Introduction to Particle Physics

Lecture in WS 2013/14 at the FSU Jena

Axel Maas

Contents

Chapter 1

Particle Physics

1.1 Introduction

Particle physics is the science of the smallest constituents of matter, and how they interact. In a sense, it evolved out of physical chemistry, where it was first resolved that all chemical substances, like molecules, are made out of set of chemical elements, of which we know currently roughly 120. At the turn of the 19th to the 20th century, it was then found that the atoms themselves were not elementary, but rather had constituents - a single nuclei, which was characteristic for the element, and a number of indistinguishable electrons. The latter number was in turn uniquely fixed by the nucleus.

In the early 20th century it was then found that the nuclei themselves are not elementary, but were made up out of just two types of constituents, the protons and neutrons, commonly denoted as nucleons. Again, and as will be described later in detail, these nucleons are not elementary, but are made up out of the quarks. At the current time, we do not know, whether either quarks or electrons do have a substructure, but substantial effort is invested to find out.

As is visible from this short historical remark, the study of elementary particles has changed subject many times over the course of time. Today, elementary particle physics is considered to be the study of the most elementary particles known to date, their interactions, and whether there could be even more elementary constituents. Today, such studies are inseparable inked to quantum mechanics, as quantum effects dominate the world of the elementary particles.

It is this field which will be presented in this lecture. After some more general remarks, the three known fundamental interactions, as well as the known particles will be presented. Their theoretical description together constitutes the standard model of particle physics. This theory is well established, and has been experimentally verified in the accessible range of parameters. But there are several reasons which prove that it can not be the final theory of particle physics. These will be briefly summarized at the end of the lecture. Thus, this lecture can only be taken as a snapshot of our current understanding of particle physics.

Suspiciously absent from these forces is gravity. The reason for this is twofold. On the one hand, so far nobody has found a reliable way how to quantize gravity, though this is a very active area of research. This quantum gravity is usually also considered to be part of particle physics. However, because of the complications involved in formulating it theoretically, there is not (yet) a standard model of quantum gravity, and just to describe the more promising candidates in a fair way is a lecture of its own, and rather speculative. The other is that also experimentally no particle has been yet observed which can be considered as an elementary particle of gravitation, the hypothesized graviton. Thus, quantum gravity is not yet part of the standard model of particle physics. This is probably the most stringent evidence for the incompleteness of the standard model of particle physics.

As the standard model of particle physics in its current form has been theoretically established at the turn of the 60ties and 70ties of the 20th century, with the final major theoretical touches added in the early 80ies, there are many excellent textbooks on its concepts. Especially, there are many theoretical textbooks on its foundations. The present lecture, however, is a more phenomenological introduction. For the concepts, this lecture is based on a number of older textbooks, but also some books on quantum-field theory have been used. However, there are numerous textbooks on the topic, with very different styles. Giving therefore a reasonable recommendation, especially in a field evolving as quickly as particle physics, is of little value. At the same time, identifying a suitable textbook to learn new concepts is a very helpful exercise for latter literature studies. Therefore, no general recommendations for this lecture will be given.

However, the latest experimental results for the standard model are from the year 2012, and two quantities still remain to be determined, as will be discussed below. Furthermore, the formulation of a theory is one problem. Determining all its consequences theoretically is a much harder problem, which is in many respects still unsolved today. We are far from the point where we can compute any given quantity in the standard model to any given accuracy. Furthermore, experimental results are never exact, and have errors. So, in many cases we know for deviations from the standard model only upper limits, and cannot exclude something is different. Thus, our knowledge of the standard model is in continuous motion. The best repository of the current knowledge is from the particle data group, which is accessible for free at pdg.lbl.gov. The most important discoveries (and information for this lecture) can also be obtained from my twitter feed, twitter.com/axelmaas.

Note that a lecture on particle physics without full-fledged quantum-field theory can necessarily often only give hints to the origin of many phenomena. The main aim is to make you acquainted with the general structure, with concepts, and with ideas. A full understanding of the details is necessarily relegated to quantum-field theory, and, in part, advanced quantum field theory courses. Wherever possible, I will try to give as much insights from these, without going into the technical details, skipping calculations for hand-waving arguments. As a result, with the knowledge of this course, you will be able to follow an experimental or theoretical overview talk at a mixed conference, but it will not be sufficient to follow a specialist talk in a parallel session, or likely an overview talk at a purely theoretical conference. Nonetheless, even in the latter cases, many of the ideas and underlying physics should sound then at least familiar to you. Eventually, the aim of this lecture is to prepare you for a full quantum-field theory course on particle physics.

1.2 Natural units

The usual units of the SI system, like meters or grams, are particularly unsuited for particle physics. The elementary particles live on very short time scales, move at high speeds, and tiny masses. E. g. typically energy and distance scales are

$$1 \text{ eV} = 1.60 \times 10^{-19} \text{ J} \tag{1.1}$$

$$1 \text{ fm} = 10^{-15} \text{ fm}, \tag{1.2}$$

which therefore have earned their own name, electron volts (eV), and Fermi (being identical to a femtometer). Typical energy scales are actually rather 10^9 eV, a GeV.

Furthermore, in particle physics the system of unit is further modified. Since units are man-made, they are not necessary, and the only thing really useful is a single scale to characterize everything. This is achieved in the system of natural units, which is obtained by setting

$$c = \hbar = k_B = 1$$

leaving only one unit. This is usually selected to be GeV, and thus everything is measured in GeV. E. g., times and lengths are then both measured in GeV^{-1}. In certain circumstances, GeV are exchanged for fm, which an be easily converted with the rule

$$\hbar c = 1 \approx 0.1973 \text{ GeV fm}$$

and thus energies are then measured in inverse fm.

These units are the standard units of particle physics. No relevant publication, textbook or scientist uses any different units, except for occasionally using a different order, like MeV, being 10^{-3} GeV, or TeV, being 10^3 GeV. Also other dimensionful quantities are then expressed in these units. E. g. Newton's constant is roughly 1.22×10^{19} GeV.

1.3 Elementary particles

From the point of view of elementary particle physics an elementary particle is any particle which, to the best of our knowledge, has no internal structure and is point-like. In quantum mechanics the Compton wave-length is often used to characterize the size of a quantum-mechanical object. Especially, this implies that an elementary particle is much smaller than its Compton wave-length. E. g., for an electron it is roughly of the order of its inverse rest mass, i. e. $(511 \text{ keV})^{-1} \approx 400$ fm. The experimental upper bounds for its size is 10^{-7} fm. Today, we know of 38 particles, which are elementary in this sense, and these will be introduced during this lecture.

On the other hand, there are composite particles, i. e. particles having an inner structure, which consists out of two or more of the elementary particles. Nucleons are an example for these. Of course, composite objects can consists out of composite objects, like nuclei, atoms, or molecules. Such composite objects are called bound states. In the course of history, many instances are known that particles previously recognized to be elementary have actually been composite. Today, nobody would be too surprised if any of the elementary particles known would, in fact, be composite. This has happened too often.

1.4 Fermions and bosons

Since there exist elementary particles which are massless, like the photon, they move necessarily with the speed of light. Thus, any adequate description of particle physics requires the use of special relativity, and only under very special circumstances is it possible to obtain reasonably accurate approximate results by using only non-relativistic physics. Besides this necessity, the union of quantum mechanics and special relativity, called quantum field theory, implies many remarkable and very general results.

One such result is the fundamental distinction of particle types into bosons and fermions. Bosons obey Bose-Einstein statistics; this implies that there can be arbitrary many of them in any given state of fixed quantum numbers. On the contrary, fermions obey Fermi-Dirac statistics and have to adhere to the Pauli principle: In any given state with fixed quantum numbers can only be a single fermion.

The spin-statistics theorem, which is a very basic property of all experimentally verified field theories, and of most hypothesized ones, implies a deep relation between the statistical properties of particles, the Lorentz group, and the spin of a particle. It essentially states that for any Lorentz-symmetric, i. e. special relativistic, quantum theory in more than two (one space and one time) dimensions every particle obeying Bose-Einstein statistics has integer spin, and any particle obeying Fermi-Dirac statistics has half-integer spin. There are no further possibilities for the spin values. Thus, bosons have integer spin, and fermions have half-integer spin. The electron, e. g., is thus a fermion, with its spin of 1/2. The photon, with its spin of 1 is therefore a boson.

Note that spin is an intrinsic property, like rest mass, of an elementary particle. It is not something which can be constructed from any properties, and has to be determined in experiment, at least inside the standard model. For a bound state, however, the spin can be computed as a function of the spins of its constituents.

A further important consequence is that the wave-function of a system with n particles has to be antisymmetric under the exchange of two fermions, while it is symmetric under the exchange of two bosons. This already illustrates the Pauli-principle. If there would be two particles of the same quantum numbers (and thus of the same type, since type is a quantum number) at the the same place, their wave-function must change sign under an exchange. At the same time, they are identical, and thus numerically it cannot change. Since the only quantity which at the same time changes and does not change sign is zero, any such state has to be zero when the particles are at the same place, and thus they cannot be - this is the Pauli principle. On the contrary, for bosons the wave-function remains unchanged under an exchange of two identical bosons.

1.5 Particles and anti-particles

Another fundamental consequence of quantum field theory is the existence of anti-particles, first noted, and experimentally verified, in the 20ies and 30ies of the 20th century. The basic statement is that for any particle there exists an anti-particle of the same mass and spin, but otherwise opposite (e. g. electric) charges. For the electron, there is a fermion with also spin 1/2, positive electric charge, and the same mass: The positron. Only if a particle has no properties except for mass and spin (like the photon) there is no anti-particle, as the anti-particle would be again the particle itself.

As a further consequence, if a particle and its corresponding anti-particle meet, they can annihilate each other. If, e. g., an electron and a positron meet, they can annihilate each other, which will result in two or more photons. This does not need to happen instantaneously. Electron and positron can also first form a bound state, positronium, which has a finite life-time. It decays because of the annihilation of the two particles, decaying into two or three photons, depending on how the spin of the electron and anti-electron had been aligned with respect to each other.

Since this is a very fundamental prediction of our understanding of particle physics, great experimental effort is invested to check whether particles and anti-particles really have the same properties, e. g. the same mass. Particular attention has been devoted to produce anti-atoms, and measure their spectrum, as this is a very sensitive test of this theoretical prediction. So far, no deviation has been found. Still, although the production of anti-matter is tedious, and yields are measured in individual atoms, rather than grams, these tests remain the most sensitive ones of the foundations of particle physics.

1.6 Interactions

The annihilation of electrons and positrons into photons is an example for an interaction, in this case an electromagnetic interaction. Writing on a black board or not falling through the floor are also examples of electromagnetic interactions. The earth orbiting the sun is an instance of a gravitational interaction. Generically, anything, except the Pauli principle, what makes a particle aware of the presence of another particle is classified as an interaction. This is called a dynamical effect. The propagation of particles, on the other hand, is just a kinematical effect.

Besides the knowledge of all the elementary particles, it is also necessary to know the interaction of them with each other to write down a theory describing them. Such an interaction is not necessarily restricted to connect only two particles. The maximum number known so far is a four-particle interaction, and any interactions involving more particles than four can be broken down to those with less particles. This does not mean that it is not possible to have more; theoretical, this is possible, but there is no experimental evidence that it is needed so far. Also, there can be interactions which appear at a coarse scale to be an interaction involving more than four particles, but on a fine scale this is not the case.

There are two particular properties of interactions, which deserve a special mentioning. One is that interactions are not totally arbitrary. Rather they only occur between specific particles. E. g. the aforementioned annihilations only proceeds with one electron, one positron, and two or more photons. It will not occur with two electrons and two photons, or one electron, a proton, and two photons. The reason are conservation laws, in this case the one of electromagnetic charge. Throughout many more conservation laws will be encountered, and any interaction will strictly respect any exactly conserved quantity.

However, mass is not a conserved quantity. As a consequence of the famous Einstein equation $E = m$ (note the natural units!), energy and mass can be freely converted into each other. Thus in the positron-electron-annihilation process, the total rest mass of both particles, roughly 1 MeV, is converted into the kinetic energy of the photons, which are themselves massless. On the other hand, colliding two particles with a large amount of kinetic energy will permit to create new particles, which have at most a total mass equal to the energy (and rest mass) of the original particles. This is of central importance to scattering experiments to be discussed next.

