmetry — the fact that there is more matter than anti-matter in the Universe.

The pions π^+, π^0, and π^- are the lightest hadrons. They are responsible for the large-distance contribution to the (still very short-ranged) nuclear force between protons and neutrons. The charged pions $\pi^\pm$ are relatively long lived, because they decay only through processes of the weak interactions. The neutral pion π^0, on the other hand, lives much shorter, because (as discussed in Chapter ???) it can decay electromagnetically into two photons. The Standard Model Lagrangian contains only one dimensionful parameter — the vacuum expectation value v of the Higgs field — which

[1]Since photons can be emitted or absorbed by charged particles, their number is not conserved.

Particle Type	Particle	Electric Charge	Mass [GeV]	Life time
Gauge Bosons	Photon γ	0	$< 10^{-35}$	stable
	$W^{\pm}$-bosons	± 1	80.398(25)	$3.07(6) \times 10^{-25}$ sec
	Z-boson	0	91.1876(21)	$2.64(3) \times 10^{-25}$ sec
Leptons	Neutrino ν_e	0	$< 2 \times 10^{-9}$	unknown
	Electron e	-1	$0.51099891(1) \times 10^{-3}$	$> 2 \times 10^{22}$ years
Baryons	Proton p	1	0.93827203(8)	$> 2.1 \times 10^{29}$ years
	Neutron n	0	0.93956536(8)	885(1) sec
Mesons	Pion π^0	0	0.1349766(6)	$8.4(6) \times 10^{-17}$ sec
	Pions $\pi^{\pm}$	± 1	0.1395702(4)	$2.6033(5) \times 10^{-8}$ sec

Table A.1: *Electric charges (in units of e), masses, and life times of some particles.*

takes the experimentally determined value

$$v = 246\,\mathrm{GeV} = 2.02 \times 10^{-17} M_{\mathrm{Planck}}\ . \tag{A.2.1}$$

There is a huge hierarchy separating the electroweak scale v from the Planck scale M_{Planck} set by the gravitational force. Since v is a free parameter of the Standard Model, at present we don't know where the hierarchy originates from. Indeed, in order to adjust v at its experimental value, the bare mass parameter of the Higgs field must be fine-tuned to a large number of decimal places. Many physicists consider this unnatural. Some theories beyond the Standard Model (*e.g.* those based on technicolor) attempt to solve the hierarchy problem by explaining the ratio v/M_{Planck} without any need for fine-tuning.

The charges and masses of the leptons are listed in table A.2. There are three generations of leptons containing the charged leptons — electron, muon, and tau — as well as the corresponding neutrinos. The masses of the charged leptons are experimentally known to a high accuracy. In the Standard Model the lepton masses are free parameters, resulting from the Yukawa couplings to the Higgs field. At present, we have no clue either why the lepton masses take their respective values. In particular, we do not understand why the masses of the electron and the tau-lepton differ by more than three orders of magnitude, or why the electron mass is more than five orders of magnitude smaller than the electroweak scale v.

Currently, only upper bounds exist for the neutrino masses. Indeed, in the Standard Model the neutrinos are exactly massless. However, the recent observations of neutrino oscillations imply that (at least some) neutrinos must have a non-zero mass. In certain extensions of the Standard Model (*e.g.* in Grand Unified Theories) the so-called see-saw mechanism, which is discussed in Chapter ???, can explain very small neutrino masses, at the expense of introducing a new GUT energy scale $\Lambda_{\rm GUT} \approx 10^{15}$ GeV. In these theories, besides additional gauge bosons, there are extremely heavy Majorana neutrinos at this energy scale.

Generation	Lepton	Electric Charge	Mass [GeV]
1.	Electron-Neutrino ν_e	0	$< 2 \times 10^{-9}$
	Electron e	-1	$0.51099891(1) \times 10^{-3}$
2.	Muon-Neutrino ν_μ	0	$< 0.17 \times 10^{-3}$
	Muon μ	-1	$0.105658367(4)$
3.	Tau-Neutrino ν_τ	0	$< 15.5 \times 10^{-3}$
	Tau τ	-1	$1.7768(2)$

Table A.2: *Electric charges (in units of e) and masses of the three generations of leptons.*

Table A.3 summarizes the charges and masses of quarks. Again, quarks appear in three generations, with the up and down quark forming the first, the charm and strange quark the second, and the top and bottom quark the third generation. The electric charges of quarks are either $2/3$ or $-1/3$ of the elementary charge e. However, since quarks do not exist as individual objects but are confined inside hadrons, in agreement with Millikan-type experiments, at a fundamental level no fractionally charged physical states seem to exist in Nature.[2] Confinement also implies that quark masses do not represent the inertia of physical objects. Only the masses of the resulting

[2]Charge fractionalization of electrons is known to occur as a collective phenomenon in the condensed matter physics of the fractional quantum Hall effect.

hadrons are truly physical masses measuring inertia and gravitational coupling strength. Like other quantities in quantum field theory, quark masses are running, *i.e.* they depend on the chosen renormalization scheme and scale. The quark masses in table A.3 are defined in the so-called $\overline{\mathrm{MS}}$ minimal substraction renormalization scheme. The masses of the light quarks up, down, and strange are quoted at a scale of 2 GeV, while the masses of the heavy quarks charm, bottom, and top are quoted at the scale of the respective mass itself. The quark masses are given by the scale v — the only dimensionful parameter in the Standard Model Lagrangian — multiplied by the respective Yukawa couplings to the Higgs field. Again, we presently do not understand why the quark masses take these specific values. In particular, we don't know why the masses of the up and the top quark differ by more than four orders of magnitude.

Generation	Quark	Electric Charge	Mass [GeV]
1.	up u	$2/3$	0.003(1)
	down d	$-1/3$	0.006(1)
2.	charm c	$2/3$	1.24(9)
	strange s	$-1/3$	0.10(2)
3.	top t	$2/3$	173(3)
	bottom b	$-1/3$	4.20(7)

Table A.3: *Electric charges (in units of e) and running masses (in the* $\overline{\mathrm{MS}}$ *scheme at the respective mass scales) of the three generations of quarks.*

It is interesting to note that the masses of the proton and other hadrons are not proportional to v. In QCD hadrons arise non-perturbatively as states containing confined quarks and gluons. Remarkably, the proton mass is still about 0.9 GeV even when the quark masses are set to zero. For massless quarks (*i.e.* in the chiral limit), the QCD action contains no dimensionful parameter and is thus scale invariant. However, scale invariance is anomalous, *i.e.* although it is present in the classical theory, it is explicitly broken at the quantum level. Since quantum field theories must be regularized and renormalized, upon quantization a dimensionful cut-off parameter enters the theory. Even when the cut-off is removed, a dimensionful scale is left behind. This phenomenon — which is known as dimensional transmutation — is visible already in perturbation theory. In particular, in the $\overline{\text{MS}}$ renormalization scheme the perturbatively defined scale $\Lambda_{\overline{\text{MS}}}$ arises, whose value in the two flavor theory (with up and down quarks only) is given by

$$\Lambda_{\overline{\text{MS}}} = 0.260(40)\,\text{GeV} . \tag{A.2.2}$$

In the chiral limit $\Lambda_{\overline{\text{MS}}}$ is the only scale of QCD, to which all hadron masses are proportional. For example, the ratio $M_p/\Lambda_{\overline{\text{MS}}}$ is a dimensionless number predicted by non-perturbative QCD without any adjustable parameters. While the proton mass is provided by Nature, the scale $\Lambda_{\overline{\text{MS}}}$ is again a man-made unit, introduced by theoretical physicists to ease perturbative calculations in QCD. Of course, in contrast to the kg, which was chosen at our human scales, $\Lambda_{\overline{\text{MS}}}$ is chosen at the relevant energy scale of the strong interaction. As mentioned before, as a consequence of the property of asymptotic freedom of QCD, it is natural that the proton mass M_p (and hence the QCD scale $\Lambda_{\overline{\text{MS}}}$) is much smaller than the Planck scale M_{Planck}. On the other hand, since the Higgs sector of the Standard Model is not asymptotically free, the hierachy problem arises: Why is v so much smaller than M_{Planck}? As long as this problem remains unsolved, we will not understand either why $\Lambda_{\overline{\text{MS}}}$ is about three orders of magnitude smaller than v.

Just as the fundamental electric charge e determines the strength of the electromagnetic interaction between photons and electrons or other charged particles, the strong coupling constant g_s determines the strength of the strong interaction between quarks and gluons. Like the quark masses, the QCD analog $\alpha_s = g_s^2/\hbar c$ of the fine-structure constant $\alpha = e^2/\hbar c$ also depends on the renormalization scale and scheme. Asymptotic freedom implies that α_s goes to zero in the high energy limit, *i.e.* the strong interaction becomes weak at high momentum transfers. At the scale of the Z-boson mass the quark-gluon coupling constant is given by

$$\alpha_s(M_Z) = 0.1176(20) . \tag{A.2.3}$$

The parameter $\Lambda_{\overline{\text{MS}}}$ sets the energy scale at which α_s becomes strong.

The electroweak interactions are described by the gauge group $SU(2)_L \times U(1)_Y$ with two corresponding gauge coupling constants g and g'. At temperature scales below v, in particular in the vacuum, the $SU(2)_L \times U(1)_Y$

symmetry is spontaneously broken down to the $U(1)_{\rm em}$ gauge group of electromagnetism with the elementary electric charge given by

$$e = \frac{gg'}{\sqrt{g^2 + g'^2}} \ . \tag{A.2.4}$$

The ratio of the W- and Z-boson masses is given by

$$\frac{M_W}{M_Z} = \frac{g}{\sqrt{g^2 + g'^2}} = \cos\theta_{\rm W} \ , \tag{A.2.5}$$

which defines the Weinberg angle $\theta_{\rm W}$. Its measured value is given by

$$\sin^2\theta_{\rm W} = \frac{g'^2}{g^2 + g'^2} = 0.2226(5) \ . \tag{A.2.6}$$

Just as the value of the fine-structure constant $\alpha \approx 1/137$ is not understood theoretically, the values of the three gauge couplings g_s, g, and g' associated with the Standard Model gauge group $SU(3)_{\rm c} \times SU(2)_L \times U(1)_Y$ are not understood either. When one uses the renormalization group to evolve the three gauge couplings from the currently experimentally accessible energy scales all the way up to the GUT scale $\Lambda_{\rm GUT}$, in a supersymmetric GUT extension of the Standard Model the couplings converge to one unified value. Hence, properly designed GUT theories are indeed able to relate the values of the gauge couplings g_s, g, and g', or equivalently $\Lambda_{\overline{\rm MS}}$, $\sin^2\theta_{\rm W}$, and α.

A.3 Fundamental Standard Model Parameters

Let us consider the fundamental Standard Model parameters. While their values can be determined experimentally, they are not understood on theoretical grounds. In fact, achieving a deeper understanding of these free parameters would require the discovery of even more fundamental structures underlying the Standard Model.

First, we concentrate on the minimal Standard Model with massless neutrinos and we consider its parameters in the order in which they appear in the book. Since the Standard Model is renormalizable, only a finite number of terms enter its Lagrangian. Consequently, the number of fundamental Standard Model parameters is also finite.

Chapter **??** addresses the Higgs sector of the Standard Model, whose Lagrangian contains two fundamental parameters — v and λ — which determine the quartic potential $V(\Phi) = \lambda/4!(\Phi^\dagger\Phi - v^2)^2$. The vacuum expectation value $v = 246$ GeV of the Higgs field is the only dimensionful parameter that enters the Lagrangian of the minimal Standard Model. In particular, in combination with the scalar self-coupling λ it determines the Higgs boson mass $m_{\mathrm{H}} = \sqrt{\lambda/3}v$.

Chapter **??** extends the Higgs sector by gauging its $\mathrm{SU}(2)_L\times$ $\mathrm{U}(1)_Y$ symmetry. In this way, two additional fundamental parameters — the gauge couplings g and g' — arise, which (together with v) determine the masses of the W- and Z-bosons, $m_W = \frac{1}{2}gv$ and $m_Z = \frac{1}{2}\sqrt{g^2 + g'^2}$, as well as the Weinberg angle $\cos\theta_W = g/\sqrt{g^2 + g'^2}$ and the electric charge $e = gg'/\sqrt{g^2 + g'^2}$.