Before moving on to them, there is one more noteworthy kind of interaction. Quantum mechanics forbids to measure energy and time at the same time arbitrarily precisely[1], $\Delta E \Delta t \geq 1$. Thus, it is possible for a brief amount of time to have more energy in the system. For this duration, it is possible to create also particles with this additional energy, though they have to be destroyed at the end of this period. Since these live only for this short amount of time, they are not real, but are called virtual particles. Such particles are often denoted by an asterisk, *.

Such virtual pairs can either appear out of the vacuum, or can be emitted and reabsorbed by a moving particle. In both cases there have been many experimental confirmations of these processes, e. g. the Casimir effect for the first and the Lamb shift for the latter. Such virtual particle fluctuations are today used in precision experiments to access particles with masses larger than those which can be directly created using available energies. This will also be discussed later.

[1]Actually, this is not an elementary relation like the Heisenberg uncertainty principle, but follows from identifying particles with wave packets.

Chapter 2

Scattering experiments

The primary tool to understand particle physics today experimentally are scattering experiments, i. e. letting particles collide with each other. The reason has been allude to before: When two particles collide they interact, and their kinetic energy can be used to create new particles. The higher the energy, the more massive particles can be created. Furthermore, because of the DeBroglie relation that wave-length and momentum are related by $\lambda = 1/p$, only high energies permit to resolve very small structures, and therefore permit to investigate whether a given particle has a substructure.

Of course, due to the presence of virtual particles, it would in principle be possible to also investigate everything using low-energy collisions or even measurements of static properties of a particle. However, as will be seen later, the required precision to resolve new particles in most cases drops with increasing energy. Therefore, to access new particles in low-energy collisions requires very precise experiments. For some cases, particular effects reverse the situation. As a consequence, today both low-energy precision experiments and high-energy collisions work hand in hand together, both having their own importance.

The formulation of scattering experiments is thus central to particle physics. Almost all discoveries in particle physics have been made with such experiments. As a consequence also most modern theoretical tools have been developed and optimized to describe scattering experiments. Thus, at least an elementary understanding of scattering processes is a necessity in particle physics.

2.1 Non-relativistic

Many of the necessary concepts can already be understood non-relativistically, which offers a simpler formalism.

2.1.1 Elastic scattering

The simplest case is that of two incoming particles colliding with each other, and leaving as the outgoing particles unchanged. This case is known as elastic scattering. A very suitable coordinate system to describe this is the center-of-mass coordinate system, in which the total momentum of the two incoming particles is zero, $\vec{p}_1 = -\vec{p}_2$ with equal length p. Other coordinate systems, like the laboratory frame, with non-zero center-of-mass momentum can be obtained using a Galileo transformation (or later a Lorentz transformation), and therefore will not be discussed here further. Furthermore, if particles have a spin, things are also more complicated, but this will be avoided for now. Especially, in the following always bosons will be assumed, and the Pauli principle will not play a role.

The initial state is thus totally described by the type of particles and their properties, say two particles with masses m_1 and m_2, and their two momenta. Non-relativistically, energy and momentum conservation hold separately, and the initial state is completely characterized by the quantities

$$\begin{aligned} E &= \frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} \\ 0 &= \vec{p}_1 + \vec{p}_2. \end{aligned}$$

Important here is the absence of any potential energy. In scattering experiments, it is assumed than in the initial and final state the particles are so far separated that they do not influence each other. This is called asymptotic states. For electromagnetic and gravitational interactions, which decay as $1/r$, this approximation is certainly reasonable. It is also found to hold true for all particles and interactions, which can so far be prepared as initial states, or detected as final sates.

The final, or outgoing, state, has to have the same total energy, and the same masses. Thus, the total final momenta $\vec{q}_i$ has the same size, $q_i = p_i$, but may have a different direction. However, there is only one undetermined quantity, since the scattering is coplanar: There is a possible scattering angle, which can be defined between any incoming and outgoing momenta, e. g. $p_1 q_1 \cos\theta = \vec{p}_1 \vec{q}_1$. There are many conventions in use.

This scattering angle θ depends on the interaction with which the two particles scatter. Measuring it and comparing it to the prediction is one of the possibilities on how a theory can be falsified.

2.1.2 Cross section

In a quantum theory, a scattering process will be statistically distributed, and therefore the measurement of an individual scattering angle is meaningless. Thus, experiments are repeated many (many billion) times to obtain a distribution of scattering angles. The number of particles scattered into a solid angle $d\Omega = \sin\theta d\phi d\theta$ is then defining this cross section as

$$\sigma(\theta, \phi) = \frac{dN}{d\Omega}.$$

In practice, there are two quantities, which can be derived from the cross section, which are of particular importance. One is the total cross section

$$\sigma_t = \int \sigma d\Omega,$$

the other are differential cross sections

$$\sigma_x = \frac{\partial\sigma(\theta, \phi, x)}{\partial x},$$

where x can be any kind of variable, including also θ and ϕ themselves, on which the cross-section can depend. Also higher order differentials could be useful, so-called multiple-differential cross sections.

The unit of a cross section is an area. A useful unit to measure it in particle physics is barn, abbreviated b, corresponding to 100 fm^2. Although nowadays usual interesting cross-sections are of the order of nb to fb.

Predicting such cross sections is one of the most important tasks when one wants to test a given theory against modern particle physics experiments. However, this is quite an indirect process. The cross-section is calculated as a function of the parameters of a theory, which includes also the type and number of elementary particles. This result is then compared to the experimental result, an by comparison of how the cross-section is then depending on external parameters (e. g. the angle or the type of particles involved), the theory is either supported or falsified.

This becomes complicated in practice, if the cross section has contribution not only from the theory in question, but especially also from other known origins. E. g., when searching for yet unobserved physics, one has to subtract any contribution of known physics from the measured cross-section. Especially if the known contribution is the dominant one, this is difficult, as it is practically impossible to calculate it to arbitrary precision. It is also not possible to derive it from experiment; there is no possibility to continue unambiguously an experimental result from one case to another, without performing a theoretical calculation. Thus, this background reduction of known processes has become one of the major challenges in contemporary particle physics.

2.1.3 Luminosity

The definition of a cross-section yields immediately another figure of merit for a modern experiment, the luminosity. Technically, it is defined as the number of particles in two beams, N_1 and N_2, which interact in an (effective) area A with frequency f,

$$L = \frac{N_1 N_2 f}{A},$$

i. e. this is the number of processes per unit time and unit area. In modern particle physics experiments, the beam is usually bunched into n packages, which act like there would be several beams, and thus multiply the luminosity by n.

More important is the integrated luminosity,

$$\mathcal{L} = \int^{T} dt L,$$

that is how many collisions per unit area occur during an experiment of temporal length T. From this quantity, the number of events N occurring with a cross-section σ can be calculated as

$$N = \sigma\mathcal{L}.$$

Especially, to obtain one event occurring with cross-section σ during the time of the experiment, an integrated luminosity of $\mathcal{L} = 1/\sigma$ is necessary. This permits to directly estimate whether an experiment is able to measure a process at all. Typical values for integrated luminosity at the LHC are 25 fb^{-1} until the end of 2012, and at least 300 fb^{-1} until 2018.

A concept which will become important later on is the so-called partial (or later parton) luminosities. Assume that the beams consist not out of a single type of projectiles, but several. Than one can define for each possible pairing an individual partial luminosity.

2.2 Relativistic

2.2.1 Repetition: Relativistic notation

A non-relativistic description is, unfortunately, not sufficient in particle physics. There are three reasons for this limitation. One is that there are massless particles, which necessarily move with the speed of light, and therefore require a relativistic description. The second is that many regularities of particle physics are very obscure when not viewed from a relativistic perspective. Finally, in the reactions of elementary particles the creation and annihilation of particles, a conversion of energy to mass and back, is ubiquitous. Only (special) relativity permits such a conversion, and is therefore mandatory.

To just repeat the most pertinent and relevant relations, the starting point are the Lorentz transformations for a movement in the 3-direction,

$$\begin{aligned} x_1' &= x_1 \\ x_2' &= x_1 \\ x_3' &= \gamma(x_3 - \beta x_0) \\ x_0' &= \gamma(x_0 - \beta x_3) \\ \gamma &= (1-\beta^2)^{-\frac{1}{2}} \end{aligned}$$

which leave the length $x^2 = x_0^2 - \vec{x}^2$ invariant. The quantity x denotes the corresponding four-vector with components x_μ in contrast to the spatial position vector $\vec{x}$ with components x_i. Note that $\beta = |\vec{\beta}|$ is in natural units just the speed.

The scalar product of two four-vectors can be obtained using the metric $g = \mathrm{diag}(1, -1, -1, -1)$, i. e.,

$$x^2 = x_\mu x^\mu = x_\mu g^{\mu\nu} x_\nu,$$

i. e., the metric lowers and raises indices, which distinguish covariant and contravariant vectors. Though there are profound geometric differences between both, these have essentially no bearing on particle physics in the context of the standard model of particle physics.

From this follows the definition of the relativistic momentum and energy,

$$\begin{aligned} p &= \gamma\beta m \\ E^2 &= \vec{p}^2 + m^2 \end{aligned} \tag{2.1}$$

where m is the (rest) mass of the particle, an intrinsic and immutable (at least within known particle physics theories) property of any given type of particle. The concept of a relativistic mass, defined as γm is actually not necessary nor relevant in particle physics, since rest mass, and spatial momentum completely characterize the state of a particle. Note that for $\vec{p} = \vec{0}$ the relation (2.1) implies $E = m$, and thus rest mass and energy in the rest frame are exchangeable concepts. Especially, this relation implies that mass and energy can be freely converted into each other in special relativity.

An important special case are massless particles, which always move at the speed of light, i. e. $\beta = 1$. In that case, $E = |\vec{p}|$ precisely. Note that γ is neither well-defined, nor needed, when describing massless particles.

The relation (2.1) is only true for real particles. The virtual particles stemming from quantum fluctuations introduced earlier do not fulfill it. Such particles are called off-shell, since the relation (2.1) defines a surface in momentum space, called the mass-shell. Particles fulfilling (2.1) are consequently called on-shell.

One further consequence relevant to particle physics is time dilatation, i. e. the time (difference) τ' observed in a frame moving relative to the rest frame of the particle is longer than the eigentime τ in the rest frame as

$$\tau' = \gamma\tau. \tag{2.2}$$

Of course, this also implies that a moving particle experiences in its own rest frame a shorter time than outside. This will be very relevant for objects with a finite life-time: They exist the longer the faster they are. This has especially consequences for the experimental accessibility of massive particles.

Elastic scattering proceeds then in the relativistic case just as in the non-relativistic case. The only exchange is that the separate conservation of spatial momentum and energy is replaced by the conservation of four momentum,

$$p_1 + p_2 = q_1 + q_2.$$

Still, the identities of the particles are conserved in an elastic scattering, and therefore their rest mass is not changed. Also, there is still only one scattering angle characterizing this (also still coplanar) reaction.

However, an interesting new possibility are now inelastic scatterings where particles are produced or destroyed[1]

2.2.2 Inelastic scattering

In particle physics, it is no longer necessary that the particles scattered retain their identity after the scattering process. Especially, they can be transformed into a number of other particles, a so-called inelastic collision. Of course, the final products must still conserve 4-momentum. Thus,

$$p_1 + p_2 = \sum_n q_i.$$

[1]In principle, e. g. by breaking up into more than two fragments, there are also non-relativistic inelastic scattering processes possible, where the total mass is conserved. Though this plays an important role in, e. g., chemistry, this possibility will be ignored here .

This implies especially

$$\begin{aligned}\sqrt{\vec{p}_1^2+m_1^2}+\sqrt{\vec{p}_2^2+m_2^2} &= \sum_i \sqrt{\vec{q}_i^2+m_i^2}\\ \vec{p}_1+\vec{p}_2 &= \sum_i \vec{q}_i.\end{aligned}$$

In the center-of-mass frame, where the initial 3-momentum vanishes, it is thus in principle possible to transform the rest energy and the kinetic energy of the initial particles only into the rest energy of a number of new particles. This is called an on-resonance production. Especially, it is possible to create particles which are heavier than the original ones, provided the kinetic energy is large enough. This explains why higher and higher energies are required in particle experiments to obtain heavier and heavier particles.

Of course, this can be generalized to more than two incoming particles. However, in practice it is almost impossible in an experiment to coordinate three particles such that they collide simultaneously. Hence, this will not play a role in the following. It is not irrelevant, though. If the number of collisions becomes much larger than currently technically achievable, this can happen for an appreciable fraction of the events. Inside the sun such huge numbers are possible, and then three-body interactions become important for the thermonuclear processes. In fact, without a certain reaction of three helium nuclei to a carbon nucleus, the production of heavier elements in the sun would like not occur in the way it does in nature.