In Chapter **??** the gluon field appears as the $\mathrm{SU}(3)_{\mathrm{c}}$ gauge field mediating the strong interaction, with the corresponding gauge coupling g_{s} as another fundamental Standard Model parameter. By dimensional transmutation, the dimensionless coupling g_{s} is traded for the dimensionful parameter $\Lambda_{\mathrm{QCD}} = 0.???$ GeV, that sets the energy scale at which the running gauge coupling becomes strong. Besides v (which enters the theory via the Lagrangian), Λ_{QCD} (which appears in the process of renormalization) is the only dimensionful parameter of the minimal Standard Model.

Many more parameters arise from the Yukawa couplings between the Higgs field and the lepton and quark fields, which are addressed in Chapters **??** and **??**. In the minimal Standard Model with massless neutrinos there are three dimensionless Yukawa couplings f_{e}, f_μ, and f_τ, which determine the masses of the charged leptons $m_{\mathrm{e}} = f_{\mathrm{e}}v$, $m_\mu = f_\mu v$, and $m_\tau = f_\tau v$. Similarly, there are six Yukawa couplings f_{u}, f_{d}, f_{c}, f_{s}, f_{t}, and f_{b}, for the different quark flavors. Besides these, there are four more parameters which determine the Cabibbo-Kobayashi-Maskawa quark mixing matrix: three angles (including the Cabibbo-angle) and one CP-violating phase. Hence, including the Higgs self-coupling λ as well as the vacuum expectation value v, altogether there are $2 + 3 + 6 + 4 = 15$ fundamental Standard Model parameters associated with the non-gauge interactions between Higgs and matter fields.

Besides the three gauge couplings g, g', and g_s, as discussed in Chapters **??** and **??**, the gauge interactions also give rise to the strong and electromagnetic vacuum angles θ and θ_QED. The weak interaction vacuum angle, on the other hand, is unphysical and can be absorbed by a field redefinition (cf. Chapter **??**). The experimentally determined QCD vacuum angle $|\theta| \leq 10^{-10}$ leads to the strong CP-problem: Why is θ consistent with zero? Like understanding the value of any other fundamental Standard Model parameter, solving this problem requires to go beyond the Standard Model. The value of the electromagnetic vacuum angle θ_QED is not known. In fact, this parameter is often ignored and indeed deserves more attention than it has received until now (cf. Chapter **??**).

Altogether the gauge and non-gauge interactions of the minimal Standard Model give rise to $5 + 15 = 20$ fundamental parameters.

Since the discovery of neutrino oscillations, it is clear that neutrinos have mass. As discussed in Chapter **??**, this can be accounted for by adding non-renormalizable dimension 5 operators to the minimal Standard Model, thus treating it as a low-energy effective theory. Once non-renormalizable terms are included, the number of free parameters grows very quickly, in principle even to infinity. Hence, in order to continue counting a finite number of fundamental parameters, we now consider the renormalizable extension of the minimal Standard Model by right-handed neutrino fields (cf. Chapter **??**). In this way, three Dirac mass parameters m_{ν_e}, m_{ν_μ}, and m_{ν_τ} (or, equivalently, three corresponding dimensionless Yukawa couplings f_{ν_e}, f_{ν_μ}, and f_{ν_τ}) as well as three dimensionful Majorana mass parameters M_i ($i \in \{1, 2, 3\}$) enter the extended Lagrangian. In this way, besides v and Λ_QCD, the high-energy scales M_i are introduced, which — via the see-saw mass mixing mechanism — give rise to small neutrino masses. In addition, as an analog of the CKM matrix, the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) lepton mixing matrix arises. As discussed in Chapter **??**, it contains three mixing angles as well as two CP-violating phases and thus it contributes five additional fundamental parameters. Hence, in the extended renormalizable Standard Model with right-handed neutrino fields, there are $3 + 3 + 5 = 11$ additional fundamental parameters.

As discussed in Chapter **??** the minimal Standard Model extended by right-handed neutrino fields may provide us with a viable dark matter candidate as well as with a satisfactory explanation of the baryon asymmetry. Hence, it may account for all observed fundamental phenomena with the exception of gravity. As discussed in Chapter **??**, perturbative quantum gravity can indeed be incorporated in the Standard Model, at least as an effective low-energy theory. At leading order, the corresponding Lagrangian is the one of classical general relativity, with Newton's constant G (or equivalently the Planck mass M_{Planck}) and the cosmological constant Λ_{c} as two additional dimensionful fundamental parameters. Further extended by gravity in this way, the Standard model then contains $20 + 11 + 2 = 33$ fundamental parameters. This model might, in fact, be valid all the way up to the Planck scale, where it would necessarily have to be replaced by a theory of non-perturbative quantum gravity.

Until now we have counted those fundamental Standard Model parameters that take continuous values. In addition, there are many hidden discrete parameters, such as, for example, the number of generations or the number of quark colors N_{c}, which indeed plays a prominent role in this book. Other discrete parameters are associated with the number of fundamental fields and their representations under the various gauge groups. What one considers a discrete parameter is a matter of choice, and thus counting them is ambiguous. There are many deep questions beyond the Standard Model related to its discrete parameters, such as: Why are there three generations? Why is the gauge group $SU(3)_{\mathrm{c}}\times SU(2) \times U(1)$? Why do quarks transform in the fundamental representation of $SU(3)_{\mathrm{c}}$? Why are there three space and one time dimension? This list could easily be extended further and we have no clue how to answer any of these deep questions.

One could have argued that the weak hypercharges of the leptons and quarks, $Y_{\mathrm{l_L}}$, $Y_{\mathrm{e_R}}$, $Y_{\mathrm{q_L}}$, $Y_{\mathrm{u_R}}$, and $Y_{\mathrm{d_R}}$ should be counted as additional continuous parameters. While this would indeed be correct at the classical level, as we have seen in Chapter **??**, at the quantum level these parameters are fixed by anomaly cancellation and should thus not be counted as free parameters.

While the total number of 33 continuous fundamental parameters may seem quite large, we should not forget that all other physical quantities (and, in fact, all other natural phenomena), at least in principle, originate from those parameters. In addition, while many of the parameters — including, for example, the masses and mixing angles of the heavy quarks or the neutrinos — are relevant in particle physics, their values have hardly any impact on the rest of physics. Only a few parameters — namely the masses of the light quarks and the electron, the Cabibbo angle, and the three gauge couplings, as well as v — at least in principle, determine all of nuclear and atomic physics (and hence condensed matter physics, chemistry, biology, and everything else that may eventually be related to fundamental physical processes). Of course, one should not forget that, in practice, the predictive power of the Standard Model is limited. While it forms a theoretical foundation for other subfields of physics, in no way does it make them any less important.

Appendix B

Basics of Quantum Field Theory

This chapter presents an introduction to the structure of quantum field theory. Classical field theories are introduced as a generalization of point mechanics to systems with infinitely many degrees of freedom — some number in each space point. Similarly, quantum field theories are just quantum mechanical systems with infinitely many degrees of freedom. In the same way as point mechanics, classical field theories can be quantized by means of the path integral — or functional integral — method. A schematic overview is sketched in Figure 1.B.

The transition to Euclidean time (Wick rotation) is favorable for the convergence of functional integrals. The resulting quantum field theories in Euclidean space have a close analogy to statistical mechanics. In this context, we also address the lattice regularization, which provides a formulation of quantum field theories beyond perturbation theory. In order to capture fermions, we introduce Grassmann variables and discuss the integration of Grassmann fields.

Figure B.1: *Overview of the transitions between different branches of physics: we proceed from mechanics to field theory (left to right), and from classical physics to quantum physics (top to bottom).*

B.1 From Point Particle Mechanics to Classical Field Theory

Point mechanics describes the dynamics of classical, non-relativistic point particles. The coordinates of the particles represent a finite number of degrees of freedom. In the simplest case — a single particle moving on a line — this degree of freedom is just given by the particle position[1] x, as a function of the time t. The dynamics of a particle of mass m moving in an external potential $V(x)$ obeys *Newton's equation*

$$F(x) = m\ddot{x} = -V'(x) \tag{B.1.1}$$

(where m is constant). Once the initial conditions are specified, this ordinary second order differential equation determines the path of the particle, $x(t)$.

[1]For the considerations here, and in Section 1.2 and 1.3, the space dimension hardly matters. For simplicity we set it to 1, but a generalization to higher dimension is trivial; one just does the same in each dimension, and replaces x by $\vec{x}$ everywhere.

Newton's equation can be obtained from the *variational principle* by minimizing the action,

$$S[x] = \int dt\ L(x, \dot{x}) \ , \tag{B.1.2}$$

in the set of all paths $x(t)$.[2] The action is a functional (a function whose argument is itself a function) that results from the time integral of the Lagrange function

$$L(x, \dot{x}) = \frac{m}{2}\dot{x}^2 - V(x) \tag{B.1.3}$$

over some particle path with fixed end-points in space and time. Now the variational condition $\delta S = 0$ implies the *Euler-Lagrange equation*

$$\partial_t \frac{\delta L}{\delta \dot{x}} - \frac{\delta L}{\delta x} = 0 \ , \tag{B.1.4}$$

which coincides with Newton's equation (B.1.1) at any time t.

Classical field theories are a generalization of point mechanics to systems with infinitely many degrees of freedom — a given number for each space point $\vec{x}$. In this case, the degrees of freedom are the field values $\phi(\vec{x}, t)$, where ϕ represents an arbitrary field. We mention a few examples:

- In the case of a *neutral scalar field*, ϕ is simply a real number representing one degree of freedom per space point.
- A *charged scalar field*, on the other hand, is described by a complex number. Hence it represents two degrees of freedom per space point.
- The *Higgs field* $\phi^a(\vec{x}, t)$ (with $a \in \{1, 2\}$), which is part of the Standard Model, is a complex doublet; it has four real degrees of freedom per space point.
- An *Abelian gauge field* $A_\mu(\vec{x}, t)$ (with index $\mu \in \{0, 1, 2, 3\}$) — in particular the photon field in electrodynamics — is a neutral vector field, which seems to have 4 real degrees of freedom per space point.

[2]More precisely, one identifies a stationary point in the set of possible paths connecting fixed end-points.

However, two of them are redundant due to the $U(1)$ gauge symmetry.[3] Hence the Abelian gauge field has two physical degrees of freedom per space point, which correspond to the two polarization states of the (massless) photon.

- A *non-Abelian gauge field* $A_\mu^a(\vec{x}, t)$ is charged and has an additional index a. For example, the *gluon field* in chromodynamics with a color index $a \in \{1, 2, \ldots, 8\}$ represents $2 \times 8 = 16$ physical degrees of freedom per space point, again because of some redundancy due to the $SU(3)$ color gauge symmetry.

  The field that represents the *W- and Z-bosons* in the Standard Model has an index $a \in \{1, 2, 3\}$ and transforms under the gauge group $SU(2)$. Thus, to start with, it represents $2 \times 3 = 6$ physical degrees of freedom. However, in contrast to the photon and the gluons, the W- and Z-bosons are massive due to the Higgs mechanism, to be discussed later. Therefore they are equipped with three (not just two) polarization states. The three extra degrees of freedom are provided by the Higgs field, which is then left with only one degree of freedom in each space point.

The analogue of Newton's equation in field theory is the *classical field equation of motion*. For instance, for a neutral scalar field it reads

$$\partial_\mu \partial^\mu \phi = -\frac{dV(\phi)}{d\phi} \ . \tag{B.1.5}$$

Again, after specifying appropriate initial conditions it determines the classical field configuration $\phi(x)$, *i.e.* the values of the field ϕ in all space-time points $x = (\vec{x}, t)$. Hence, the rôle of `time` in point mechanics is played by `space-time` in field theory, and the rôle of the point particle `coordinates` is now played by the `field values`. As before, the classical equation of motion results from minimizing the action, which now takes the form

$$S[\phi] = \int d^4x \ \mathcal{L}(\phi, \partial_\mu \phi) \ , \quad \text{where} \quad d^4x = d^3x \, dt \ . \tag{B.1.6}$$

[3] For instance, the Lorenz gauge condition $\partial^\mu A_\mu = 0$ leaves the Abelian gauge field with 3 degrees of freedom (in each space point), but it does still not fix the gauge completely. Doing so swallows yet another degree of freedom.