This also yields another important way of how to deconstruct cross sections. Assume a collision of two particles of type x as the so-called initial state. As a quantum mechanical process, the reaction will not only lead to a single outcome, a single final state, i. e. only one set of final particles. If it is possible, given the conservation of four-momentum and any other constraints, that there are more than one possible outcomes y_i, they all will occur. Their relative frequency is given by the details of both the reaction and the underlying interactions. These rates n_i for n final states yield the partial cross-sections,

$$\sigma_{2x\to i} = \lim_{n_i\to\infty} \frac{n_i}{\sum_i^n n_i}\sigma.$$

Partial cross sections can also be defined in a broader sense. Instead of having precisely defined particles in the final state it is e. g., possible to define a partial cross-section for n particles of any type.

Partial cross-sections play an important role in experiment and theory alike. Measuring a total cross-section is called an inclusive measurement, while identifying some or all final particles and their properties are called exclusive measurements. At the same time, theoretical calculations permit to make a judicious choice of a final state in which the background due to known physics is small or vanishing, thereby increasing the signal-to-noise ratio. Determining partial cross-sections is therefore a very important task.

2.3 Decays

In a sense a very particular type of inelastic cross-section is the decay of a single particle, i. e. a cross-section with a single initial particle, and two or more final particles. A decay occurs if a massive particle can, in principle, be converted into two or more lighter particles. For this to be possible, the rest mass of the original, or parent, particle must be larger than the sum of the rest masses of the daughter particle, plus possible a certain amount of energy to satisfy spatial momentum conservation. Furthermore, there may be additional constraints like that electric charge is conserved.

Since the decay of a particle is a quantum-mechanical process, it does not occur after a fixed decay time τ, but occurs at a time τ with an exponentially decaying probability, $\exp(-\Gamma\tau/2)$, where Γ is the so-called decay width. The value of the decay width is determined by details of the interaction, but is in most cases the larger the larger the difference between the rest mass of the initial particle and the total rest mass of the final particles is. Furthermore, the ratio Γ/m is used to characterize how unstable a particle is. If this ratio is small, the particle is rather stable, as it can traverse many of its Compton wave-lengths before decaying. Such particles are also known as resonances. If the ratio is close to or larger than one, the particle is not really behaving as a real particle, and there is no final consensus on whether such objects should be regarded as particles or not. Most particles well accessible in experiments have small widths.

However, quantum mechanically all possible outcomes can always be realized, just with differing probabilities. Therefore, if it is kinematically, and with respect to all other constraints, possible for a particle to decay into different final states, this will occur. Each possible outcome is called a decay channel, and is characterized by a corresponding partial decay width Γ_i, The sum of these widths together can be shown to yield the total decay width, which determines the life-time of the particle.

For the experimental measurement of the properties of resonances time dilatation (2.2) plays an important role. Since in experimental particle physics decay products are often produced at considerable speeds, their life-time in the experiment's rest frame is usually much larger than the life-time in their respective rest-frame. As a consequence, they can travel over considerable distances away from where the original particle has been produced, which permits to disentangle the processes of resonance production and decay. This also permits to much better determine the decay products, which in turn permit to identifying the particle which has decayed.

2.4 Cross-sections and particles

The formation of resonances yields, at least to leading order, a unique signature in the cross-section. Consider the situation that two particles of momenta p_1 and p_2 are collided, and react, elastically or inelastically, with each other, yielding two final particles with final momenta q_1 and q_2 with $p_1 + p_2 = q_1 + q_2$. If this process occurs such that the two original particles annihilate and form a resonance, which consecutively decays into the final products, it can be shown that the cross section behaves as

$$\sigma \sim \frac{1}{((p_1 + p_2)^2 - m^2)^2} \tag{2.3}$$

If the initial momenta are tuned such that the center of mass energy s fulfills $s = (p_1 + p_2)^2 = m^2$, the cross-section diverge, and thus the name resonance for the intermediate particles. Away from this condition, the cross-section quickly drops. Thus the formation of an intermediate resonance is signaled by a sharp peak in the cross-section. Since the relevant variable is s, this is also called an s-channel process.

In contrast, later processes will be encountered, where one particle can be transmuted into another by emitting a third particle. In these cases the relevant variables to be tuned are differently, either in a so-called $t = (p_1 - q_1)^2$ channel or, if the final particles are exchanged, in the $u = (p_1 - q_2)^2$ channel. These three variables together, s, t, and u are denoted collectively as Mandelstam variables. Observing a resonance in any of the channels is already providing insights into its production process, as the different kinematics can be shown to stem from different physical origins.

Of course, in nature there is not a real divergence in the cross-section, since any resonances has a finite, though possible small, width. Including this width leads to a Breit-Wigner cross section in, e. g. the s-channel

$$\sigma \sim \frac{\Gamma^2}{(\sqrt(s) - m)^2 + \frac{\Gamma^2}{4}}. \tag{2.4}$$

Thus, the divergences becomes a peak with a width determined by the decay width of the particle, which is therefore also accessible to an experimental measurement.

Of course, this is an idealized situation. Higher order processes or processes of a different origin can all contribute to the cross-section. Already a second resonance close by could distort the cross-section appreciably. Furthermore, quantum mechanically interference processes could suppress a peak or, even worse, create a fake peak. This has been observed, and is therefore not just an academic possibility.

Finally, if the process is inelastic, there could be more than just two particles in the final states. Then the cross-section becomes more involved, and the presence of resonances is not just a simple peak. There are advanced techniques to deal with such situations, like the Dalitz analysis, which now lead too far.

One should not that the formula (2.4) is only accurate if Γ/m is small. For resonances with a large width, the shape can be significantly distorted to the point where no real peak is observable at all.

2.5 Feynman diagrams

As has now been repeatedly emphasized, scattering experiments are the central experimental tool in particle physics. Accordingly, theoretical calculations of them have become the single most important technique. The usual task is therefore to have some initial state, e. g. two protons at the LHC, and then determine the partial cross-section for a given final state, inclusive or exclusive.

The most well-known tool for doing so is perturbation theory, in which the interactions between the particles is expanded in a power series of its strength. Though it is well-known that such an expansion is not able to capture all features of such a theory, there are very many experimentally accessible processes in which already low orders in perturbation theory, usually one to four, yield the dominant part of the relevant cross-section. The lowest order is under the assumption of quick convergence of the series, usually called leading order, abbreviated by LO. The next order is called next-to-leading order, NLO, and so on as N...NLO.

Such perturbative cross-section calculations in quantum field theories can be organized similar to quantum mechanics, i. e. as a mathematical series. Such a series can be graphically represented, which is called Feynman diagrams. In such diagrams a line corresponds to the propagation of a particle, a so-called propagator, while a vertex corresponds to an interaction. Thus, the number of vertices corresponds directly to the order of the series expansion. It can then be shown that each graph can be uniquely mapped to a mathematical expression describing the contribution of a given physical process to a cross-section. Thus one can use the mathematical tool of graph theory to construct all contributions of a given order to a physical process.

However, the possible vertices and propagators are not arbitrary, but are fixed for each theory in the form of so-called Feynman rules, which uniquely determine all the possible lines, vertices, and composition rules. These have to be determined for each theory anew, though this is an algorithmic procedure which a simple computer algebra system can perform. Especially vertices are restricted by conservation laws, like four-momentum conservation.

However, in practice for a reasonable complicated theory the number of diagrams grows factorial with the order of the expansion. Thus higher-order calculations require sophisticated methods to deal with the logistical problems involved.

Despite the power of these expression, it should always be kept in mind that already a hydrogen atom cannot be described with these methods.

Finally, it has turned out that also beyond perturbation theory a graphical language can very often be introduced in a mathematical precise way. Hence, Feynman-graph-like representations are ubiquitous in particle physics, though may mean very different mathematical objects. Especially, they are not necessarily restricted to a series expansion, and it is possible to encode also the full content of a theory graphically. Thus, one should always make sure what entities a given set of graphs corresponds to.

Chapter 3

A role model: Quantum electrodynamics

The first example of a particle physics theory, the first part of the modern standard model of particle physics, to emerge was quantum electrodynamics, QED. It already contains many of the pertinent features of the more complex standard model, while at the same time is much simpler. It is therefore the ideal starting point to get acquainted with a particle physics theory.

Quantum electrodynamics is describing the existence and interaction of electromagnetically active particles. As such, it is the quantum generalization of the classical electrodynamics, which already had been implementing special relativity.

3.1 Electrons and photons

QED, like any other particle physics theory, implements two roles for particles, though particles may later also implement both roles simultaneously.

The first role is that of charge carrier. Classically, these are objects which carry an electromagnetic charge. They therefore constitute the quanta of the electric current $j = (\rho, \vec{j})$. The prime example of these charge carriers are the electrons. These elementary particles are fermions, having spin 1/2, and have a rest mass of $m \approx 511$ keV. To the best of our current knowledge they are point-like[1], or at least smaller than 10^{-7} fm. Each electron carries exactly one unit of electric charge, which is in natural units dimensionless, and of a size of roughly $e = -0.3$. Its sign is chosen by convention, giving electrons a negative charge, in accordance with classical physics.

[1]The problem of infinite energy of point-like objects is resolved in quantum field theory, at least in an effective way. This will be discussed below in section 3.7.

The second role is that of the force carrier. Such particles are exchanged by charge carriers to transmit the force. The force carriers are thus the quanta of the electromagnetic field, which transmit electromagnetic interactions classically. Thus, the particles carrying the electromagnetic force are the quanta of the electromagnetic field, called photons. These are massless[2] bosons of spin one, and therefore so-called vector bosons.

In classical electrodynamics it is the electromagnetic fields $\vec{E}$ and $\vec{B}$, which are associated with the force. These fields can be derived from the vector potential A,

$$\begin{aligned} B_i &= \epsilon_{ijk}\partial_j A_k \\ E_i &= -\partial_i A_0 - \partial_0 A_i. \end{aligned}$$

Classically, this potential does not appear in the Maxwell equations

$$\begin{aligned} \vec{\nabla}\vec{E}(\vec{x}) = &\quad \partial_i E_i &= 4\pi\rho(\vec{x}) \\ \vec{\nabla}\times\vec{B}(\vec{x}) - \partial_t\vec{E} = &\quad \epsilon_{ijk}\partial_j B_k(\vec{x})\vec{e}_i - \partial_t\vec{E} &= 4\pi\vec{j}(\vec{x}) \\ \vec{\nabla}\times\vec{E}(\vec{x}) + \partial_t\vec{B}(\vec{x}) = &\quad \epsilon_{ijk}\partial_j E_k(\vec{x})\vec{e}_i + \partial_t\vec{B}(\vec{x}) &= 0 \\ \vec{\nabla}\vec{B}(\vec{x}) = &\quad \partial_i B_i &= 0 \\ \vec{\nabla}\vec{j} + \partial_t\rho = &\quad \partial_i j_i + \partial_t\rho &= 0, \end{aligned}$$

which describe the interaction between the electromagnetic fields and the sources. However, already in the quantum-mechanical Hamilton operator of a (spin-less) electron

$$H = \frac{1}{2m}(i\partial_i - eA_i)^2 + eA_0, \tag{3.1}$$

it is instead the vector potential that couples to the particle. Thus, the photons are actually the quanta of the vector potential A, instead of the E and B fields[3]. Classically, for any given $\vec{E}$ and $\vec{B}$ fields the corresponding vector potential is not uniquely determined. It is always possible to perform a gauge transformation

$$A_\mu \to A_\mu + \partial_\mu\phi, \tag{3.2}$$

[2]The concept of mass in quantum field theory is a non-trivial one. What precisely is meant by the statement that the photon has zero mass involves a significant amount of technical subtleties, which will be glossed over here. A full discussion requires a treatment of quantum field theory beyond perturbation theory.