The integral over a time interval in eq. (B.1.2) is extended to an integral over a volume in space-time, and the Lagrange function L of point mechanics is replaced by the Lagrange density, or *Lagrangian*,[4]

$$\mathcal{L}(\phi, \partial_\mu \phi) = \frac{1}{2}\partial_\mu \phi \partial^\mu \phi - V(\phi) \ . \tag{B.1.7}$$

A prominent interacting field theory is the $\lambda\Phi^4$ model with the potential[5]

$$V(\phi) = \frac{m^2}{2}\phi^2 + \frac{\lambda}{4!}\phi^4 \ . \tag{B.1.8}$$

Classically m is the mass of the scalar field ϕ, and λ is the coupling strength of its self-interaction. The mass term[6] corresponds to a harmonic oscillator potential in the point mechanics analogue, while the interaction term corresponds to an anharmonic perturbation.

Here the condition $\delta S = 0$ leads to the Euler-Lagrange equation

$$\partial_\mu \frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi)} - \frac{\delta \mathcal{L}}{\delta \phi} = 0 \ , \tag{B.1.9}$$

which is the equation of motion. In particular, based on the Lagrangian (B.1.7) we arrive at the scalar field equation (B.1.5). The analogies between point mechanics and field theory are summarized in Table B.1.

B.2 The Quantum Mechanical Path Integral

The quantization of field theories is conveniently performed using the *path integral approach* [**?**, **?**]. We first discuss the path integral in quantum mechanics — quantized point mechanics — using the real time formalism. A

[4]Throughout this book the derivatives ∂_μ act only on the immediately following field. We add that other authors use the term “Lagrangian” also for the Lagrange function (B.1.3).

[5]Part of the literature restricts “potentials” to the interaction terms, so the mass term is not included, but this is just a matter of terminology.

[6]For the moment we assume m^2 to be positive. Later we will also consider the case $m^2 < 0$.

Point Mechanics	**Classical Field Theory**
time t	space-time $x = (\vec{x}, t)$
particle coordinate x	field value ϕ
particle path $x(t)$ (for all t in some interval)	field configuration $\phi(x)$ (for all x in some volume)
Lagrange function $L(x, \dot{x}) = \frac{m}{2}\dot{x}^2 - V(x)$	Lagrangian $\mathcal{L}(\phi, \partial_\mu\phi) = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\Phi)$
action $S[x] = \int dt\ L(x, \dot{x})$	action $S[\phi] = \int d^4x\ \mathcal{L}(\phi, \partial_\mu\phi)$
equation of motion $\partial_t \frac{\delta L}{\delta \dot{x}} - \frac{\delta L}{\delta x} = 0$	field equation $\partial_\mu \frac{\delta \mathcal{L}}{\delta(\partial_\mu\phi)} - \frac{\delta \mathcal{L}}{\delta\phi} = 0$
Newton's equation $m\ddot{x} = -V'(x)$	scalar field equation $\partial_\mu\partial^\mu\phi = -\frac{dV(\phi)}{d\phi}$
kinetic energy $\frac{m}{2}\dot{x}^2$	kinetic energy density $\frac{1}{2}\partial_\mu\phi\partial^\mu\phi$
harmonic oscillator potential $\frac{m}{2}\omega^2x^2$	mass term $\frac{m^2}{2}\phi^2$
anharmonic potential $\frac{\lambda}{4!}x^4$	quartic self-interaction term $\frac{\lambda}{4!}\phi^4$

Table B.1: *A dictionary that translates 1-d point mechanics into the language of classical field theory in* 3 *spatial dimensions. Thus we proceed from one degree of freedom to an infinite number of degrees of freedom. Moreover, we consider field theories where* $\mathcal{L}$ *consists of Lorentz invariant terms, hence this translation also provides special relativity.*

mathematically safer formulation uses an analytic continuation to the so-called Euclidean time. This will be addressed in Section 1.3.

We use the Dirac notation, where a *ket* $|\Psi\rangle$ describes some state as a unit vector in a Hilbert space, and a *bra* $\langle\Psi|$ its Hermitian conjugate. Thus the *bracket* $\langle\Psi'|\Psi\rangle$ is a scalar product. The corresponding wave functions in (one dimensional) coordinate space and in momentum space are obtained as

$$\Psi(x,t) = \langle x\,|\Psi(t)\rangle\ , \qquad \Psi(p,t) = \langle p\,|\Psi(t)\rangle\ , \tag{B.2.1}$$

where $|x\rangle$ and $|p\rangle$ are the coordinate and momentum eigenstates, $\hat{x}|x\rangle = x|x\rangle\ ,\ \hat{p}|p\rangle\ = p|p\rangle\ .$ We further denote the energy eigenstates as $|n\rangle$, *i.e.* $\hat{H}|n\rangle = E_n|n\rangle$, where $\hat{H}$ is the Hamilton operator and E_n are the energy eigenvalues. The spatial energy eigenfunctions are then given by $\langle x|n\rangle$.

The eigenstates $|x\rangle$, $|p\rangle$ and $|n\rangle$ all build complete orthonormal sets,

$$\int dx\,|x\rangle\langle x| = \int \frac{dp}{2\pi\hbar}\,|p\rangle\langle p| = \sum_n\ |n\rangle\langle n| = \hat{1\!\!1}\ . \tag{B.2.2}$$

(Of course, the meaning is to build sums and integrals whenever the set of states is discrete resp. continuous). So we can write a scalar product as

$$\langle\Psi'|\Psi\rangle = \int dx\,\langle\Psi'|x\rangle\langle x|\Psi\rangle = \int dx\,\Psi'^{\dagger}(x)\,\Psi(x)\ . \tag{B.2.3}$$

The wave functions in coordinate and momentum space can be converted into one another by the Fourier transform and its inverse,

$$\begin{aligned}
\Psi(p,t) &= \int dx\,\langle p\,|x\rangle\,\langle x\,|\Psi(t)\rangle = \int dx\,e^{-\mathrm{ipx}/\hbar}\,\Psi(x,t)\ , \\
\Psi(x,t) &= \int \frac{dp}{2\pi\hbar}\,e^{\mathrm{ipx}/\hbar}\,\Psi(p,t)\ ,
\end{aligned} \tag{B.2.4}$$

so that $\hat{p}\Psi(x,t) = -\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{dx}}\Psi(\mathrm{x},\mathrm{t})$ and $\hat{x}\Psi(p,t) = \mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{dp}}\Psi(\mathrm{p},\mathrm{t})$.

A Hermitian operator $\hat{O}(\hat{x})$, which may represent some observable, takes the expectation value

$$\langle\Psi|\hat{O}|\Psi\rangle = \int dx\,\Psi^*(x)\,O(x)\,\Psi(x)\ . \tag{B.2.5}$$

The time evolution of a quantum system — described by a Hamilton operator $\hat{H}$ — is given by the time-dependent *Schrödinger equation*

$$\mathrm{i}\hbar\partial_{\mathrm{t}}|\Psi(\mathrm{t})\rangle = \hat{\mathrm{H}}|\Psi(\mathrm{t})\rangle \ . \tag{B.2.6}$$

Like Newton's equation, the Schrödinger equation describes the evolution over an infinitesimal time. As in Section 1.1 we proceed to its integrated form, *i.e.* to *finite time* steps, for which we write the ansatz

$$|\Psi(t')\rangle = \hat{U}(t',t)|\Psi(t)\rangle \qquad (t' \geq t) \ . \tag{B.2.7}$$

$\hat{U}(t',t)$ is the *evolution operator.* For a Hamilton operator without explicit time dependence it takes the simple form[7]

$$\hat{U}(t',t) = \exp\Big(-\frac{\mathrm{i}}{\hbar}\hat{H}(t'-t)\Big) \ . \tag{B.2.8}$$

Let us consider the transition amplitude $\langle x'|\hat{U}(t',t)|x\rangle$ of a non-relativistic point particle that starts at space-time point (x,t) and arrives at (x',t'). Using eqs. (B.2.1) and (B.2.2) we obtain

$$\Psi(x',t') = \int dx \ \langle x'|\hat{U}(t',t)|x\rangle \ \Psi(x,t) \ , \tag{B.2.9}$$

i.e. $\langle x'|\hat{U}(t',t)|x\rangle$ acts as a *propagator* for the wave function, if we assume $t' > t$.

The propagator is a quantity of primary physical interest. In particular it contains information about the *energy spectrum:* let us consider the propagation from an initial position eigenstate $|x\rangle$ back to itself,

$$\begin{aligned} \langle x|\hat{U}(t',t)|x\rangle &= \langle x|\exp\Big(-\frac{\mathrm{i}}{\hbar}\hat{H}(t'-t)\Big)|x\rangle \\ &= \sum_n |\langle x|n\rangle|^2 \exp\Big(-\frac{\mathrm{i}}{\hbar}E_n(t'-t)\Big) \ , \end{aligned} \tag{B.2.10}$$

where we applied the last relation in eq. (B.2.2). Hence the inverse Fourier transform of the propagator yields the energy spectrum as well as the energy eigenstates.

[7] In the general case the evolution operator has to be expanded by the Dyson series. In the present case we could also just write $\hat{U}(t'-t)$. We stay with the general notation $\hat{U}(t',t)$, however, since the crucial decomposition in eq. (B.2.15) and the central result (B.2.17) hold in fact generally.

Inserting now a complete set of position eigenstates at some time t_1, with $t < t_1 < t'$, we obtain

$$\begin{aligned}
\langle x'|\hat{U}(t',t)|x\rangle &= \langle x'|\exp\Big(-\frac{\mathrm{i}}{\hbar}\hat{H}(t'-t_1)\Big)\exp\Big(-\frac{\mathrm{i}}{\hbar}\hat{H}(t_1-t)\Big)|x\rangle \\
&= \int dx_1 \langle x'|\exp\Big(-\frac{\mathrm{i}}{\hbar}\hat{H}(t'-t_1)\Big)|x_1\rangle \\
&\times \ \langle x_1|\exp\Big(-\frac{\mathrm{i}}{\hbar}\hat{H}(t_1-t)\Big)|x\rangle \\
&= \int dx_1 \langle x'|\hat{U}(t',t_1)|x_1\rangle\langle x_1|\hat{U}(t_1,t)|x\rangle \ . \qquad \text{(B.2.11)}
\end{aligned}$$

This expression is illustrated in Figure B.2 (on top).

Obviously 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 equidistant time steps of size ε such that

$$t' - t = N\varepsilon \ . \qquad \text{(B.2.12)}$$

Inserting a complete set of position eigenstates at the intermediate times $t_j = t + j\varepsilon$, $j = 1, 2, \ldots, N-1$, we arrive at

$$\begin{aligned}
\langle x'|\hat{U}(t',t)|x\rangle &= \int dx_1 \int dx_2 \ldots \int dx_{N-1}\, \langle x'|\hat{U}(t',t_{N-1})|x_{N-1}\rangle \ldots \\
&\times \ \langle x_2|\hat{U}(t_2,t_1)|x_1\rangle\langle x_1|\hat{U}(t_1,t)|x\rangle \ . \qquad \text{(B.2.13)}
\end{aligned}$$

Now we are summing over *all paths,* as depicted (symbolically) in Figure B.2 (below).

In the next step we focus on one of these factors. We consider a single non-relativistic point particle moving in an external potential $V(x)$ such that

$$\hat{H} = \frac{\hat{p}^{\,2}}{2m} + \hat{V}(\hat{x}) \ . \qquad \text{(B.2.14)}$$

Using the Baker-Campbell-Hausdorff formula, and neglecting terms of order

Figure B.2: *A transition amplitude with one intermediate time t_1 (on top), and with a set of equidistant intermediate times $t_1 \dots t_{N-1}$ (below).*

ε^2, we obtain[8]

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

[8]This decomposition of $\exp(-\mathrm{i}\hat{\mathrm{H}}\varepsilon/\hbar)$ is also known as *Trotter's formula.* The fact that it holds only up to $O(\varepsilon^2)$ is the reason why we have to proceed in infinitesimal time steps. This formula is obvious for bounded operators $\hat{p}^{\,2}$, $\hat{V}$, but it is highly relevant that it also holds if these operators are only semi-bounded.

$$
\begin{aligned}
&\times \quad \exp\Big(-\frac{\mathrm{i}\varepsilon}{\hbar}\hat{V}(\hat{x})\Big)|x_i\rangle \\
&= \quad \frac{1}{2\pi\hbar}\int dp \exp\Big(-\frac{\mathrm{i}\varepsilon \mathrm{p}^2}{2m\hbar}\Big)\, \exp\Big(-\frac{\mathrm{i}}{\hbar}p(x_{i+1}-x_i)\Big) \\
&\times \quad \exp\Big(-\frac{\mathrm{i}\varepsilon}{\hbar}V(x_i)\Big)\,. \qquad \text{(B.2.15)}
\end{aligned}
$$

• **Question : Baker-Campbell-Hausdorff formula**

Let $\hat{A}$ and $\hat{B}$ be bounded operators, and ε is an infinitesimal parameter. For the product $\exp(\varepsilon\hat{A})\cdot\exp(\varepsilon\hat{B})$ we make the ansatz

$$\exp(\varepsilon\hat{A})\cdot\exp(\varepsilon\hat{B}) = \exp\Big(\varepsilon\hat{X} + \varepsilon^2\hat{Y} + \varepsilon^3\hat{Z} + O(\varepsilon^4)\Big)\,.$$

Compute the operators $\hat{X}$, $\hat{Y}$, and $\hat{Z}$, and express them in a compact form in terms of commutators.

This integral over p is ill-defined because the integrand is a rapidly oscillating function.[9] To make this expression well-defined we replace the time step ε by $\varepsilon-\mathrm{i}a$, $0 < a \ll 1$, *i.e.* we step a little bit into a complex time plane. After performing the integral we take the limit $a \to 0$. We keep in mind that the definition of the path integral requires an analytic continuation in time. One arrives at

$$\langle x_{i+1}|\hat{U}(t_{i+1},t_i)|x_i\rangle = \Big(\frac{m}{2\pi\mathrm{i}\hbar\varepsilon}\Big)^{1/2} \exp\Big(\frac{\mathrm{i}}{\hbar}\varepsilon\Big[\frac{m}{2}(\frac{x_{i+1}-x_i}{\varepsilon})^2 - V(x_i)\Big]\Big)\,. \qquad \text{(B.2.16)}$$

Inserting this back into the expression for the propagator we obtain

$$\boxed{\langle x'|\hat{U}(t',t)|x\rangle = \int \mathcal{D}x \, \exp\Big(\frac{\mathrm{i}}{\hbar}S[x]\Big)}\,. \qquad \text{(B.2.17)}$$

[9]Moreover, there is an ambiguity in the last factor of eq. (B.2.15): one could also argue that it should be $\exp\{-\frac{\mathrm{i}\varepsilon}{\hbar}V(x_{i+1})\}$ instead. In the present case this difference does not matter for the result that we obtain in the limit $\varepsilon \to 0$. This difference does matter, however, if the potential also depends on $\dot{x}$. This is the case for the electrodynamic vector potential, where one runs into an ambiguity, which corresponds to the ordering problem in operator Quantum Mechanics.

The action has been identified in the time continuum limit as

$$\begin{aligned} S[x] &= \lim_{\varepsilon\to 0} \varepsilon \sum_i \Big[\frac{m}{2}\Big(\frac{x_{i+1}-x_i}{\varepsilon}\Big)^2 - V(x_i)\Big] \\ &= \int dt \, \Big[\frac{m}{2}\dot{x}^2 - V(x)\Big] \, . \end{aligned} \tag{B.2.18}$$

The integration measure in eq. (B.2.17) is given by

$$\int \mathcal{D}x = \lim_{\varepsilon\to 0} \Big(\frac{m}{2\pi \mathrm{i}\hbar\varepsilon}\Big)^{\frac{N}{2}} \int dx_1 \int dx_2 \dots \int dx_{N-1} \, . \tag{B.2.19}$$

This means that we integrate over all possible particle positions at each intermediate time t_i. In this way we integrate over *all possible paths* of the particle starting at $x(t)$ and ending at $x'(t')$. Each path is weighted with a phase factor $\exp(\frac{\mathrm{i}}{\hbar}S[x])$. As in classical point mechanics, a finite time interval is handled by the Lagrangian. In quantum mechanics this formulation eliminates the operators, but it employs a (somewhat mysterious) functional measure $\mathcal{D}x$.

If the path is varied, this phase factor undergoes an extremely fast oscillation, because $\hbar$ is very small. The classical path of minimal action has the least oscillations, hence its vicinity provides the largest contribution to the path integral. In the limit $\hbar \to 0$ only the contribution of the classical path survives, and we are back at the Euler-Lagrange equation (B.1.4). At finite (but tiny) $\hbar$ the contributions of non-classical paths are still suppressed (or "washed out") by the rapidly oscillating phase; their remaining contributions to the path integral are the quantum effects.

Eq. (B.2.17) is the key result for the path integral formulation of quantum mechanics. It provides a transparent transition from classical physics to quantum physics as we turn on $\hbar$ gradually to include fluctuations around the path of minimal action. This transition has an analogue in optics, if we proceed from Fermat's principle to the more fundamental Huygens principle. Along this line, also the transition behavior of quantum particles through double slits (and multi-slits) is obvious in view of the path integral description. More detailed presentations can be found in *e.g.* Ref. [?, ?, ?, ?, ?, ?, ?].

B.3 The Path Integral in Euclidean Time

As we have seen, it takes at least a small excursion into the complex time plane to render the path integral well-defined. Now we will perform a radical step into that plane and consider purely imaginary time, the so-called *Euclidean time*. Remarkably, this formulation has a direct physical interpretation in the framework of statistical mechanics, as discussed comprehensively *e.g.* in Ref. [**?**].

Let us consider the quantum statistical partition function

$$\boxed{Z = \mathrm{Tr}\exp(-\beta\hat{H})}\ , \tag{B.3.1}$$

where $\beta = 1/T$ is the inverse temperature. It is mathematically equivalent to the time interval that we discussed in the real time path integral. In particular, the operator $\exp(-\beta\hat{H})$ turns into the time evolution operator $\hat{U}(t', t)$ in eq. (B.2.8) if we identify

$$\beta = \frac{\mathrm{i}}{\hbar}(t' - t)\ . \tag{B.3.2}$$

In this sense the system at finite temperature corresponds to a system propagating in purely imaginary time, *i.e.* in Euclidean time. The rotation of the time coordinate by $\pi/2$ in the complex plane is denoted as the *Wick rotation*. Clearly it transforms Minkowski's metric tensor $g_{\mu\nu}$ into a Euclidean metrics $\propto \delta_{\mu\nu}$, cf. Chapter 4.

By dividing the Euclidean time interval into N equidistant time steps, *i.e.* by writing $\beta = Na/\hbar$ — and by inserting again complete sets of position eigenstates — we now arrive at the *Euclidean time path integral* [10]

$$\boxed{Z = \int \mathcal{D}x \exp\Big(-\frac{1}{\hbar}S_E[x]\Big)}\ . \tag{B.3.3}$$

Here the action takes the Euclidean form

$$S_E[x] \ = \ \int_t^{t'} d\tau\ \Big[\frac{m}{2}\dot{x}^2 + V(x)\Big]$$

[10] Note that here the momentum integral corresponding to eq. (B.2.15) is well-defined from the beginning.

$$= \lim_{a \to 0} a \sum_i \Big[\frac{m}{2} (\frac{x_{i+1} - x_i}{a})^2 + V(x_i) \Big] \ . \tag{B.3.4}$$

Unlike the real time case, the measure now involves N integrals,

$$\int \mathcal{D}x = \lim_{a \to 0} \Big(\frac{m}{2\pi\hbar a} \Big)^{\frac{N}{2}} \int dx_1 \int dx_2 \dots \int dx_N \ . \tag{B.3.5}$$

The extra integration over $x_N = x'$ is due to the trace in eq. (B.3.1). Note that there is no extra integration over $x_0 = x$ because the trace implies periodic boundary conditions in the Euclidean time direction, $x_0 = x_N$.

The Euclidean path integral allows us to evaluate *thermal expectation values.* For example, let us consider an operator $\hat{\mathcal{O}}(\hat{x})$ that is diagonal in the position state basis $\{|x\rangle\}$. By inserting this operator into the path integral we obtain an expression for its expectation value,

$$\langle \hat{\mathcal{O}}(\hat{x}) \rangle = \frac{1}{Z} \mathrm{Tr}[\hat{\mathcal{O}}(\hat{x}) \exp(-\beta \hat{H})] = \frac{1}{Z} \int \mathcal{D}x \ \mathcal{O}(x(0)) \exp \Big(-\frac{1}{\hbar} S_E[x] \Big) \ . \tag{B.3.6}$$

Since the theory is translation invariant in Euclidean time, one can place the operator anywhere in time, *e.g.* at $t = 0$ as it is done here.

When we take 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, $Z \sim \exp(-\beta E_0)$. In this limit the path integral is formulated in a very long Euclidean time interval, which describes the *vacuum expectation values.* For instance, for the 1-point function it reads

$$\langle 0 | \hat{\mathcal{O}}(\hat{x}) | 0 \rangle = \lim_{\beta \to \infty} \frac{1}{Z} \int \mathcal{D}x \ \mathcal{O}(x(0)) \exp \Big(-\frac{1}{\hbar} S_E[x] \Big) \ . \tag{B.3.7}$$

In addition, it is very often of interest to consider *2-point functions* of operators at different instances in Euclidean time,

$$\langle \hat{\mathcal{O}}(\hat{x}(t)) \, \hat{\mathcal{O}}(\hat{x}(0)) \rangle = \frac{1}{Z} \mathrm{Tr} \Big[\exp(-\hat{H}t) \, \hat{\mathcal{O}}(\hat{x}) \exp(\hat{H}t) \, \hat{\mathcal{O}}(\hat{x}) \exp(-\beta \hat{H}) \Big]$$
$$= \frac{1}{Z} \int \mathcal{D}x \ \mathcal{O}(x(t)) \, \mathcal{O}(x(0)) \exp \Big(-\frac{1}{\hbar} S_E[x] \Big) \ . \tag{B.3.8}$$

Again we consider the limit $\beta \to \infty$, but we also separate the operators by a large difference in Euclidean 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 part, and thus forming the *connected 2-point function*, one obtains asymptotically

$$\lim_{\beta,\, t\to\infty} \langle \mathcal{O}(x(t))\mathcal{O}(x(0))\rangle - |\langle \mathcal{O}(x)\rangle|^2 = |\langle 1|\mathcal{O}(x)|0\rangle|^2 \exp\left(-(E_1 - E_0)t\right) \ . \tag{B.3.9}$$

Here $|1\rangle$ is the first excited state of the quantum system, with energy E_1. The connected 2-point function decays exponentially at large Euclidean time separations. This decay is governed by the *energy gap* $E_1 - E_0$.

At this point we anticipate that in a quantum field theory E_1 corresponds to the energy of the lightest particle. Its mass is determined by the energy gap $E_1 - E_0$ above the vacuum. Hence, in Euclidean field theory particle masses are evaluated from the exponential decay of connected 2-point functions.

B.4 Spin Models in Classical Statistical Mechanics

So far we have considered quantum systems, both at zero and at finite temperature. We have represented their partition functions by means of Euclidean path integrals over configurations on a time lattice of length β. We will now take a new start and consider *classical discrete systems at finite temperature*. We will see that their mathematical description is very similar to the path integral formulation of quantum systems. The physical interpretation, however, is basically different in the two cases. In the next section we will set up another dictionary that allows us to translate quantum physics language into the terminology of statistical mechanics. For further reading about spin models and critical phenomena we recommend the text books listed in Refs. [**?**, **?**, **?**].

For simplicity, let us concentrate on **classical spin models**. Here the term “spin” does not mean that we deal with quantized angular momenta.

All we do is work with classical vectors as field variables.[11] We denote these vectors as $\vec{s}_x = (s_x^1, \dots, s_x^N)$. In these models the spins live on the sites of a d-dimensional spatial grid. Hence x denotes a corresponding lattice site. The latter is often meant to be a *crystal lattice* (so typically $d = 3$), and here the lattice spacing has a physical meaning. This is in contrast to the time step that we have introduced before as a regularization in order to render the path integral mathematically well-defined, and that we finally send to zero to reach the temporal continuum limit.