[3]It is actually possible to write down a quantum theory also in terms of the $\vec{E}$ and $\vec{B}$ fields. However, such a theory is much more complicated, especially it contains non-local interactions, and is almost intractable analytically. Hence the formulation in terms of the vector potential is used essentially exclusively. It is called the localized version, since no non-local interactions are present.

with some arbitrary function ϕ. This is also true quantum-mechanically, except that also the wave-function ψ of the electron needs to be modified as

$$\psi \to \exp(-ie\phi)\psi. \tag{3.3}$$

Thus, the quanta describing electrons and photons are not uniquely defined, but can be altered. It can, however, be shown than any experimentally observable consequence of the theory does not depend on the the function ϕ, and thus on the choice of gauge. Theories with these features are called gauge theories. All theories, which have been experimentally supported, which we know today contain at least one gauge field. They therefore represent the archetype of particle physics theories. QED is just the simplest example for such a gauge theory. Later on, much more complicated gauge theories will appear, which eventually form the standard model of particle physics.

Since the gauge symmetry requires that the ordinary derivative in the Hamilton (3.1) is replaced by the combination of the ordinary derivative, ∂, and the gauge field A_i, $D_i = \partial_i - eA_i$, this combination is called the covariant derivative. The name covariant stems from the fact that the whole operator transforms in such a way under a gauge transformation, as to make the whole Hamiltonian gauge-invariant[4].

It is worthwhile to mention that the gauge symmetry is essentially what was called a redundant variable in classical mechanics. In classical mechanics, the usual first step is to eliminate all redundant variables by enforcing all constraints. This is not done in particle physics, and the presence of gauge transformations (3.2-3.3) is just a manifestation of this redundancy. Keeping this redundancy is, in contrast to classical mechanics, technically much more advantageous than eliminating it. The presence of this gauge degree of freedom is thus rather a mathematical convenience, than a genuine feature of nature.

As noted before in section 1.5, there are anti-particles to every particle, and especially the positron as an anti-particle to the electron. It is thus also a spin 1/2 fermion with the same mass, but opposite electric charge $e = 0.3$. The photon is uncharged. It is therefore its own anti-particle, and there is no anti-photon which could be distinguished by any means from the photon.

3.2 Formulation as a field theory

So far, the formulation at the level of (3.1) is quantum-mechanical. As noted, quantum mechanics cannot deal with relativistic effect, and one has to pass to quantum field theory

[4]Note here the distinction between gauge-invariance, i. e. unchanged under a gauge transformation, and gauge-independent, i. e. not transforming at all under a gauge transformation.

(QFT). QFT is technically much more complicated than ordinary quantum mechanics, and is beyond the scope of this lecture. However, it is very useful to understand and use the way QFTs are formulated, and this provides a quick way of representing a theory. Furthermore, significant insight can be gained already by treating the QFT classically, and thus just like classical electrodynamics, which is a classical field theory.

This should now be exemplified for the case of QED.

The simplest part is just starting with Maxwell theory, i. e. the theory of free electromagnetic fields. Since the Hamilton operator is not a Lorentz-invariant it is usually not the best way to formulate a relativistic theory. A better starting point is the Lagrange function L. For a field theory, this changes to a Lagrangian density $\mathcal{L}$, or briefly denoted by Lagrangian. Classical Maxwell theory takes the form

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{3.4}$$

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{3.5}$$

which is explicitly invariant under the gauge transformation (3.2). The anti-symmetric tensor $F_{\mu\nu}$ is the field-strength tensor, with components formed by the electromagnetic fields. The fact that the photon is a spin one boson can be read off from the Lagrangian since it carries a single Lorentz index, and therefore transforms like a vector, and is described by an ordinary field.

Adding fermions, and thus electrons, is subtle. Fermions obey the Pauli principle. There is no classical object, which does the same. To formulate a field theory for them, it was necessary to construct mathematical entities which encode this property, so-called Grassmann numbers. A Grassmann number ξ has the property that it squares to zero, i. e. $\xi^2 = 0$, and anti-commutes with any other Grassmann number η, $\xi\eta = -\eta\xi$. This looks rather strange at first,but can be dealt with quite straightforwardly in practice. Here, however, it is only necessary to remember that one cannot commute fermion fields, which are fields of Grassmann numbers, at will.

Spin 1/2 fermions, like electrons, are described by four complex Grassmann variables, which are collected into a column matrix ψ. The anti-particle can then be shown to be described by the field $\bar{\psi} = \psi^+\gamma_0$ with γ_0 a four-dimensional matrix, which is unitarily equivalent to $\mathrm{diag}(1,1,-1,-1)$. The total Lagrangian of QED then becomes

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}\gamma_\mu(i\partial^\mu - ieA^\mu)\psi - m\bar{\psi}\psi, \tag{3.6}$$

where m is the mass of the electron. The fields representing the electron, the so-called spinors ψ, transform under gauge-transformations like the quantum-mechanical wave-functions, and therefore this Lagrangian becomes gauge-invariant. These spinors are four-dimensional vectors of complex Grassmann numbers. Though they are four-dimensional they do not transform like vectors under Lorentz transformations, but differently, hence the name spinors. The details of this will not be relevant in the lecture, and are therefore skipped.

Furthermore, the matrices γ_i form together with γ_0 the Dirac matrices γ_μ, and fulfill the so-called Clifford algebra

$$\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}.$$

Note that despite the suggestive notations these matrices are invariant under Lorentz transformations, but play an important role in showing that this theory is, indeed, Lorentz-invariant. Nonetheless, the metric can be used to define γ-matrices with raised indices as well. The details o this are beyond the present scope. However, it is important to note that the quantity $\bar{\psi}\psi$ is a scalar and $\bar{\psi}\gamma_\mu\psi$ is a vector under Lorentz transformations. It is furthermore possible to define the matrix $\gamma_5 = -i\gamma_0\gamma_1\gamma_2\gamma_3$, which can be used to create a pseudo scalar $\bar{\psi}\gamma_5\psi$ and an axial vector $\bar{\psi}\gamma_5\gamma_\mu\psi$.

Besides the presence of photons and (massive) electrons and positrons another property of the theory can be read off the Lagrangian (3.6). There is a term which involves not two, but three fields, the interaction term $ie\bar{\psi}\gamma_\mu A^\mu\psi$. This is the first example of an interaction vertex, which specifies an interaction of three particles, an electron, a positron, and a photon. They interact with the coupling strength e, the electromagnetic coupling. The remainder factors $i\gamma_\mu$ ensure that this respects Lorentz invariance, and that it yields a meaningful quantum theory. Note that there is no term describing interactions with more than three particles in this Lagrangian. This will change later on. Furthermore, the relation between positron and electron requires them that a mass term includes both of them, and there are no separate mass terms for both particle species.

It is a further remarkable feature that the involved fields are functions not only of space, but also of time. Thus, a field configuration, i. e. any given field, represents a complete space-time history of a universe containing only QED.

As emphasized before, the Lagrangian is a classical object, i. e. the fields are not operator-valued. The quantization yields a QFT, a topic treated in a separate lecture. Here, it suffices to use the language of Lagrangians to specify a theory.

3.3 Indistinguishable particles

The formulation in terms of the Lagrangian (3.6) also answers the question why elementary particles cannot be distinguished in any way. This Lagrangian contains only a single electron field, a single positron field, and a single photon field. Each of these fields can describe an arbitrary number of particles of the corresponding type. Thus, since the corresponding particles are only excitations of the same field, they cannot be distinguished.

How can such particles then be identified at all? This requires to understand what a particle in the context of a field theory really means. A single particle is usually identified with an (almost) Gaussian excitation of the field[5], and therefore exponentially localized. Two separate particles are therefore two Gaussian excitations, where the two peaks are separated far compared to the widths of the peaks. Thus, the two particles can be identified, but since they are just the same excitation type of the field, they are nonetheless indistinguishable. This can be repeated for as many particles as desired, as long as the separation is large compared to the widths, creating many indistinguishable particles.

If the two particles come close to each other, the peaks overlap, and it no longer makes sense to speak of individual particles. Interactions play a role, and particles in the conventional sense only reemerge when the particles have moved far away from each other. In a perturbative description of a scattering process, the starting point are two such Gaussians sufficiently far apart as to ignore any remaining overlap. The same applies to the final states. That is what is called asymptotic states.

3.4 Gauge symmetry

The gauge transformations (3.2-3.3) have indeed a quite deeper structure than visible at first sight. Especially, they implement a symmetry, the gauge symmetry. The electron fields are modified by an arbitrary phase factor at every space-time point. This phase factor is a complex number, and therefore corresponds to a two-dimensional rotation. It is therefore an explicit implementation of the two-dimensional rotation group SO(2)≈U(1). As a consequence, the gauge symmetry is called a U(1) gauge symmetry, and the group U(1) is called the gauge group. Since the elements of U(1) commute, this is also called an Abelian gauge theory.

The electron field transforms like an element of this group, i. e. by multiplication. The electron is therefore said to be in a representation of the gauge group, the so-called fundamental representation.

The photon field does not transform by a phase factor, but instead by an additive function, which is related to the phase factor by exponentiation. It can be shown that these functions implement an algebra, the so(2)≈u(1) algebra. Therefore, the photon field is said to be in a different representation of the gauge group, the so-called adjoint

[5] As usual, the reality of field theory is far more complicated than this, but this picture is sufficiently close to the true mechanism to give the correct idea.

representation.

Such structures will again reappear for the other parts of the standard model, and are also present in gravity. So far, only the fundamental and adjoint representation play a role in particle physics, though several proposals exist for theories beyond the standard model in which particles in different representations would appear. They would then transform differently under gauge transformations. The precise form is given by the relevant gauge group, and is mainly an exercise in group theory.

For comparison, gravity can also be thought of as a gauge theory, with as gauge group the non-Abelian Poincare group SO(3,1)×T, where T is the (Abelian) translation group. This is, however, beyond the scope of this lecture.

3.5 Bound states and the positronium

Already the simple system of QED in form of (3.6) yields a plethora of interesting phenomena, which will be discussed in the following. The first one is that besides the electrons and photons[6] it is possible to construct bound states. The simplest bound state usually encountered is an atom. Lacking a proton or nuclei so far, the simplest bound state, which can be constructed in QED is positronium. This bound state consists out of an electron and a positron. In principle, it is just like a hydrogen atom, but its nuclei is just as heavy as the electron. This state has also been observed in experiments.

Of course, such a bound state is not stable. As stated, matter and anti-matter can annihilate each other, and so can electrons and positrons. Thus, after a while, the electron and the positron will annihilate each other into two or three photons (one is not permitted due to four-momentum conservation). The two different options depend on the relative alignment of the electron's and positron's spin, ether parallel or anti-parallel.

Thus, even such a simple theory as QED has already highly non-trivial structures.

3.6 Muons and Taus

Before delving further into the theoretical features of QED, it is time to introduce a property of nature which was not foreseen at the time QED was invented, and for which till today no convincing explanation has been found. This is the fact that the electron has two heavier siblings, and likewise does the positron. One is the muon, having a mass of about 105 MeV, and the other the tauon, with a mass of around 1777 MeV. Thus, both

[6] Again here some subtleties arise in a quantum-field theory when one talks about electrons or photons as particles. This is also beyond the scope of this lecture.

are much heavier than the electron. Otherwise their properties are the same as for the electron: They are fermions with spin 1/2 and negative electric charge. They can be easily included into the Lagrangian (3.6),

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \sum_f \bar{\psi}_f\gamma_\mu(i\partial^\mu - ieA^\mu - m_f)\psi_f, \quad (3.7)$$

where f now counts the so-called flavor, i. e. type, of the fermions, and can be electron, muon, and tauon. In QED, this flavor is conserved, i. e. it is not possible to convert the heavier muon into an electron. Later, additional interactions will be encountered, which change this. Then neither the muon nor the tauon are stable, and thus are encountered in nature only as short-lived products of some reaction.

The electron, muon, and tauon are otherwise exact replicas of each other. They form the first members of three so-called families or generations of elementary particles, which will have, except for the mass, otherwise identical properties. Their presence, and the relative sizes of their masses form one of the most challenging mysteries of modern particle physics, subsumed under the name of flavor physics. However, several interesting phenomena we observe in nature are only possible, if there are at least three such families, and some will be encountered later on.

To distinguish the electron, muon, and tauon from other particles to be discussed later, they are also collectively denoted as leptons.

3.7 Radiative corrections and renormalization

When one tries to perform perturbative calculations in QED, this is actually rather straightforward to leading order. However, a serious problem occurs if an attempt is made to go beyond tree-level. These were first encountered when people attempted to calculate corrections due to the emission and absorption of photons from electrons bound in atoms. Therefore, such contributions are nowadays collectively called radiative corrections, though in most cases this name is less than accurate.

The problem encountered in the calculation stems from the combination of two features of quantum field theories. One is that energy may be borrowed for some times in the form of virtual particles. The other is that quantum mechanically everything what could happen happens and mixes.