One often normalizes the spin vectors,

$$|\vec{s}_x| = 1 \quad \text{at all sites } x \ . \tag{B.4.1}$$

The simplest spin model of this kind is the *Ising model* with classical spin variables $s_x = \pm 1$. Below we list some spin models, all of them with the constraint (B.4.1) and a global $O(N)$ spin rotation symmetry (which turns into $Z(2)$ for $N = 1$), see also Appendix A:

$$\begin{aligned} N = 1 \quad &: \quad \text{Ising model} \\ N = 2 \quad &: \quad \text{XY model} \\ N = 3 \quad &: \quad \text{classical Heisenberg model} \\ N = \infty \quad &: \quad \text{spherical model} \end{aligned}$$

The XY model is of theoretical interest, but it does hardly match experimental phenomena. On the other hand, the classical Heisenberg model is used for the description of ferromagnets, where the electron spins in some crystal cell are summed up to act collectively like a classical spin. In this case the spin space — which is completely abstract in general — is linked to the ordinary space. The $O(3)$ and $O(4)$ model also occur in particle physics (due to their local isomorphy to $SU(2)$ and $SU(2) \otimes SU(2)$), respectively, as we will see. The spherical model enables a number of analytical calculations, which are not feasible at finite N. Along with a $1/N$ expansion one can then hope to capture some features of the physically relevant cases $N \leq 4$.

These $O(N)$ spin models are characterized by a (classical) Hamilton function $\mathcal{H}$ (not a quantum Hamilton operator), which specifies the energy

[11] Quantum spin models also exist, but they are far more complicated: for instance, in those models it is already hard to identify the ground state.

of any spin configuration. The couplings between different spins are often limited to nearest neighbor sites, which we denote as $\langle xy \rangle$. The standard form of the Hamilton function reads

$$\mathcal{H}[\vec{s}\,] = J \sum_{\langle xy \rangle} \vec{s}_x \cdot \vec{s}_y - \vec{B} \cdot \sum_x \vec{s}_x \ . \tag{B.4.2}$$

It is *ferromagnetic* for a coupling constant $J < 0$ (which favors parallel spins), and anti-ferromagnetic for $J > 0$. In addition the spins prefer to be aligned with the external "magnetic field" $\vec{B} = (B^1, \dots, B^N)$.

In particular, the partition function of the Ising model is given by

$$Z = \prod_x \sum_{s_x = \pm 1} \exp(-\mathcal{H}[s]/T) := \int \mathcal{D}s \ \exp(-\mathcal{H}[s]/T) \ , \tag{B.4.3}$$

where we set again the Boltzmann constant $k_B = 1$. The *sum over all spin configurations* corresponds to the summation over all possible orientations of individual spins. For $N \geq 2$ the measure $\mathcal{D}s$ can be written as

$$\mathcal{D}s = \prod_x \int_{-1}^{1} ds_x^1 \dots \int_{-1}^{1} ds_x^N \ \delta(\vec{s}_x^{\,2} - 1) \ . \tag{B.4.4}$$

Thermal averages are computed by inserting appropriate quantities in the functional integrand. For example, the *magnetization* is given by

$$\langle \vec{s}_x \rangle = \frac{1}{Z} \int \mathcal{D}s \ \vec{s}_x \exp(-\mathcal{H}[\vec{s}\,]/T) \ . \tag{B.4.5}$$

Due to the translation invariance of the measure $\mathcal{D}s$, the result does not depend on x. So we can simply write the magnetization as $\langle \vec{s}\, \rangle$, in analogy to the time independence of the 1-point function (B.3.6).

Similarly, the spin *correlation function* is defined as

$$\langle \vec{s}_x \cdot \vec{s}_y \rangle = \frac{1}{Z} \int \mathcal{D}s \ \vec{s}_x \cdot \vec{s}_y \ \exp(-\mathcal{H}[\vec{s}\,]/T) \ , \tag{B.4.6}$$

which only depends on the distance $|x - y|$. Subtracting again the leading contribution, we obtain the *connected* spin correlation function. At large distances it typically decays exponentially,

$$\langle \vec{s}_x \cdot \vec{s}_y \rangle_{\rm c} = \langle \vec{s}_x \cdot \vec{s}_y \rangle - \langle \vec{s}\, \rangle^2 \sim \exp\left(- \frac{|x - y|}{\xi} \right) \ , \tag{B.4.7}$$

where ξ is called the *correlation length.* At high — or even moderate — temperatures the correlation length can well be just a few lattice spacings; strong thermal noise is destructive for long-range correlations.

When one models real materials, the Ising model appears as a drastic simplification, for example because real magnets also involve couplings beyond nearest neighbor spins. However, the details of the Hamilton function at the scale of the lattice spacing do not always matter. There may be a *critical temperature* $T_{\rm c}$ at which ξ diverges and a universal behavior arises. At this temperature a *second order phase transition* sets in. Then the details of the model at the scale of a few lattice spacings are irrelevant for the long-range physics that takes place at the scale of ξ. In fact, some real materials do behave close to their critical temperatures just like the simple Ising model of eq. (B.4.3). This is why this model attracts so much interest. It was introduced in the $19^{\rm th}$ century, solved in $d = 1$ by Ernst Ising in 1928, and in $d = 2$ by Lars Onsager in 1944 (this means that observables like the correlation functions could be computed explicitly). In higher dimensions an analytic solution has not be found so far, but there are analytical approximation techniques as well as accurate numerical results.

In $d = 1$ there is no finite temperature $T_{\rm c}$, and Ernst Ising concluded from this that the model is an over-simplification. There are, however, finite critical temperatures in $d > 1$, as first argued qualitatively by Rudolf Peierls (1941) for the 2-d case. Hence the Ising model *is* of interest, and it is in fact the most successful spin model in statistical mechanics.

The Ising model is just a very simple member of a large *universality class* of different models, which all share the same critical behavior. This does not mean that they have the same values of their critical temperatures. However, as the temperature T approaches T_c from below, their magnetization $M = |\langle \vec{s} \rangle|$ vanishes with the same power of $T_c - T$.[12] This universal behavior (at $B = |\vec{B}| = 0$) is characterized by the *critical exponent* β,

$$M \sim (T_c - T)^\beta \quad , \quad i.e. \quad \lim_{T \nearrow T_c} \frac{\ln M}{\ln(T_c - T)} = \beta \ . \qquad \text{(B.4.8)}$$

[12]In terms of thermodynamics one would write $M = -\frac{\partial f}{\partial B}|_{T=\rm const.}$, where f is the free energy density (and the free energy is given as $\mathcal{F} = -T \ln Z$).

Alternatively, if we fix the temperature T_c and include an external magnetic field $\vec{B}$, which is then gradually turned off, the magnetization goes to zero as $M \sim B^{1/\delta}$.

On the other hand, the susceptibility $\chi = T \frac{\partial M}{\partial B}|_{T=\mathrm{const.}}$ diverges for $T \to T_c$ as $\chi \sim |T - T_c|^{-\gamma}$. Also the specific heat $C = T \frac{\partial S}{\partial T}|_{B=\mathrm{const.}}$ diverges at the critical temperature; in that case we write $C \sim |T - T_c|^{-\alpha}$ (and $S = -\frac{\partial f}{\partial T}|_{B=\mathrm{const.}}$ is the entropy).

The parameters β, δ, γ and α are all critical exponents. For a large variety of materials — with different T_c, different crystal structure etc. — experimentalists found within a few percent

$$\beta \approx 1/3 \quad , \quad \gamma \approx 4/3 \quad , \quad \delta \approx 4.2 \quad , \quad \alpha \gtrsim 0 \ . \tag{B.4.9}$$

Therefore these values describe a universality class which plays a prominent rôle in Nature. A Table with explicit results for these critical exponents is displayed in Appendix A.

Note that dimensional quantities — like T_c — will clearly change if, for instance, the lattice spacing of the crystal is altered, as it often happens for materials with different kinds of molecules. On the other hand, the critical exponents are dimensionless — as all exponents in physics — hence these are the parameters which are suitable for an agreement within a universality class.

B.5 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.

The path integral for the quantum system is defined on a 1-dimensional Euclidean time lattice, while a spin model can be defined on a d-dimensional spatial lattice. In the path integral we integrate over all paths, *i.e.* over all "configurations of intermediate points" $x_i = x(t_i)$. In the spin model we sum over all spin configurations $\vec{s}_x$. Paths are weighted by their Euclidean action $S_E[x]$, while spin configurations are weighted with their Boltzmann factors based on the classical Hamilton function $\mathcal{H}[\vec{s}\,]$.

The prefactor of the action is $1/\hbar$, and the prefactor of the Hamilton function is $1/T$. Indeed $\hbar$ determines the strength of quantum fluctuations, while the temperature T controls the strength of thermal fluctuations. The classical limit $\hbar \to 0$ (only the path with the least action survives) corresponds to the limit $T \to 0$ (only the ground state contributes). A difference is, of course, that T is variable in Nature, in contrast to $\hbar$.

The kinetic energy $\frac{1}{2}((x_{i+1} - x_i)/a)^2$ in the path integral is analogous to the nearest neighbor spin coupling $\vec{s}_x \cdot \vec{s}_{x+\hat{\mu}}$ (where $\hat{\mu}$ is a vector in μ-direction with the length of one lattice unit). The potential term $V(x_i)$ is similar to the coupling $\vec{B} \cdot \vec{s}_x$ to an external magnetic field (or $\vec{B}_x \cdot \vec{s}_x$ to make it more general).[13]

The magnetization $\langle \vec{s}\,\rangle$ corresponds to the vacuum expectation value of an operator $\langle \mathcal{O}(x) \rangle$, also denoted as a *condensate* or *1-point function*, and the spin correlation function $\langle \vec{s}_x \cdot \vec{s}_y \rangle$ corresponds to the 2-point correlation function $\langle \mathcal{O}(x(t))\mathcal{O}(x(0)) \rangle$.

The inverse correlation length $1/\xi$ is analogous to the energy gap $E_1 - E_0$ (and hence to a particle mass in a Euclidean quantum field theory). Finally, the Euclidean time continuum limit $a \to 0$ corresponds to a second order phase transition where $\xi \to \infty$. The lattice spacing in the path integral is an artifact of our regularized description. We send it to zero at the end, and the physical quantities emerge asymptotically in this limit. In 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. Nevertheless, since ξ sets the relevant scale, the lattice spacing has to be measured as the ratio a/ξ, which vanishes in both cases. Hence the second order phase transition is indeed equivalent to a continuum limit. All these analogies are summarized in Table B.2.

[13]Of course, the potential and the vacuum expectation value of any operator are more general than the corresponding quantities that we mention for the spin models.

Quantum Mechanics	**Statistical Mechanics**
Euclidean time lattice	d-dimensional spatial lattice
elementary time step	crystal lattice spacing
particle position $x_i = x(t_i)$	classical spin variable $\vec{s}_x$
particle path $\{x_i\}$ $(i = 1 \dots N)$	spin configuration $[\vec{s}] = \{\vec{s}_x\}$ $(x \in \text{lattice})$
path integral $\int \mathcal{D}x$	sum over configurations $\int \mathcal{D}s$
Euclidean action $S_E[x]$	Hamilton function $\mathcal{H}[\vec{s}]$
Planck constant $\hbar$	temperature T
quantum fluctuations	thermal fluctuations
classical limit	zero temperature
kinetic energy $\frac{1}{2}(\frac{x_{i+1}-x_i}{a})^2$	nearest neighbor coupling $J\vec{s}_x \cdot \vec{s}_{x+\hat{\mu}}$
potential energy $V(x_i)$	external field energy $\vec{B}_x\, \vec{s}_x$
weight of a path $\exp(-\frac{1}{\hbar}S_E[x])$	Boltzmann factor $\exp(-\mathcal{H}[\vec{s}]/T)$
1-point function $\langle \mathcal{O}(x) \rangle$	magnetization $\langle \vec{s}_x \rangle$
2-point function $\langle \mathcal{O}(x(t))\, \mathcal{O}(x(0)) \rangle$	correlation function $\langle \vec{s}_y \vec{s}_x \rangle$
energy gap $E_1 - E_0$	inverse correlation length $1/\xi$
continuum limit $a \to 0$	critical behavior $\xi \to \infty$

Table B.2: *A dictionary that translates quantum mechanics into the language of statistical mechanics. The points x are located in physical space, whereas $\vec{s}_x$ is an unit vector in an abstract N-dimensional spin space (and the index x represents a site on some crystal lattice).*

B.6 Lattice Field Theory

So far we have restricted our attention to quantum mechanical problems and to statistical mechanics. The former were defined by a path integral on a 1-dimensional 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 paths $\vec{x}(t)$ of a 3-d quantum particle, we now integrate over all configurations $\phi(x)$ of a quantum field defined at any Euclidean space-time point $x = (\vec{x}, x_4)$.[14] Again the weight factor in the path integral is given by the action $S_{\mathrm{E}}[\phi]$. Let us illustrate this for a free neutral scalar field $\phi(x) \in \mathbb{R}$. Its Euclidean action reads

$$S_{\mathrm{E}}[\phi] = \int d^4x \, \left[\frac{1}{2} \partial_\mu \phi \partial_\mu \phi + \frac{m^2}{2} \phi^2 \right] . \qquad \text{(B.6.1)}$$

Interactions can be included, for example, by adding a $\lambda\phi^4$ term to the action, as we have seen before. The partition function for this system is formally written as

$$Z = \int \mathcal{D}\phi \, \exp(-S_E[\phi]) \ . \qquad \text{(B.6.2)}$$

Note that we have set $\hbar = c = 1$, *i.e.* from now on we use *natural units*, which we discuss in Appendix X. The physical units can be reconstructed at any point unambiguously by inserting the powers of $\hbar$ and c which match the dimensions. In natural units (excluding the Planck scale) we only deal with one scale, which can be considered either as `length` or `time`, or its inverse that corresponds to `mass` or `energy` or `momentum` or `temperature`.

The integral $\int \mathcal{D}\phi$ extends over *all field configurations,* which is a divergent expression if no *regularization* is imposed. One can make the expression mathematically well-defined by using the lattice regularization. Starting from well-defined terms is essential from the conceptual point of view. Moreover, this formulation extends to the interacting case, including finite field couplings. Thus it is also essential in practice, if the interaction does not happen to be small. That situation occurs in particular in QCD

[14]In contrast to Section 1.3, we now denote the Euclidean time as x_4, following the standard convention of field theory. In addition we adapt the usual convention to write only lower indices in Euclidean space, cf. Chapter 4.

at moderate or low energies, which dominate our daily life.

On the lattice, the continuum field $\phi(x)$ is replaced by a lattice field ϕ_x, which is restricted to the sites x of a d-dimensional space-time lattice of spacing a. Then the continuum action (B.6.1) has to be approximated by discretized continuum derivatives, such as

$$S_{\mathrm{E}}[\phi] = \frac{a^d}{2} \sum_x \left[\sum_{\mu=1}^{d} \Big(\frac{\phi_{x+\hat{\mu}} - \phi_x}{a} \Big)^2 + m^2 \phi_x^2 \right] , \qquad \text{(B.6.3)}$$

where $\hat{\mu}$ is the vector of length a in the μ-direction. This is the standard lattice action, but different discretized derivatives (with couplings beyond nearest neighbor sites) are equivalent in the continuum limit. The corresponding lattice actions belong to the same universality class.

The integral over all field configurations now becomes a multiple integral over all values of the field at all lattice sites,

$$Z = \prod_x \int_{-\infty}^{\infty} d\phi_x \exp(-S_E[\phi]) . \qquad \text{(B.6.4)}$$

For a free field theory the partition function is just given by Gaussian integrals. In fact, we can write its lattice action as

$$S_E[\phi] = \frac{a^d}{2} \sum_{x,y} \phi_x \mathcal{M}_{xy} \phi_y , \qquad \text{(B.6.5)}$$

with a symmetric matrix $\mathcal{M}$, which contains the couplings between the field variables at the lattice sites. We can diagonalize this matrix by an orthogonal transformation matrix Ω,

$$\mathcal{M} = \Omega^T D \Omega , \quad D = \mathrm{diag}\,(d_1, \dots, d_N) , \qquad \text{(B.6.6)}$$

where N is the number of lattice sites. We choose $\Omega \in SO(N)$, and with the substitution

$$\phi'_x = \Omega_{xy} \phi_y \qquad \text{(B.6.7)}$$

we arrive at (note that the Jacobian is $\det \Omega = 1$)

$$Z = \prod_x \int_{-\infty}^{\infty} d\phi'_x \exp\Big(- \frac{a^d}{2} \sum_x \phi'_x D_{xx} \phi'_x \Big) = \Big(\frac{2\pi}{a^d} \Big)^{N/2} \frac{1}{\sqrt{\det \mathcal{M}}} . \qquad \text{(B.6.8)}$$

To extract the energy eigenvalues of the corresponding (quantum) Hamilton operator, we study the 2-point function of the lattice field,

$$\langle \phi_x \phi_y \rangle = \frac{1}{Z} \int \mathcal{D}\phi \; \phi_x \phi_y \exp(-S_E[\phi]) \; . \tag{B.6.9}$$

This can be achieved elegantly by introducing a *source field* $j(x)$ in the partition function,

$$Z[j] = \int \mathcal{D}\phi \exp(-S_E[\phi] + j\phi) \; , \tag{B.6.10}$$

where we use the short-hand notation $j\phi = a^d \sum_x j_x \phi_x$. Similarly we are going to write below $\phi \mathcal{M} \phi = a^{2d} \sum_{x,y} \phi_x \mathcal{M}_{xy} \phi_y$, etc.

The connected 2-point function is given by

$$\langle \phi_x \phi_y \rangle_{\mathrm{c}} = \langle \phi_x \phi_y \rangle - \langle \phi \rangle^2 = \frac{\delta^2}{\delta j_x \delta j_y} \ln Z[j]|_{j=0} \; . \tag{B.6.11}$$

In our case $\langle \phi \rangle$ vanishes, *i.e.* $\langle \phi_x \phi_y \rangle_{\mathrm{c}} = \langle \phi_x \phi_y \rangle$. We eliminate the linear term in the exponent by another substitution

$$\phi' = \phi - \mathcal{M}^{-1} j \; , \tag{B.6.12}$$

so that the Boltzmann factor characterizing $Z[j]$ in eq. (B.6.10) is given by the exponent

$$-\frac{1}{2}\phi \mathcal{M} \phi + j\phi = -\frac{1}{2}\phi' \mathcal{M} \phi' + \frac{1}{2} j \mathcal{M}^{-1} j \; . \tag{B.6.13}$$

Performing now the functional integral over ϕ', we obtain

$$Z[j] = \left(\frac{2\pi}{a^d}\right)^{N/2} \frac{1}{\sqrt{\det \mathcal{M}}} \exp\left(\frac{1}{2}\, j \mathcal{M}^{-1} j\right) \; , \tag{B.6.14}$$

and from eq. (B.6.11) we infer

$$\langle \phi_x \phi_y \rangle = (\mathcal{M}^{-1})_{xy} \; . \tag{B.6.15}$$

It is instructive to invert the matrix $\mathcal{M}$ by going to Fourier space,

$$\phi_x = \left(\frac{a}{2\pi}\right)^d \int_B d^d p \; \phi(p) \exp(\mathrm{ipx}) \; , \quad (\text{where } \mathrm{xp} = \sum_{\mu=1}^{\mathrm{d}} \mathrm{x}_\mu \mathrm{p}_\mu) \; . \tag{B.6.16}$$

Due to periodicity, the momentum integration on the lattice is restricted to the (first) Brillouin zone

$$B = (-\pi/a, \pi/a]^d \ . \tag{B.6.17}$$

The impact of finite lattice spacing and finite volume is illustrated below. The virtue of momentum space is here that the action becomes diagonal,

$$S_{\rm E}[\phi] = \frac{a^d}{2}\Big(\frac{a}{2\pi}\Big)^d \int d^dp\, \phi(-p)\mathcal{M}(p)\phi(p) \ , \quad \mathcal{M}(p) = \hat{p}^2 + m^2 \ , \quad \hat{p}_\mu = \frac{2}{a}\sin\Big(\frac{p_\mu a}{2}\Big) \ . \tag{B.6.18}$$

$\mathcal{M}(p)$ must be periodic over B, which is achieved by the "lattice momentum" $\hat{p}$ (we adapt here a usual notation, but the "hat" does not indicate an operator in this case). A general 2-point function in momentum space reads

$$\langle \phi(q)\phi(p)\rangle = \mathcal{M}^{-1}(p)\, \delta(p+q) \ . \tag{B.6.19}$$

This is the lattice version of the (field theoretic) propagator in the continuum

$$\lim_{a\to 0}\ \langle \phi(-p)\phi(p)\rangle = \frac{1}{p^2+m^2} \ . \tag{B.6.20}$$

From the propagator (B.6.19) 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,

$$\phi(\vec{p}\,)_{x_d} = a^d \sum_{\vec{x}} \phi_{\vec{x},x_d} \exp(-{\rm i}\vec{\rm p}\cdot\vec{\rm x}) = \frac{\rm a}{2\pi}\int {\rm dp_d}\, \phi({\rm p}) \exp({\rm ip_d x_d}) \ , \tag{B.6.21}$$

and we consider its 2-point function

$$\langle \phi(-\vec{p}\,)_0 \phi(\vec{p}\,)_{x_d}\rangle = \frac{a}{2\pi}\int_{-\pi/a}^{\pi/a} dp_d\, \langle \phi(-p)\phi(p)\rangle\ \exp({\rm ip_d x_d}) \ . \tag{B.6.22}$$

Inserting now the lattice propagator we can compute this integral. We encounter poles in the propagator when $p_d = {\rm iE}$ with

$$\Big[\frac{2}{a}\sinh\Big(\frac{Ea}{2}\Big)\Big]^2 = \sum_{i=1}^{d-1} \hat{p}_i^2 + m^2 \ . \tag{B.6.23}$$

The integral (B.6.22) captures the pole with $E > 0$. Hence the 2-point function decays exponentially with rate E,

$$\langle \phi(-\vec{p}\,)_0 \phi(\vec{p}\,)_{x_d} \rangle \propto \exp(-E x_d) \ , \tag{B.6.24}$$

This allows us to identify E as the energy of the lattice scalar particle with spatial momentum $\vec{p}$. In the continuum limit we obtain the dispersion relation

$$E^2 = \vec{p}^{\,2} + m^2 \ . \tag{B.6.25}$$

At finite a, the lattice dispersion relation differs from the continuum result, *i.e.* we are confronted with lattice artifacts.

The lattice literature often uses "lattice units", where one sets the lattice spacing $a = 1$. In these terms, agreement with continuum physics is found in the limit where E, $\vec{p}$ and m are small. In particular m corresponds to the inverse correlation length $1/\xi$, and we discussed in Appendix A that $\xi \to \infty$ characterizes the continuum limit.

A free scalar particle has the same propagator in quantum mechanics and in classical mechanics. We see now the same property holds for the free scalar field, since eq. (B.6.25) is in accordance with the (classical) Klein-Gordon equation. Of course, this agreement is strictly limited to the free case.

Now that we have a safe continuum limit for the free field, we could proceed to the interacting case, *e.g.* to the $\lambda\phi^4$ theory. Keeping only the free part of the action in the exponent, the interacting part could be expanded as a power series in λ, if λ is small. These terms are then evaluated as n-point functions (n even). They are naturally decomposed into 2-point functions, as the above technique with the source derivatives shows. However, the corresponding propagators

$$\langle \phi_x \phi_y \rangle = \frac{1}{(2\pi)^d} \int d^d p \, \frac{\exp(\mathrm{ip}(\mathrm{y}-\mathrm{x}))}{p^2 + m^2} \tag{B.6.26}$$

diverge as they stand in $d \geq 2$. They can be treated by continuum regularizations like the *Pauli-Villars method:* it subtracts another propagator with a very large mass. This renders the propagators UV finite, and it maintains covariance. At the end the Pauli-Villars mass is sent to infinity. An alternative, which is more fashionable now, is *dimensional regularization,* that we are going to discuss in Chapter 4. However, one should keep in mind that these methods are restricted to perturbation theory, whereas the lattice provides a definition of the functional integral also at finite field couplings.

In lattice units all quantities appear dimensionless. To return to physical units one inserts the suitable power of a, such as m/a or $x\,a$. In practice, the physical value of a — and thus of all dimensional terms — is fixed by simulation measurements of one reference quantity, which is phenomenologically known. In QCD simulations the appropriate lattice spacings are in the order of 0.1 fm.