Therefore, in the next order of perturbation theory generically the possibility appears that an electron can emit a virtual photon and then reabsorb it later. Or a photon can split into a virtual electron-positron pair, which later annihilates to form again the photon. So far, this is not a problem. But these virtual particles can borrow energy, and therefore may be created with arbitrarily large energies. These arbitrarily large energies then induce divergences in the mathematical description, and all results become either zero or infinite.

This is, of course, at first sight a catastrophe, and for a while it seemed to mark the end of quantum field theory. It is actually no an artifact of perturbation theory, as one might suspect; though there are theories where this is the case, this is not so for those relevant for the standard model of particle physics.

Today, this failure has been understood not as a problem, but rather as a signal of the theory itself that it cannot work at arbitrarily high energies. It signals its own breakdown, and the theory is, in fact, only a low-energy approximation, or a low-energy effective theory, of the true theory. This is then as expected, since gravity is not yet part of the standard model, but is expected to play an important role at very high energies. Hence the prediction of failure is consistent with our experimental knowledge on the presence of gravity.

This insight alone does not cure the problems, however. Fortunately, it turns out that a certain class of theories exist, so-called renormalizable theories, where this problem can be cured in exchange for having a finite number of undetermined parameters. In case of the full QED Lagrangian (3.7) these are four: The strength of the electromagnetic coupling and the three masses. We exchange the divergences in favor of having these four numbers as an input to our theory, which we have to measure in an experiment. Once these parameters are fixed, reliable predictions at not too high energies can be made, where not too high for the standard model covers an energy range larger than our current experimental reach, probably much larger.

This exchange, a process known for historical reasons as renormalization, was the final savior of quantum field theory, making it again a viable theory. It also implies that a renormalizable theory cannot be a complete description of physics: The parameters of the theory cannot be predicted from inside the theory, and external input is needed. Whether a completely parameter-free theory can be formulated, which describes nature, is not known.

3.8 Fixing the inputs and running quantities

In the previous section 3.7, it was sloppily said that it is necessary to fix the parameters, in case of QED three masses and one coupling. This is a far less trivial process than it seems at first.

The problem is most easily seen when it comes to fixing the electric charge. To do so requires to define what is meant by the term electrical charge. Classically, this is simple. Together with the electron mass, it can be determined by the deflection of an electron in both an electric an a magnetic field. Quantum-mechanically, this is not so obvious. As stated, what is observed in nature is not just the electron. What actually looks like an electron, is an electron which is constantly emitting virtual photons, which themselves may emit virtual electron-positron pairs. It is this electromagnetic cloud, which surrounds the electron, what is measured. When the electron is probed at shorter and shorter distances, the probe travels farther and farther into this cloud. Therefore, what is measured, depends on the distance traversed, and thus the energy of the probe. Hence, it is necessary to make a definition of what electrical charge is, since it is not possible to measure only the process at tree-level.

For QED, this is usually done by measuring Thompson scattering. This is the scattering cross-section of a photon of small energy from an electron at rest. Aside from some trivial geometric and kinematic factors, the size of this scattering is determined by the fine-structure constant $\alpha = e^2/(4\pi)$, and thus by the electric charge. By fixing the parameters in the Lagrangian of the theory (3.7) such as to reproduce this scattering cross-section, one of the parameters is fixed, in this case to the value of roughly $1/137$.

This at first sight simple prescription has two inherent subtleties. The first is rather of a technical nature. For no relevant theory it is possible to calculate things like a scattering cross-section exactly. As a consequence, the dependence of the calculated cross-section on the parameters of the Lagrangian is only approximate, and therefore their determination can only be approximate. Moreover, two different calculations done with different approximations will not necessarily yield coinciding values for the parameters. It is therefore mandatory to make sure that any comparison has to be done in such a way as to take such approximation artifacts into account.

The second subtlety is far more complicated. It has just been defined what the electric charge is. However, it is not forbidden to instead define the electric charge by measuring the scattering cross-section for an electron and photons having all a certain energy, say μ. This is the so-called symmetric configuration. In general[7], the scattering cross-section will depend on this energy μ, and hence the so-defined electric charge will be different, and depend on the actual value of μ, $\alpha = \alpha(\mu)$. Thus, there is no unique definition of the electric coupling, but a coupling depending on the energy where it is measured. This is called a running coupling, a concept of great importance.

The change of the running coupling with the energy scale μ is also called the renormalization evolution, and it is possible to write down a mathematical group where the elements are changes in this scale. This is the so-called renormalization group, a very

[7]There are some theories, so-called conformal theories, where this is not the case. None of the theories realized in nature is of this type, though. This can actually be excluded exactly from experiment.

powerful tool when performing actual calculations in quantum field theory.

Similarly, the electromagnetic cloud of an electron will be deformed, if it is probed at higher and higher energies. As a consequence, the effective mass of an electron is also not independent of the energy[8], and the mass depends similarly on the energy $m_e(\mu)$, a running mass.

This running of quantities is a characteristic of quantum field theories. In fact, testing the energy-dependence of the running quantities is one of the most stringent tests, which can be performed on the validity of a theory. For the standard model, this has been done for the running over several orders of magnitude, a marked success of the theory.

It now may appear that one parameter has been traded in for a function. This is not the case. It is actually sufficient to measure the value of four running quantities at a single energy to fix all parameters of QED uniquely. The remainder of the functions is then uniquely determined by the theory and the four parameters. It is thus sufficient to probe a theory at a single energy scale, and it is then possible to make predictions about it at completely different energy scales.

A final word of caution must be added. In the present case, the electric charge was defined at the symmetric point, to obtain a single running coupling. There is no reason to make this choice. An equally well acceptable choice would be to chose the photon's energy twice as large as the one of the electron, and deduce everything from this configuration. In the end, physics must be independent of such choices. This is indeed what happens.

Such different choices are called renormalization schemes, and the change between two such schemes is called a renormalization scheme transformation. Though a mathematical well-defined process, it is important to compare only results in the same scheme, as otherwise the comparison is meaningless.

3.9 Landau poles

One of the mainstays of modern particle physics remains perturbation theory. However, perturbative calculations can, strictly speaking, never be entirely correct[9], and there are always non-perturbative contributions to any quantities at all energies. Still, in many relevant cases, the perturbatively calculable contribution is by far the dominant part.

One of the arguably most famous results of perturbation theory is the running of the

[8] Here, again, a subtlety of advanced quantum field theory plays a role when it comes to define what an electron is. This will also be skipped, as it is of little effective importance.

[9] Formally, a perturbative series is expected to break down roughly at order $\mathcal{O}(1/x)$, where x is the expansion parameter.

coupling $\alpha(\mu)$. It shows that the running coupling satisfies the differential equation

$$\frac{de}{d\ln\mu} = \beta(e) = -\beta_0 \frac{e^3}{16\pi^2} + \mathcal{O}(e^5), \tag{3.8}$$

where β is the so-called β-function, and β_i are its Taylor coefficients. The latter can be calculated in perturbation theory. Truncating the series at the lowest order, the differential equation can be integrated, and yields

$$\alpha(\mu^2) = \frac{e(\mu^2)^2}{4\pi} = \frac{\alpha(\mu_0^2)}{1 + \frac{\alpha(\mu_0^2)}{4\pi}\beta_0 \ln\frac{\mu^2}{\mu_0^2}} \equiv \frac{4\pi}{\beta_0 \ln\frac{\mu^2}{\Lambda^2}}, \tag{3.9}$$

where μ_0 is the energy where the experimental input $\alpha(\mu_0)$ has been obtained. The so introduced quantity Λ is often called the scale of the theory, and is a characteristic energy of the described interaction. The value of β_0 for QED is -4. This implies that the running coupling becomes larger with increasing energy, until it eventually diverges at a very large energy scale.

This divergence is the so-called Landau pole. Similar problems arise also in the other running quantities. The presence of such a Landau pole can be traced back to the use of perturbation theory. Before reaching the Landau pole, the running coupling, which is the expansion parameter, becomes larger than one, and therefore the perturbative expansion breaks down. Thus, a Landau pole is a signal of the breakdown of perturbation theory.

When this point is reached, it is necessary to resort to non-perturbative methods. Even for QED such a full non-perturbative treatment is not simple. The results, however, indicate that QED is also breaking down beyond perturbation theory, and the only quantum version of QED which make sense is the one with $e = 0$, a so-called trivial theory. This is the triviality problem. It is assumed that this effect is cured, once QED is embedded into a larger theory, in this case the standard model of particle physics. This is a recurring problem, and will be discussed in more details later.

Chapter 4

Invariances and quantum numbers

4.1 Symmetries, groups, and algebras

As has already been seen in section 3.4, symmetries and groups play an important role in particle physics. This becomes even more pronounced when eventually the whole standard model will be formulated. A technically convenient and tractable formulation of particle physics appears so far only possible by using gauge symmetries. However, these are no the only kind of symmetries which play an important role in particle physics. The most important ones will be discussed in the following sections.

Invariances have two important consequences. One is that the observation of an invariance strongly restricts how the underlying theory can look like: The theory must implement the invariance. The second is that if such an invariance is implemented, then as a consequence different processes can be connected with each other. E. g., one of the invariances to be encountered below is connected with the electric charge. As a consequence the scattering of two particles with given charges will have the same cross section as for the anti-particles with electric charge reversed. Hence, invariances are powerful tools in theoretical calculations, and powerful experimental tools to unravel the structure of the underlying theory.

An important classification of symmetries should be mentioned beforehand. In particle physics, there are three most important classes of symmetries. The first are discrete symmetries, like reflection in a mirror. They act, and do not involve any parameters. In contrast, continuous symmetries have transformations which depend on one or more parameters. If this parameter is constant, this is called a global symmetry. If it depends on space-time, as for the gauge transformations of QED, it is called a local or gauge symmetry.

Since symmetries imply the existence of an operator which commutes with the Hamilton operator, they lead in general to conserved quantum numbers. These can be either multiplicatively or additive. This means that a for a composite state the quantum numbers of the constituents are either multiplied or added. Discrete symmetries usually lead to multiplicative ones, while continuous ones to additive ones. Note that not all states have well-defined quantum numbers, even if a symmetry exists, and that a composite state can have well-defined quantum numbers even if the constituents do not have.

Furthermore, it is necessary to introduce a few notions of group theory. Later, in section 5.15, this will be further elaborated. For now, it is sufficient to introduce a few concepts. Most of these concepts are already familiar from quantum mechanics, especially from the theory of atomic spectra, and thus angular momenta. However, these concepts require generalization in the context of particle physics.

The basic property of a symmetry is that there exists some transformation t, which leaves physics invariant. Especially, such a transformation must leave therefore a state $|s\rangle$ invariant, up to a non-observable phase. Thus, a transformation has to be implemented by a unitary or anti-unitary operator, T. Especially in cases of continuous symmetries, it is possible to act on any state repeatedly with different values of the relevant parameter(s), $T(\alpha_1) = T_1$, $T(\alpha_2) = T_2$ yielding $T_2T_1|s\rangle$. There is also always a unit element, i. e. one transformation which leaves every state invariant. It is also furthermore for a unitary transformation possible to reverse it. Hence, such transformations have a group structure, and it is said that they are the corresponding symmetry group.

An example has been the group of the gauge transformations from section 3.4. There, the group was SO(2)≈U(1). The group elements were the phase rotations $\exp(i\alpha(x))$, with space-time-dependent parameters. It was thus a local symmetry. The unit operator is thus just the one, with $\alpha = 0$. The inverse one is just $\exp(-i\alpha(x))$, the complex conjugated one, as expected for a unitary operator.

Since the application of two gauge transformations commute, this symmetry group is Abelian. However, a consecutive application of more than one group element need not be commutative,

$$T_1T_2|s\rangle \neq T_2T_1|s\rangle.$$

If it is not commutative, the symmetry group is called non-Abelian.

The gauge field did not transform in the same way, see equation (3.2). In fact, the transformation did not seem to be unitary at all. It is, however, only the manifestation of a different mathematical version of the symmetry. Since interesting symmetry transformations are (anti-)unitary, they can always be decomposed into a complex conjugation or a one, for anti-unitary or unitary transformations, respectively, and a unitary transformation. Any unitary transformation can be written as[1]

$$T = e^{i\alpha_i \tau^i},$$

where the τ^i are Hermitian operators. These operator form the so-called generators of the symmetry group, and form themselves the associated symmetry algebra. Hence, the gauge fields transform not with the group, but with the algebra instead.