Exercises to Section 1.1

- **Klein-Gordon Equation**

Consider a neutral scalar field ϕ with the Lagrangian

$$\mathcal{L}(\phi) = \frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - \frac{m^2}{2}\phi^2 - \frac{\lambda}{4!}\phi^4 \ .$$

Apply the variational condition $\delta S = 0$ to derive the equation of motion of for field ϕ.

Assume $\lambda > 0$ and discuss for arbitrary $m^2 \in \mathsf{R}$ the stability of the constant solutions $\phi(x) = \phi_0$.

- **The non-linear σ-model**

The non-linear σ-model deals with a multiple scalar field $\vec{s}(x) = (s_1(x), \dots, s_n(x))$, $s_i(x) \in \mathsf{R}$, which obeys in each point x the condition $\vec{s}(x)^2 \equiv 1$. The Lagrangian reads

$$\mathcal{L}(\vec{s}) = \frac{1}{2}\partial_\mu\vec{s}\cdot\partial^\mu\vec{s} \ .$$

We now assume one component to dominate everywhere and denote it by σ, $\vec{s} = (\sigma, \underline{\pi})$, $\underline{\pi} = (\pi_1, \dots, \pi_{n-1})$, $1 \gtrsim \sigma^2 \gg \underline{\pi}^2$, and we count each small component as $|\pi_i| = O(\varepsilon)$.

Compute the Lagrangian $\mathcal{L}(\underline{\pi})$ up to $O(\varepsilon^2)$.

Derive for this (approximate) Lagrangian the equations of motion for the field components π_i.

Which are the symmetries of the actions $S[\vec{s}]$ and $S[\underline{\pi}]$?

How many generators have the corresponding groups of continuous field transformations ?

<u>**Exercises to Section 1.2**</u>

- **Free propagator**

A free, non-relativistic quantum mechanical particle of mass m moves on a line. Compute the transition amplitude from point position state $|x\rangle$ at time $t = 0$ to $|y\rangle$ at time $t = T$ with the path integral formalism, and check the result with the canonical operator formalism. Compare this result to the one obtained in the classical limit.

- **Quantum rotor**

Another non-relativistic quantum particle of mass m moves freely on a closed curve of length L. Which is the transition amplitude from some position eigenstate back to itself after time T ? Compute the energy spectrum of this particle.

Can you interpret this result, *e.g.* in view of a vibrating string ?

Hint: it is useful to express $\sum_{n\in\mathbb{Z}} \dots$ in terms of the integral $\int_{-\infty}^{\infty} d\alpha\, \delta(\alpha - n) \dots$, and then to apply the Poisson formula

$$2\pi \sum_{n\in\mathbb{Z}} \delta(\alpha - n) = \sum_{n\in\mathbb{Z}} \exp(2\pi \mathrm{i} n\alpha) \ .$$

- **Imaginary Gauss integral**

Investigate the imaginary Gauss integral

$$\int_{-\infty}^{+\infty} dx \, \exp(-\mathrm{i}\alpha \mathrm{x}^2) \ , \quad \alpha > 0 \ ,$$

by considering as the integration contour in the complex plane the polygon connecting the points $-R,\ R(-1+\mathrm{i}),\ \mathrm{R}(1-\mathrm{i}), \mathrm{R}$ with $R \gg 1$.

- **Dominance of the classical path**

We reduce the set of paths in a path integral to the trajectories, which can be parameterized by a single, real variable u. The action of such a path is $S(u)$, and the path integral is reduced to

$$Z_{\rm red} = \int_{-\infty}^{\infty} du \, \exp\Big(\frac{\rm i}{\hbar} S(u)\Big) \ .$$

This set of trajectories also contains the (unique) classical path at u_0.

Evaluate $Z_{\rm red}$ in the approximation, which neglects $O((u-u_0)^3)$, assuming $S''(u)|_{u=u_0} \neq 0$.

This approximation implies that the vicinity of the classical path (in units of $\hbar$) dominates the path integral. To justify this claim, compare the above approximate result to the contribution by some interval $u \in [a,b]$, which does *not* contain u_0. (Method: Substitute $\sigma = S(u)$ and assume $1/S'(u)$ to be differentiable in $[a,b]$.)

Exercise to Section 1.3

- **Random walk**

At time $t=0$ a point particle is located at position x_0. Now it starts to move on a 1-dimensional lattice of spacing a. In each time unit Δt it jumps over one lattice spacing, with equal probability for both directions.

Which is the differential equation that the probability distribution for the location of this particle, $P(x,t)$, obeys in the limit

$$a \to 0 \ , \quad \Delta t \to 0 \ , \quad D := \frac{a^2}{2\,\Delta t} = \text{constant ?}$$

How is it related to the Schrödinger equation ?

What does the condition $D =$ constant imply on the random walk in the continuum limit regarding continuity and velocity of the particle path ?

Exercises to Section 1.4

- **The CP^{n-1} model**

The two dimensional CP^{n-1} model is based on a field with n complex components,

$$z = \begin{pmatrix} z_1(x) \\ \cdot \\ \cdot \\ z_n(x) \end{pmatrix} \ , \quad z_i \in \mathbb{C} \ ,$$

with the constraint $z^\dagger z = 1$ in each point $x \in \mathbb{R}^2$.

The action reads

$$S[\vec{z}] = \int d^2x \, (D_\mu z)^\dagger D_\mu z \ , \quad D_\mu := \partial_\mu - z^\dagger \partial_\mu z \ .$$

Which is the global (x-independent) symmetry of this action ?

Show that there is in addition even a local (x-dependent) $U(1)$ symmetry.

- **Critical Exponents**

The simplest theoretical description of critical exponents is the *mean field theory* (or molecular field theory). It expands the magnetic field perceived by a specific spin as

$$\vec{B}_{\rm eff} = \vec{B} + a\vec{M} - bM^2\vec{M} \ ,$$

where $\vec{M}$ is the magnetic field due to the magnetization of the remaining spins. As a phenomenological input, we further use the *Curie Law* (Pierre Curie, 1859-1906)

$$\vec{M} = \frac{c}{T}\vec{B}_{\rm eff}$$

where a, b and c are constants.
(a) Neglect $O(M^3)$ and derive the prediction for the critical exponent γ.
(b) Now include $O(M^3)$ and derive predictions for β and δ.

Discuss the quality of this ansatz. What happens if you include $O(M^3)$ in the determination of γ ?

`[perhaps also determination of Ising model critical exponents in MFA]`

Exercises to Section 1.6

- **Natural units**

In particle physics one quantifies for instance the mass of the proton as $m_p = 938$ MeV. Identify the power of c and $\hbar$ which are needed to convert this quantity into kg. How much is m_p in kg ?

Which is the Compton wave length $\lambda\!\!\!^{-}_c$ of the proton in cm ?
(In natural units $\lambda\!\!\!^{-}_c = 1/m_p$).

The value of the electron mass of $m_e = 0.511003$ MeV expressed in American pounds has been referred to in the US congress as an argument for preserving this unit, guess why.

At an early stage of nuclear physics Hideki Yukawa made the following observation: he knew the pion masses ($m_{\pi^0} \approx 135$ MeV, $m_{\pi^\pm} \approx 140$ MeV) and assumed their wave function to be governed by the Klein-Gordon equation $(\Box - m^2)\Psi = 0$ (relativistic Quantum Mechanics of free, spin 0 particles). The stationary, spherically symmetric solutions take the form $\Psi(r) \propto \frac{1}{r}e^{-mr}$ (please check !). Explain why this inspired Yukawa to postulate that nuclear forces are due to pion exchange.

- **Connected 3-point function**

In the presence of a source field j the partition function of some quantum field theoretical model on the lattice reads

$$Z[j] = \int \mathcal{D}\phi \exp\Big(\mathrm{iS}[\phi] + \sum_{\mathrm{x}} \mathrm{j_x}\phi_{\mathrm{x}}\Big) \ .$$

Show that the connected 3-point function corresponds to

$$\langle \phi_x \phi_y \phi_z \rangle_{\mathrm{c}} = \frac{\delta}{\delta j_x}\frac{\delta}{\delta j_y}\frac{\delta}{\delta j_z} \ln Z|_{j=0} \ .$$

- **The lattice propagator for a free scalar field**

We have considered the case of a neutral scalar field $\phi_x \in \mathbb{R}$ on an infinite lattice of spacing a, and we computed its propagator.
(a) In which order of a is the corresponding dispersion relation plagued by lattice artifacts ?
(b) Plot the dispersion relations for masses $ma = 0$, 1 and 2 for the field on the lattice and in the continuum.
(c) We now allow for couplings of the field variable ϕ_x to $\phi_{x+\hat{\mu}}$ *and to* $\phi_{x+2\hat{\mu}}$. Construct a new discrete derivative in form of a linear combination of these couplings, such that the lattice artifacts in the leading order (identified in (a)) are eliminated. (This method is known as *Symanzik's program.*)

Exercises to Section 1.7

- **Grassmann integrals**

(a) Discuss in which sense functions of Grassmann variables can be partially integrated.
(b) How would you set up a δ-function for Grassmann integrals ?

- **Fermion determinant**

The components of $\Psi = \begin{pmatrix} \psi_1 \\ \cdot \\ \cdot \\ \psi_N \end{pmatrix}$ and of $\bar{\Psi} = (\bar{\psi}_1, \dots, \bar{\psi}_N)$ are Grassmann variables, and M is an $N \times N$ matrix (its elements are any complex numbers).

Show that the following equations hold

(a) $\displaystyle\int D\bar{\Psi} D\Psi \; e^{-\bar{\Psi} M \Psi} = \det\, M$,

(b) $\displaystyle\int D\bar{\Psi} D\Psi \; \bar{\psi}_i \psi_j \; e^{-\bar{\Psi} M \Psi} = (M^{-1})_{ij} \det\, M$, where $\displaystyle D\bar{\Psi} D\Psi = \prod_{i=1}^{N} d\bar{\psi}_i \, d\psi_i$.

Hints:
(a) Start from the special case $M = \mathbb{1}$ and generalize the result by means of a suitable substitution.
(b) Introduce external sources.

- **The Pfaffian**

We consider a set of Grassmann numbers $\eta_1, \dots, \eta_n$, and an anti-symmetric matrix $A = -A^T$. The term

$$\mathrm{Pf}A = \int d\eta_1 \dots d\eta_n \ \exp\Big(- \frac{1}{2} \sum_{i,j=1}^{n} \eta_i A_{ij} \eta_j \Big)$$

is the *Pfaffian* of the matrix A (this name refers to Johann Friedrich Pfaff, 1765 - 1825).

(a) Show that the Pfaffian vanishes if n is odd.

(b) Compute explicitly the Pfaffian for the cases $n = 2$ and $n = 4$. The results should be expressed in terms of the matrix elements A_{ij}, $i > j$.

(c) We now address the case $n = 2n'$ ($n' \in \mathbb{N}$) and matrices with the structure

$$A = \begin{pmatrix} 0 & a \\ -a^T & 0 \end{pmatrix} ,$$

where a is a $n' \times n'$ matrix. How are $\mathrm{Pf}A$ and $\mathrm{Det}\, a$ related ?

Appendix C

Group Theory of S_N and $SU(n)$

We will soon complete the formulation of the standard model by adding the gluons as the last remaining field, thus introducing the strong interactions which are governed by an $SU(3)_c$ gauge symmetry. To first familiarize ourselves a bit with the relevant group theory, we will now make a short mathematical detour. Once we add the strong interactions to the standard model, the quarks will get confined inside hadrons. In the so-called constituent quark model (which is at best semi-quantitative) baryons are made of three quarks, while mesons consist of a quark and an anti-quark. In the group theoretical construction of baryon states the permutation group S_3 of three quarks plays an important role. In general, the permutation group S_N of N objects is very useful when one wants to couple arbitrary $SU(n)$ representations together.

C.1 The Permutation Group S_N

Let us consider the permutation symmetry of N objects — for example the fundamental representations of $SU(n)$. Their permutations form the group S_N. The permutation group has $N!$ elements — all permutations of N objects. The group S_2 has two elements: the identity and the pair permutation. The representations of S_2 are represented by Young tableaux

$$\begin{array}{ll} \begin{array}{|c|c|}\hline &\\\hline\end{array} & \text{1-dimensional symmetric representation,} \\ \begin{array}{|c|}\hline \\\hline \\\hline\end{array} & \text{1-dimensional antisymmetric representation.}\end{array} \tag{C.1.1}$$

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

$$\begin{array}{ll} \begin{array}{|c|c|c|}\hline &&\\\hline\end{array} & \text{1-dimensional symmetric representation,} \\ \begin{array}{|c|c|}\hline &\\\hline &\multicolumn{1}{c}{}\\\cline{1-1}\end{array} & \text{2-dimensional representation of mixed symmetry,} \\ \begin{array}{|c|}\hline \\\hline \\\hline \\\hline\end{array} & \text{1-dimensional antisymmetric representation.}\end{array} \tag{C.1.2}$$

The representations of the group S_N are given by the Young tableaux with N boxes. The boxes are arranged in left-bound rows, such that no row is longer than the one above it. For example, for the representations of S_4 one finds

$$\begin{array}{|c|c|c|c|}\hline &&&\\\hline\end{array}\,, \quad \begin{array}{|c|c|c|}\hline &&\\\hline &\multicolumn{2}{c}{}\\\cline{1-1}\end{array}\,, \quad \begin{array}{|c|c|}\hline &\\\hline &\\\hline\end{array}\,, \quad \begin{array}{|c|c|}\hline &\\\hline &\multicolumn{1}{c}{}\\\cline{1-1} &\multicolumn{1}{c}{}\\\cline{1-1}\end{array}\,, \quad \begin{array}{|c|}\hline \\\hline \\\hline \\\hline \\\hline\end{array}\,. \tag{C.1.3}$$

The dimension of a representation is determined as follows. The boxes of the corresponding Young tableau are enumerated from 1 to N such that the numbers grow as one reads each row from left to right, and each column from top to bottom. The number of possible enumerations determines the dimension of the representation. For example, for S_3 one obtains

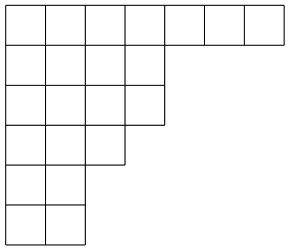
$$\begin{array}{ll} \begin{array}{|c|c|c|}\hline 1&2&3\\\hline\end{array} & \text{1-dimensional,} \\ \begin{array}{|c|c|}\hline 1&2\\\hline 3&\multicolumn{1}{c}{}\\\cline{1-1}\end{array}\,, \quad \begin{array}{|c|c|}\hline 1&3\\\hline 2&\multicolumn{1}{c}{}\\\cline{1-1}\end{array} & \text{2-dimensional,} \\ \begin{array}{|c|}\hline 1\\\hline 2\\\hline 3\\\hline\end{array} & \text{1-dimensional.}\end{array} \tag{C.1.4}$$

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

$$\sum_{\Gamma} d_{\Gamma}^2 = N! \ . \tag{C.1.5}$$

In particular, for S_2 we have $1^2 + 1^2 = 2 = 2!$ and for S_3 one obtains $1^2 + 2^2 + 1^2 = 6 = 3!$.

A general Young tableau can be characterized by the number of boxes m_i in its i-th row. For example the Young tableau

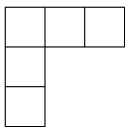


(C.1.6)

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

$$d_{m_1,m_2,\ldots,m_n} = N! \frac{\prod_{i<k}(l_i - l_k)}{l_1!l_2!\ldots l_n!}, \ \ l_i = m_i + n - i. \tag{C.1.7}$$

Applying this formula to the following Young tableau from S_5



(C.1.8)

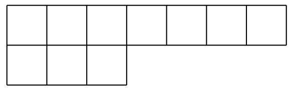
with $m_1 = 3$, $m_2 = 1$, $m_3 = 1$ and $n = 3$ yields $l_1 = 3 + 3 - 1 = 5$, $l_2 = 1 + 3 - 2 = 2$, $l_3 = 1 + 3 - 3 = 1$ and hence

$$d_{3,1,1} = 5! \frac{(l_1 - l_2)(l_1 - l_3)(l_2 - l_3)}{l_1!l_2!l_3!} = 5! \frac{3 \cdot 4 \cdot 1}{5!2!1!} = 6. \tag{C.1.9}$$

The permuted objects can be the fundamental representations of $SU(n)$. For $SU(2)$ we identify

$$\square = \{2\}. \tag{C.1.10}$$

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.



(C.1.11)

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

$$S = \frac{1}{2}(m_1 - m_2), \tag{C.1.12}$$

which is also denoted by $\{m_1 - m_2 + 1\}$. The above Young tableau hence represents $S = 2$ — a spin quintet $\{5\}$. Young tableaux with more than two rows have no realization in $SU(2)$ since among just two distinguishable objects no more than two can be combined anti-symmetrically.

C.2 The Group $SU(n)$

The unitary $n \times n$ matrices with determinant 1 form a group under matrix multiplication — the special unitary group $SU(n)$. This follows immediately from

$$\begin{aligned} &UU^\dagger = U^\dagger U = 1, \ \det U = 1. \\ &\det UV = \det U \det V = 1. \end{aligned} \tag{C.2.1}$$

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

$$(UV)^\dagger UV = V^\dagger U^\dagger UV = 1, \ UV(UV)^\dagger = UVV^\dagger U^\dagger = 1 \tag{C.2.2}$$

also holds. The group $SU(n)$ is non-Abelian because in general $UV \neq VU$. Each element $U \in SU(n)$ can be represented as

$$U = \exp(iH), \tag{C.2.3}$$

where H is Hermitean and traceless. The matrices H form the $su(n)$ algebra. One has $n^2 - 1$ free parameters, and hence $n^2 - 1$ generators η_i, and one can write

$$H = \alpha_i \eta_i, \ \alpha_i \in R. \tag{C.2.4}$$

The structure of the algebra results from the commutation relations

$$[\eta_i, \eta_j] = 2ic_{ijk}\eta_k, \tag{C.2.5}$$

where c_{ijk} are the so-called structure constants.

The simplest nontrivial representation of $SU(n)$ is the fundamental representation. It is n-dimensional and can be identified with the Young tableau $\square$. Every irreducible representation of $SU(n)$ can be obtained from coupling N fundamental representations. In this way each $SU(n)$ representation is associated with a Young tableau with N boxes, which characterizes the permutation symmetry of the fundamental representations in the coupling. Since the fundamental representation is n-dimensional, there are n different fundamental properties (e.g. u and d in $SU(2)_L$ and $c \in \{1, 2, 3\}$ in $SU(3)_c$). Hence, we can maximally anti-symmetrize 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

$$\begin{array}{|c|c|c|c|c|c|c|}
\hline
n & n+1 & n+2 & n+3 & n+4 & n+5 & n+6 \\
\hline
n-1 & n & n+1 & n+2 \\
\cline{1-4}
n-2 & n-1 & n & n+1 \\
\cline{1-4}
n-3 & n-2 & n-1 \\
\cline{1-3}
n-4 & n-3 \\
\cline{1-2}
n-5 & n-4 \\
\cline{1-2}
\end{array} \quad . \tag{C.2.6}$$

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

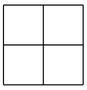
$$D^n_{m1,m_2,\ldots,m_n} = \frac{(n+m_1-1)!}{(n-1)!}\frac{(n+m_2-2)!}{(n-2)!}\ldots\frac{m_n!}{0!}\frac{1}{N!}N!\frac{\prod_{i<k}(l_i-l_k)}{l_1!l_2!\ldots l_n!}$$

$$= \frac{\prod_{i<k}(m_i - m_k - i + k)}{(n-1)!(n-2)!...0!}. \tag{C.2.7}$$

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

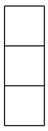
$$D^2_{m_1,m_2} = \frac{m_1 - m_2 - 1 + 2}{1!0!} = m_1 - m_2 + 1 = q_1 + 1 \tag{C.2.8}$$

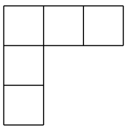
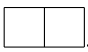
in agreement with our previous result. For a rectangular Young tableau with n rows, e.g. in $SU(2)$ for

 , (C.2.9)

all $q_i = 0$, and we obtain

$$D^n_{m,m,...,m} = \frac{\prod_{i<k}(m_i - m_k - i + k)}{(n-1)!(n-2)!...0!} = \frac{(n-1)!(n-2)!...0!}{(n-1)!(n-2)!...0!} = 1, \tag{C.2.10}$$

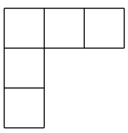
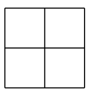
and therefore a singlet. This shows that in $SU(3)$  corresponds to a singlet. It also explains why the dimension of an $SU(n)$ representation depends only on the differences q_i. Without changing the dimension we can couple a representation with a singlet, and hence we can always add a rectangular Young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

 $\cong$ . (C.2.11)

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

$$\bar{m}_i = m_1 - m_{n-i+1}, \ \bar{q}_i = \bar{m}_i - \bar{m}_{i+1} = m_{n-i} - m_{n-i+1} = q_{n-i}. \tag{C.2.12}$$

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

 and  (C.2.13)

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

$$\ydiagram{3,1} \quad \text{and} \quad \ydiagram{2} \tag{C.2.14}$$

are anti-representations of one another, but

$$\ydiagram{3,1} \quad \cong \quad \ydiagram{2}\ . \tag{C.2.15}$$

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

$$D^n_{\bar{m}_1,\bar{m}_2,\ldots,\bar{m}_n} = D^n_{m_1,m_2,\ldots,m_n}. \tag{C.2.16}$$

For general n the so-called adjoint representation is given by $q_1 = q_{n-1} = 1$, $q_i = 0$ otherwise, and it is identical with its own anti-representation. The dimension of the adjoint representation is

$$D^n_{2,1,1,\ldots,1,0} = n^2 - 1. \tag{C.2.17}$$

Next we want to discuss a method to couple $SU(n)$ representations by operating on their Young tableaux. Two Young tableaux with N and M boxes are coupled by forming an external product. In this way we generate Young tableaux with $N + M$ boxes that can then be translated back into $SU(n)$ representations. The external product is built as follows. The boxes of the first row of the second Young tableau are labeled with 'a', the boxes of the second row with 'b', etc. Then the boxes labeled with 'a' are added to the first Young tableau in all possible ways that lead to new allowed Young tableaux. Then the 'b' boxes are added to the resulting Young tableaux in the same way. Now each of the resulting tableaux is read row-wise from top-right to bottom-left. Whenever a 'b' or 'c' appears before the first 'a', or a 'c' occurs before the first 'b' etc., the corresponding Young tableau is deleted. The remaining tableaux form the reduction of the external product.

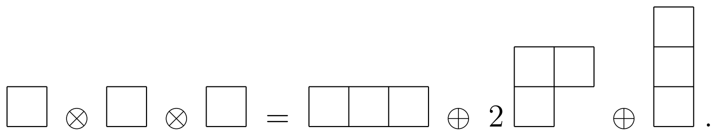
We now want to couple N fundamental representations of $SU(n)$. In Young tableau language this reads

$$\{n\} \otimes \{n\} \otimes \ldots \otimes \{n\} = \ydiagram{1} \otimes \ydiagram{1} \otimes \ldots \otimes \ydiagram{1}\ . \tag{C.2.18}$$

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

$$\{n\} \otimes \{n\} \otimes ... \otimes \{n\} = \sum_{\Gamma} d_{m_1,m_2,...,m_n} \{D^n_{m_1,m_2,...,m_n}\}. \tag{C.2.19}$$

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



(C.2.20)

Translated into $SU(n)$ language this reads

$$\begin{aligned} \{n\} \otimes \{n\} \otimes \{n\} &= \{\frac{n(n+1)(n+2)}{6}\} \oplus 2\{\frac{(n-1)n(n+1)}{3}\} \\ &\oplus \{\frac{(n-2)(n-1)n}{6}\}. \end{aligned} \tag{C.2.21}$$

The dimensions test

$$\frac{n(n+1)(n+2)}{6} + 2\frac{(n-1)n(n+1)}{3} + \frac{(n-2)(n-1)n}{6} = n^3 \tag{C.2.22}$$

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

$$\{2\} \otimes \{2\} \otimes \{2\} = \{4\} \oplus 2\{2\} \oplus \{0\}, \tag{C.2.23}$$

and in $SU(3)$

$$\{3\} \otimes \{3\} \otimes \{3\} = \{10\} \oplus 2\{8\} \oplus \{1\}. \tag{C.2.24}$$