To separate both cases, it is said that the electron states and the gauge fields transform in different representations of the gauge algebra.

4.2 Noether's theorem

One of the central reasons why symmetries are so important in quantum physics is that it can be shown that any continuous symmetry entails the existence of a conserved current. Such conserved quantities are of central importance for both theoretical and practical reasons. Practical, because exploiting conservation laws is very helpful in constructing experimental signatures which are not too contaminated by known physics or making theoretical calculations more feasible. Theoretical, conserved currents mainly conserve quantum numbers. Since states of different conserved quantum numbers do not mix, they can be used to classify states.

The connection of symmetry and conserved currents is established by Noether's theorem. It is already a classical statement. Essentially it boils down to the fact that if the Lagrangian is invariant under a variation of the field, the variation can be used to derive a conserved current. To keep it simple, start with a Lagrangian $\mathcal{L}(\phi, \partial_\mu \phi)$ of a single real field ϕ, and its derivative $\partial_\mu \phi$. The Lagrangian is assumed to be invariant under the transformation

$$\phi \to \phi + \delta\phi,$$

i. e.

$$\delta\mathcal{L} = \mathcal{L}(\phi + \delta\phi, \partial_\mu\phi + \delta\partial_\mu\phi) - \mathcal{L}(\phi, \partial_\mu\phi) = \partial_\mu K^\mu, \tag{4.1}$$

where $\partial_\mu K^\mu$ is a total divergence, which is neither classically nor quantum-field theoretically able to alter the behavior of the system[2]. The corresponding equation of motions of the field ϕ are the usual Euler-Lagrange equations

$$\partial_\mu \frac{\delta\mathcal{L}}{\delta\partial_\mu\phi} - \frac{\delta\mathcal{L}}{\delta\phi} = 0. \tag{4.2}$$

[1]Note that this relation is in general only unique up to a discrete group.
[2]A least in flat space-time and for most quantum-field theories.

The explicit form of the variation (4.1) is then

$$\begin{aligned}\delta\mathcal{L} &= \delta\phi\frac{\delta\mathcal{L}}{\delta\phi}+(\delta\partial_\mu\phi)\frac{\delta\mathcal{L}}{\delta\partial_\mu\phi}\\ &= \delta\phi\partial_\mu\frac{\delta\mathcal{L}}{\delta\partial_\mu\phi}+(\delta\partial_\mu\phi)\frac{\delta\mathcal{L}}{\delta\partial_\mu\phi}\\ &= \partial_\mu\left(\delta\phi\frac{\delta\mathcal{L}}{\delta\partial_\mu\phi}\right),\end{aligned}$$

where the equation of motion (4.2) were used in the derivation and differentiations have been exchanged. This implies that the Noether current

$$j_\mu = K_\mu - \delta\phi\frac{\delta\mathcal{L}}{\delta\partial_\mu\phi},$$

or its corresponding extension to arbitrary sets of fields

$$j_\mu = K_\mu - \delta\phi_i\frac{\delta\mathcal{L}}{\delta\partial_\mu\phi_i} - \delta\phi_i^\dagger\frac{\delta\mathcal{L}}{\delta\partial_\mu\phi_i^\dagger}$$

is conserved, $\partial_\mu j^\mu = 0$. Hence, the associated charge

$$Q = \int d^3\vec{x}j_0,$$

is conserved and thus time-independent,

$$\frac{dQ}{dt} = \int d_V^3\vec{x}\partial_t j_0 = \int d_V^3\vec{x}\vec{\nabla}\vec{j} = \int\limits_{\partial V} d^2\vec{n}\vec{j},$$

where $\vec{n}$ is the unit vector on the surface ∂S, stemming from the application of Gauss' law[3]. Thus, the only change in the total charge is by the migration of charge carriers through the surface in which the charge is enclosed. Hence, if the surface is send to infinity, where no more charges can come from the outside, and the charge becomes time-independent.

In this way, conserved currents are linked to symmetries of the theory. This will be encountered throughout the standard model of particle physics. Another useful feature of symmetries in quantum-field theories is that they entail relations between different quantities, so-called Ward-Takahashi identities for global symmetries and Slavnov-Taylor identities for local (gauge) symmetries. This is a rather technical topic in detail, and will therefore not be addressed further here.

[3]Note that it was throughout tacitly assumed that the current and fields vanish sufficiently fast at infinity to permit the integral manipulations. In quantum-field theory, this is not obviously true anymore. But appropriate generalizations hold true in most cases, but not always, and care has to be taken. For the content of this lecture, this can be assumed to be true.

4.3 Electric charge

It is worthwhile to apply this mathematical exercise to derive the electromagnetic current of (Q)ED. There are now three terms, the variations of the fermions, anti-fermions, and the photon- field. In this case, there appears no surface current K_μ. The three terms are

$$\begin{aligned}\delta\psi\frac{\delta\mathcal{L}}{\delta\partial_\mu\psi} &= (-ie\psi)\times(i\bar{\psi}\gamma_\mu)\\ \delta\psi^\dagger\frac{\delta\mathcal{L}}{\delta\partial_\mu\psi^\dagger} &= (ie\psi^\dagger)\times 0\\ \delta A_\nu\frac{\delta\mathcal{L}}{\delta\partial_\mu A_\nu} &= (\partial_\nu)\times F_{\mu\nu}=0,\end{aligned}$$

where the transformation function ϕ has been dropped, as it is arbitrary, and all identities have to hold anyway for any arbitrary choice. The product $\times$ symbolizes the adequate composition rules for the various tensors involved. Note that the term for the gauge field vanishes because of the antisymmetry of the field-strength tensor. Thus the current is

$$j_\mu = -e\bar{\psi}\gamma_\mu\psi, \tag{4.3}$$

where use has been made of the fact that the fermion fields anticommute, and there is an implicit γ_0 in the anti-fermion field. As expected, the current is carried only by the fermion fields, and the photons do not contribute. It thus reproduces the expected form.

Thus electric charge is conserved in QED, as it is in classical physics. This will also not be altered when passing to the standard model of particle physics. However, there is a remarkable feature. Since the gauge-transformation of the photon field (3.2) is independent of the electric charge, it is in principle possible that the three generations, i. e. electron, muon, and tauon, could have different and arbitrary electric charges. Furthermore, there is no reason that the proton has the opposite and positive charge as the electron, as it is not made of of positrons, as discussed below. Still, all experimental results agree to very good precision with this equality, especially the latter one. In QED, there is absolutely no reason for this fact. A discussed later, the standard model of particle physics is only a consistent quantum-field theory if and only if all the electric charges fulfill certain relations.

However, the fact that a theory only works if certain experimental facts are taken into account is supporting the theory. But is does not dictate that the experimental facts have to be this way. If a small deviation of these relation would be observed tomorrow, it is the theory which has the problem. Therefore, the condition that the electric charges are as they are for the theory to work is not an explanation of why this has to be. It is an experimental fact which can be described, but not explained inside the standard model.

Though there are several proposals why this could be the case, especially so-called grand-unified theories. There is not yet any experimental support for any explanation, and it remains one of the great mysteries of modern particle physics.

4.4 Implications of space-time symmetries

Noether's theorem is not limited to so-called internal symmetries, i. e. symmetries which leave the space-time unchanged. It can also be applied to space-time transformations like translations and rotations. The consequences of both is that four-momentum and total angular momentum are conserved.

The spin of a particle is also a consequence of space-time symmetry, but in a much more subtle way. It is possible to show that the Poincare group only admits certain values of spin for particles, half-integer and integer[4]. This does not yet imply anything about the statistics which these particles obey. This requires further symmetries to be exploited later.

All this depends on having Minkowski space-time implementing the Poincare symmetry. These are two inputs to particle physics, and therefore the above listed properties are again not a prediction but a feature of particle physics. Other space-times can endow a theory with quite different properties, but for the subject of this lecture, it is fixed.

It also should be noted that the individual components of the spin have no meaning as conserved quantities in particle physics. This is most immediately visible when performing a Lorentz boost, under which they are not invariant. But for massless, and thus light-like particles, there is an exception. Since there is no way by making a Lorentz-boost to move into their rest-frame, the projection of the spin upon their momentum is always the same. This projection is called helicity, and is sometimes a useful concept in particle physics, when dealing with massless, or nearly massless, particles.

Note that space-time symmetries are global. Making them local leads to general relativity, which is not yet part of the standard model.

4.5 Parity

Another symmetry, closely related to space-time symmetry, is the discrete parity symmetry. The discrete transformation parity P is essentially a reflection in the mirror, i. e.

[4]In two space-time dimensions this is not the case, and the spin can take on any real number, leading to so-called anyonic particles. They are of no relevance in the standard model, but play a certain role in solid-state physics.

$\vec{x} \to -\vec{x}$. Time is unaltered. A distinction between space and time appears here, because of the signature of the Minkowski metric, which makes time and space (in a loose sense) distinguishable. Note that applying a parity transformation twice yields again the original system. In fact, a parity transformation inverts the sign of each vector, e. g., coordinate r or momenta p

$$Pp = -p.$$

Pseudo-vectors or axial vectors, however, do not change sign under parity transformation. Such vectors are obtained from vectors, e. g., by forming a cross product. Thus the prime example of an axial vector are (all kind of) angular momenta

$$PL = P(r \times p) = Pr \times Pp = r \times p = L.$$

As a consequence, fields can either change sign or remain invariant under a parity transformation[5]. Hence, a quantity can be assigned a positive or a negative parity, $+1$ and -1. Generically, when there are only these two possibilities, the quantities are called even or odd, respectively, in this case under parity transformations.

The fields describing particles have definite transformation properties under parity. It is necessary to define the parity of some states, to remove some ambiguities regarding absolute phases. Thus, the absolute parity of a particle is a question of convention. Usually, the electrons, and thus also muons and tauons, are assigned positive parity, while the photon has negative parity. Note that the anti-particles have the opposite parity for fermions, but the same for bosons. The reason is the different statistics.

In QED, parity is conserved. Classically that can be read off the Lagrangian (3.7), though there are some subtleties involved concerning fermions when performing the transformation. Furthermore, it it is important to note that also differential operators have definite transformation properties under parity transformation, and this is the same as the one of the photon field. As a consequence all terms have a definite positive parity, and the total Lagrangian has even parity. In principle, this could be changed in the quantization procedure, but this does not occur for QED. Later, additional interactions of particle physics will break this symmetry already classically.

4.6 Time reversal

Of course, the corresponding partner to parity is the discrete time-reversal symmetry T, i. e. the exchange of $t \to -t$, without changes to the spatial coordinates. This is

[5]For fermions, there is a multiplication by γ_0 involved.

physically equivalent to reversing all speeds and momenta, and as a consequence also angular momenta: Either objects move backward or the time moves backward. Time-reversal symmetry is anti-unitary, and therefore involves besides a factor of 1 or -1 also a complex conjugation of the object it acts upon. Similarly to parity, fields can be classified as being even or odd under time-reversal. However, because time-reversal is anti-unitary, there is no quantum number associated with it.

As with parity, time reversal symmetry is respected both classically and quantum-mechanical by QED, and again later on theories will be encountered which do not respect it.

4.7 Charge parity

In classical electrodynamics, it is possible to reverse the sign of all electric charges without changing the physics. This is a classical example of charge parity.

In particle physics, it is the presence of particles and anti-particles, and the fact that all their intrinsic (additive) properties are identical, that permits to define another discrete symmetry, the quantum version of charge parity C. Charge parity defines how a quantity behaves under the exchange of particles and anti-particles. Fields are also complex-conjugated, making charge-parity an anti-unitary operation. Note that properties like rest mass, momentum, spin, angular momentum or parity are not affected by charge parity.

It is again possible for a system to either change sign or not, defining even and odd charge parity. But it is also possible to not have a particular property at all, which is particularly true for charged state. That can be understood in the following way: To have a definite charge parity, a state must be an eigenstate under the charge parity transformation. However, charge parity transforms a particle with, say, electric charge +1 into a state with charge -1, and since such states are not identical, these states cannot have a definite charge parity. However, neutral states like the photon can have. It charge parity, e. g., is -1. Also, states made up from multiple charged states which are in total neutral can have a definite charge parity.

Again, QED respects charge parity, and again, later on theories will be encountered, where this is not the case.

4.8 CPT

It is now a very deep result that quantum-field theories on Minkowski space-time have the following property: A combination of charge parity transformation, parity and time reversal, i. e. the application of the operator CPT (or any other order, as they commute) will always be a symmetry of the theory[6]. Especially, causality is deeply connected with this property, and one implies the other.

This implies that if one mirrors physics and exchanges particles by anti-particles, and runs everything backwards there will not be any difference observed in any physical process.

From this, it can also be shown that necessarily all integer spin particles respect Bose-Einstein symmetry, while all half-integer spin particles will respect Fermi-Dirac statistics. The connection is made using special relativity: On the one hand it implies that only certain spins may exist, and at the same time it implies that certain properties under the exchange of particles have to be respected. Connecting this leads to the aforementioned classification. Therefore, in the context of particle physics, integer-spin particles are synonymously called bosons and half-integer spin particles fermions. Since this is a particular property of three-(or higher) dimensional quantum-field theory endowed with special relativity on flat Minkowski space-time, which is the arena of the standard-model, this association is fixed in particle physics. One must be careful, if one moves to a different field.

As a consequence of this so-called CPT-theorem the behavior of a state or field under one of the symmetries is fixed once the other two are given. Conventionally therefore only the quantum numbers under P and C are recorded for a particle, which leads to the J^{PC} classification, where J is the total, orbital and intrinsic spin, angular momentum of a particle, and P and C are only denoted as $+$ or $-$. E. g. a photon has $J^{PC} = 1^{--}$. The charge parity quantum number of particles with anti-particles of half-integer spin is not entirely trivial.

4.9 Symmetry breaking

Before continuing to further symmetries a brief intermission about breaking symmetries is necessary. Take a Lagrangian $\mathcal{L}_0$ with a symmetry at the classical level. There are now three possibilities how this symmetry may be affected.

The first is a somewhat obvious possibility, the explicit breaking. In this case an additional term is added to the Lagrangian, $\mathcal{L}_0 + \delta\mathcal{L}_b$, with δ a parameter, which does not obey the symmetry. Then , of course, already at the classical level the theory no longer

[6] There are some attempts, whether it is possible to formulate quantum-field theories which do not respect this, as experimentally there are only upper limits for any violation. However, so far none emerged with a convincing concept of how this really defines a reasonable field theory. It is, of course, possible that when going beyond the framework of quantum-field theory such possibilities may exist.

has the symmetry. It can be shown that such an explicit breaking can only be performed for global symmetries. Any attempt to do this for a local symmetry, though classically possible, cannot be quantized consistently.

The relevance of explicit symmetry breaking comes about in two special cases. One is that of soft symmetry breaking. In this case the terms in $\mathcal{L}_b$ are such that they are only taking effect in a certain energy regime, but become negligible in another one. The most common case will be seen to be mass terms, which break symmetries of most particle physics theories at low energies, but become irrelevant at large energies. Therefore, if the theory is probed at sufficiently large energies, the symmetry appears to be effectively restored.

The second possibility is, if δ is small. In this case the symmetry may not be exact, but relations due to the symmetry may still hold true approximately also in the full theory. In such a case it is said that the symmetry is only weakly broken. This concept plays an important role later for the strong nuclear interaction.

The next possibility is that a classical symmetry is no longer a symmetry of the quantized theory. Again, this only leads to a consistent quantum theory if the affected symmetry is a global one. Such an effect is also called an anomaly. In that case the symmetry is just not present at the quantum level. There are several examples to be encountered in particle physics. The simplest one is given by massless QED. The classical theory is invariant under a scale transformation, i. e. when rescaling all dimensionful quantities by a fixed number. This symmetry, the so-called dilation symmetry, is broken in the quantization. As a consequence, the theory has an intrinsic mass scale. This mass scale is e. g. the one encountered in the running of the coupling, see section 3.8. If the dilatation symmetry would be unbroken, the coupling would be independent of energy. Incidentally, it can be shown that this is deeply connected to the masslessness of the photon.

If a symmetry exists both classically and survives the quantization process, it may still be broken spontaneously. It can be proven that this cannot occur in usual particle physics theories for local symmetries, but only for global symmetries. This will manifest itself in the possible outcomes of the theory. E. g., if a symmetry predicts that two states should be the same, they will no longer be after spontaneous symmetry breaking. The spontaneous magnetization of a magnet below th Curie temperature is an often cited example for this phenomenon. Again, there is some fine-print in a full quantum-field theory[7], but the simpler picture will suffice for this lecture.

[7] E. g. the total magnetization of a magnet is always zero, because there is no preferred direction in space-time. Spontaneous magnetization manifests itself rather in the relative magnetization of two neighboring spins in the magnet. These are uncorrelated without spontaneous magnetization, but correlated in the case of spontaneous magnetization.

A more detailed treatment of what symmetry breaking really is will follow in section 5.7.

4.10 Fermion number

An inspection of the QED Lagrangian (3.7) shows that there is another symmetry. If the fermion field is multiplied by a constant phase $\exp(i\alpha)$ and the anti-fermion with $\exp(-i\alpha)$, the Lagrangian remains unbroken. This is a continuous U(1) symmetry, and therefore a Noether current has to exist. Calculating it, its is found to be proportional to the electric current (4.3). However, it is not the electric current itself, but is actually the current of fermion number, which just in the particular case of all fermions having the same electric charge coincides with the former.

Therefore, the number of fermions, i. e. the difference of the number of fermions minus the number of anti-fermions, is conserved. This does not require that these fermions remain of the same type. E. g. a state with one electron or with one muon both have the same fermion number, one. Any process respects this property. In fact, within the standard model this number of fermions is always respected.

Note that there is no similar condition for photons, and they can be created and destroyed at will, provided other conditions like conservation of total angular momentum and four-momentum conservation are satisfied. This is not so because they are bosons, but because they are their own anti-particles. Later on bosons will be encountered, which carry charge, and there charge conservation will restrict this possibility. Still then no analogue in the form of a boson number conservation exist.

4.11 Flavor

QED has, however, not only a single fermion species, but three ones, so-called three flavors. Multiplying only one of them with a phase still is a symmetry of the Lagrangian. Each of these three symmetries has its own conserved current. Therefore, the total number electrons, muons, and tauons are conserved in QED separately, it constitutes a flavor quantum number. This flavor quantum number is, like electric charge or fermion number, an internal symmetry, and therefore flavor is an internal quantum number. QED is said to respect flavor symmetry, where the flavors are uniquely identified by their mass. Formally, this is a $\mathrm{U}(1)^3$ symmetry, i. e. a product of three independent U(1) groups.

If the three flavors would be mass-degenerate, this symmetry would be enlarged. Since all particles have the same mass and the same electric charge, they become indistinguishable, and they can also be exchanged. Thus, it is possible to perform rotations in the internal flavor space. If the leptons would be described by real fields, this would be just the usual orthogonal group of three-dimensional rotations, O(3). But because they are complex, this group is enlarged to SU(3). The total flavor symmetry of mass-degenerate QED is hence SU(3), or, when supplemented by the additional fermion number symmetry, SU(3)×U(1).

4.12 Chiral symmetry

The QED Lagrangian (3.7) exhibits one interesting additional symmetry if the masses of all the fermions are set to zero, besides the then manifest flavor symmetry. This additional symmetry emerges by the combination of a flavor or fermion number transformation and an axial transformation. Axial transformations are a special property of fermions, and there is no analogue for bosons of arbitrary spin. The combination of fermion number and the axial transformation is the simplest case. It is mediate by multiplying every fermion field by $\exp(i\alpha\gamma_5)$, where α is a real parameter, and $\gamma_5 = -i\gamma_0\gamma_1\gamma_2\gamma_3$ is a combination of the γ-matrices. This can be shown using the fact that γ_5 anti-commutes with all γ_μ. The anti-fermion field is transformed by the corresponding hermitian conjugated phase factor.

This phase symmetry adds an additional U(1) symmetry to the theory, which is called an axial symmetry $\mathrm{U}_A(1)$. In addition, like the generalization of the fermion number symmetry U(1) to the flavor symmetry $\mathrm{SU}(N_f)\times\mathrm{U}(1)$ for N_f flavors, it is possible to enlarge the axial symmetry to an axial flavor symmetry, called the chiral symmetry. This name stems from the fact that it turns out that it connects fermions with spin projections along and opposite to their momentum direction, i. e. of different helicities. Since these projections yield classically a left-handed and right-handed screw[8], the name chiral, Greek for handedness, is assigned. The total symmetry of the theory is therefore $\mathrm{SU}(N_f)\times\mathrm{SU}_A(N_f)\times\mathrm{U}(1)\times\mathrm{U}_A(1)$ for N_f flavors of leptons.

Of these symmetries, the axial symmetry is actually broken by an anomaly during quantization. Non-zero lepton masses break the chiral symmetry, and the non-degenerate lepton masses then finally break the flavor symmetry just to a flavor number symmetry. Hence, little is left from the classical symmetries of massless QED.

This symmetry is therefore not realized in nature, where these masses are non-zero and different. Furthermore, since the masses are large compared to the strength of elec-

[8]Note that the question of what is left-handed or right-handed depends upon whether you look at a screw from the top or the bottom. Since different conventions on how to look at a screw are in use, care should be taken.

tromagnetic interactions, the symmetry is so baldy broken that it has almost no relevance for QED. This will change later when discussing the strong nuclear interaction.

4.13 Generators

A final remark on symmetries is the notion of generators, as it appears throughout the literature. It is a technically very useful observation that for a continuous symmetry, which changes $\phi \to \phi+\delta\phi$, the associated Noether charge Q plays a special role. It can be shown that

$$[\phi, Q]_{\pm} \sim \delta\phi,$$

where additional constants of proportionality may or may not appear, depending on conventions, and the $\pm$ indicates that this may be a commutator or anti-commutator, depending on whether the involved quantities are bosonic or fermionic, respectively. As a consequence, the charges Q are also referred to as the generators of the symmetry. In this sense, the electric charge creates the electromagnetic (gauge) symmetry.

Chapter 5

Strong interactions

It was very early on recognized that QED cannot be used to describe atomic nuclei, since QED can never sustain a bound state of the positive charge carriers which make up the nuclei, the so-called protons with their mass of roughly 938 MeV, and electrically neutral particles, the neutrons, with their just about 1.5 MeV smaller mass. Since also gravity was not an option, there had to be another force at wok, which is called the strong nuclear force. This became an ever more pressing issue, as it became experimentally clear that neither protons nor neutrons can be elementary particles, as they have a finite extension, about 1 fm, and behave in scattering experiments like having a sub-structure. To understand the details was, however, a rather complicated challenge.

The formulation of the strong interactions in its currently final form of quantum chromo dynamics (QCD) is, in a sense, the latest addition to the standard model. Because of the phenomenon of confinement, discussed below in section 5.8, it was much longer formulated than experimentally established. The reason behind is that the strong nuclear force is, as it names suggest, strong. It is therefore only in very limited, though relevant, circumstances possible to use perturbation theory to calculate anything in QCD. This made theoretical progress slow. The mainstay of QCD calculations today are in the form of numerical simulations, which for many problems in QCD physics has become the method of choice. Still, since computational power is limited, though such simulations being among the top ten contender for computation time in the world, it is not yet possible to calculate complex objects like even a helium nucleus. Still, for many purposes, many of the more fundamental features of the theory remain an area of active research today. In this context, QCD and theories which are (slightly) modified versions of QCD also serve as role models for generic strongly interacting theories in particle physics.

5.1 Nuclei and the nuclear force

As noted, all atomic nuclei consists out of differing numbers of protons and neutrons, the former having a positive electric charge of the same size as the electron and the latter no electric charge at all. The simplest atomic nucleus is a single proton. However, single neutrons are not stable, but decay in a few minutes, and are therefore only appearing bound in nuclei. Because of electromagnetic repulsion it became quickly clear that the nuclear force must be very much stronger than QED to create quite compact nuclei, not much larger than the protons and neutrons, with more than one proton. It must also act on different charges, as the neutron is electrically neutral.

One additional observation was made, when investigating nuclei and the nuclear force. It was not only because of its strength very different from QED and gravity, but it was also not of an infinite range. Though being much stronger than electrodynamics, it dropped almost to zero within a very short range of a few Fermi.

How can such a short-range force emerge? One suggestion comes from the relativistic description of bosons. The simplest example is the Klein-Gordon equation for a scalar boson of mass m. It is obtained when replacing in the relativistic energy-momentum relation (2.1) the energy and momentum canonically by $i\partial_t$ and $-i\partial_i$.

$$(\partial^2 + m^2)\psi = 0, \tag{5.1}$$

for the wave-function ψ. Investigating a static situation yields the solution

$$\psi(\vec{r}) = \frac{g}{4\pi|\vec{r}|}e^{-m|\vec{r}|}, \tag{5.2}$$

where g is an integration constant. Thus, the wave-function decays exponential towards large distances, where the characteristic distance of propagation is given by the inverse mass of the particle. Hence, a short-range interaction can arise if the exchange particles are massive. This lead to the prediction of an exchange boson for the nuclear force of order a few hundred MeV, given the range of the nuclear force of about a few fm. Such a potential is called a Yukawa potential.

The interpretation of the integration constant g becomes clear when the zero-mass limit is considered. In this case the wave-function becomes Coulomb-like, with g corresponding to the fine-structure constant. Thus, also in the massive case g can be interpreted as a coupling constant, which characterizes the strength of the strong nuclear force. It turns out that characteristic values for it are about two orders of magnitude larger than for the electromagnetic force.

Before taking this as more than a hand-waving motivation to look for a massive particle, one should add a word of caution with respect to the Klein-Gordon equation. It is, strictly speaking, only valid if the energy scales involved are small compared to the mass. Furthermore, it does not lead to a stable vacuum state, and therefore it can only describe physics relative to an (undetermined) vacuum. Both sicknesses are only cured when moving to quantum field theory, though this is beyond the scope of this lecture.

Nonetheless, the Yukawa potential yields the right idea. Instead of having massless photons as force particles, massive particles must mediate the strong nuclear force. They were indeed found in the form of the mesons.

5.2 Mesons

While the protons and neutrons are fermions with spin 1/2, the force carrier of the nuclear force were identified to be actually bosons. The lightest of them are the pions with quantum numbers $J^P = 0^-$, i. e. they are pseudoscalars. They come as a neutral one, π^0, and two oppositely charged ones, $\pi^\pm$. The range of the nuclear force is about 1 fm, which indicates that the mass of the force carrier, according to the Yukawa potential (5.2), should have a mass around 100-200 MeV. Indeed, the pions are found to have masses of 135.0 and 139.6 MeV for the uncharged and charged ones, respectively, and are thus much lighter than either protons or neutrons. These pions are not stable, but decay either dominantly electromagnetically into photons for the neutral one or like the neutron for the charged ones. Their life-time is of the order of 10^{-8} seconds and 10^{-17} seconds for the charged and uncharged ones, respectively. Therefore the charged ones live long enough to be directly detectable in experiment.

One of the surprises is that the neutral one decays into two photons, as usually photons are expected to couple only to electromagnetically charged objects. While this can be thought of as a neutral pion virtually splitting into two charged pions, and then annihilation under emission of photons, this is somewhat awkward. A more elegant resolution of this will be given in the quark model below in section 5.4.

With these pions it was possible to describe the overall properties of nucleons, especially long-range properties. At shorter range and for finer details it turned out that a description only with pions as force carriers was impossible. This was resolved by the introduction, and also observation, of further mesons. Especially the vector meson ρ with a mass of 770 MeV, spin one, and a very short life-time of roughly 10^{-24} seconds and the vector meson ω with a mass of about 780 MeV, but with a 20 times longer life-time than the ρ, play an important role. This larger number of mesons is also at the core of apparent three-body forces observed in nuclear interactions, which are, e. g., necessary to describe deuterium adequately. In fact, many more mesons have been discovered, and some more will appear later.

Describing how these various mesons create the strong nuclear force is in detail very complicated, but it in principle can be systematically performed, e. g. in the form of the so-called chiral perturbation theory. This will lead too far astray from particle physics itself, and will therefore not be detailed here. What is, however, remarkable is that out of nowhere appear several different mesons, all contributing to the nuclear force, and actually all of them also affected by the nuclear force. Such a diversity of force carriers is distinctively different from the case of QED, where only the photon appears.

5.3 Nucleons, isospin, and baryons

In the endeavor to find the carriers of the nuclear force, several other observations have been made. The first is that most nuclear reactions show an additional approximate symmetry, the isospin symmetry. This symmetry is manifest in the almost degenerate masses of the proton and the neutron. Both particles can therefore be considered to be a doublet of this new symmetry, which is similar to the flavor symmetry of QED, just much less broken because of the very similar masses. It is furthermore found that also the three pions fit into this scheme.

That once two particles and once three particles appear can be understood in the following way, very similar to the case of angular momentum or spin in quantum mechanics. There the rotational symmetry existed, and states could have any number of total spin l, and then there existed a degeneracy of the states into $2(l+1)$ states with different third components. Similar for the isospin, a state can have a particular isospin value, and then there are different charge states. In fact, isospin can be related to an SU(2) group once more, just like spin in quantum mechanics, hence the name. The proton and neutron then belong to the case of isospin I being $1/2$, and thus there are two states with different third components of the isospin I_3, which are essentially charge states. Such a situation is called a doublet, and therefore proton and neutron are sub-summed under the name of nucleons. Since the symmetry is only approximate, the different charge states are not degenerate. This is similar to the Zeemann effect in quantum mechanics, only that the mass now takes the role of the magnetic field.

The pions then belong to a state with $I = 1$, and thus three states. This is called a triplet, with approximately the same mass. Generically, such collections are called multiplets. In group theory, they correspond to different representations.

The natural question is then, whether there are higher representation, e. g. $I = 3/2$, with four states. Based on the fact that isospin seems to be related to electric charge, since the different states of an isospin multiplet all have different charge, and that half-integer or integer isospin corresponds to fermions or bosons, the properties of such a quadruplet should be predictable. In fact, the relation

$$Q = I_3 + S \tag{5.3}$$

seems to hold so far, where I_3 is the charge of the state, and S its spin[1]. For the proton and neutron, the assignments $I_3 = 1/2$ and $I_3 = -1/2$, respectively, yield the correct result. For the three pions, the I_3 assignments of -1, 0, and 1 reproduce the correct electric charge for π^-, π^0, and π^+.

However, this fails for the ρ meson, which has spin one, and is an isospin singlet, but in contrast to the rule (5.3) is uncharged. This can be remedied by replacing the spin S by $B/2$, where B is the so-called baryon number,

$$Q = I_3 + \frac{B}{2} \tag{5.4}$$

which is one for the nucleons and zero for the mesons. This is so far a phenomenological identification, but will become quite relevant in section 5.4. Of course, the anti-particle of the nucleons, the anti-proton and anti-neutron, carry negative baryon number.

According to this rule, it is possible to attempt to construct a quadruplet, having four states with $I_3 = -3/2, -1/2, 1/2, 3/2$. To get integer charges, it must then have a baryon number, like the nucleons. These particles should therefore have electric charge -1, 0, $+1$, and $+2$, and corresponding anti-particles. These particles have been observed experimentally, and again the different states have almost the same mass. They are called Δ, have masses of about 1232 MeV, and are fermions, as are the nucleons. However, their spin is 3/2. Since both nucleons and baryons carry baryon number, there are called commonly baryons, to distinguish them from the mesons. In fact, they are not the only baryons, and many more have been found experimentally.

Together, mesons and baryons are denoted by hadrons, and are identified as those particles directly affected by the strong nuclear force.

5.4 The quark model

The number of baryons and mesons found by now numbers several hundreds. Already decades ago, when only a few dozens were known, it appeared unlikely that all of them should be elementary. This was very quickly confirmed by experimental results which

[1]The spin will already in the next step be replaced by a different quantum number.

showed that the proton had a finite size of about 1 fm, and Rutherford-like experiments found that there are scattering centers inside the proton, which appeared point-like. However, in contrast to the atoms these constituents were not almost isolated in essentially free space, but very tightly packed. Furthermore, while the neutron is electrically neutral, it was found to have a magnetic dipole moment, a feature beforehand believed to be only existing if there is an electrically charged substructure present.

This evidence together suggested that the elementary particle zoo could possibly be obtained from simpler constituent and put into a scheme like the periodic table of chemical elements, which originates from just three different particles.

Playing around with quantum numbers showed a number of regular features of the hadrons. This gave rise to the quark model, where in the beginning two quarks were needed to explain the regularities observed, the up quark and down quark, abbreviated by u and d, as well as their anti-particles. Since both the bosonic mesons and fermionic hadrons must be constructed from them, it requires them to be fermions themselves. Since all of the hadrons have an extension, none of them can be identified with a single quark, just like the periodic table does not contain a single proton, neutron or electron. However, in contrast to the latter no free quarks are observed in nature, a phenomenon to be discussed below in section 5.8.

The simplest possibility to construct then a hadron would be from two quarks. This must be a boson, as two times a half-integer spin can only be coupled to an integer spin, and therefore a meson. Since no free quarks are seen, the nucleons must contain at least three quarks to get a half-integer spin. These considerations turn out to be correct. However, they lead to the conclusion that quarks cannot have integer electric charges. This is most easily seen by looking at the nucleons.

Scattering experiments identified that the nucleons have no uniform sub-structure, but have a two-one structure, that is two quarks of one type and one quark of the other type. Since it is found that the down quark is heavier than the up quark, the heavier one, i. e. the neutron, should have two down quarks. This yields a composition of uud for the proton and udd for the neutron. The only solution for the observed electric charges of the proton and the neutron are then an assignment of $2/3$ of the (absolute value of the) electron charge for the up quark, and $-1/3$ of the (absolute value of the) electron charge for the down quark. This consistently yields the required positive and neutral proton and neutron charges. This also explains the magnetic dipole moment of the neutron. At the same time, the baryon number of quarks must be $1/3$ for up quark and down quark. This implies also that the isospin of the up quark is $+1/2$ and that of the down quark is $-1/2$.

The pions are then constructed as a combination of a quark and an anti-quark, $u\bar{d}$ for the π^+, $\bar{u}d$ for the π^-, and a mixture of $\bar{u}u$ and $\bar{d}d$ for the π^0, i. e. the state of the π^0 is

$$\left|\pi^0\right\rangle = \cos\alpha\left|\bar{u}u\right\rangle + \sin\alpha\left|\bar{d}d\right\rangle,$$

where α is a mixing angle. This mixing angle can be experimentally or theoretically determined. Experimentally, this is possible by using the different decay patterns for different mixing angles. Theoretically, by calculating the left-hand side and the right-hand side in different bases. However, in practice both possibilities are highly challenging. An assignment of two quarks instead of a quark and anti-quark is not possible, as this cannot give the required baryon number of zero.

Particles like the ρ meson are then also combinations of a quark and anti-quark, but where the quarks have relative orbital angular momentum, creating their total spin of one. The Δ, however, turns out to pose a serious challenge.

5.5 Color and gluons

At first glance, the Δ appears simple enough. The double-positive state Δ^{++} is just three up quarks, and with decreasing charge always one up quark is replaced by one down quark, until reaching the Δ^- with three down quarks. To obtain the observed $3/2$ spin requires to align the spin of all three quarks. Of course, it could be possible that there would be a relative orbital angular momentum, but experimentally this is not found. In fact, there exists an excited version of the Δ with such an orbital angular momentum and total angular momentum of $5/2$, which is also experimentally confirmed.

And this is, where the problem enters. Since the Δ is a fermion, its wave-function must be totally antisymmetric. Since the spins are aligned and all three quarks are of the same type in the ground-state, no wave-function can be constructed which is anti-symmetric. Thus the existence of the Δ appears to violate the Pauli principle at first sight. But this is not so. Originally introduced to resolve that problem, and later experimentally verified, another conserved quantum number is attached to quarks: Color. The wave-function can then be anti-symmetric in this new quantum number, saving the Pauli principle and the quark model at the same time.

Since this new quantum number of the quarks is not observed for the Δ, or any other hadron, the hadrons must all be neutral with respect to this new quantum numbers. For the mesons, consisting of a particle and an anti-particle, this is simple enough, as just both have to have the same charge. This is not the case for baryons. Assigning just positive or negative charges, like the electrical charge, it is not possible to construct neutral states out of three particles. Attempts to do so with fractional charges also do not succeed in the attempt to make the proton and neutron color-neutral simultaneously. It is therefore necessary to depart from the simple structure of the electromagnetic charge.

As a consequence, it is assumed that there are three different charges, suggestively called red, green (or sometimes, especially in older literature, yellow), and blue. It is furthermore assumed that not only a color and the corresponding anti-color is neutral, but also a set of each of the colors is neutral. Then there are three quarks for each flavor: red, green, and blue up quarks, and red, green, and blue down quarks, totaling six quarks. A color-neutral baryon is then containing a quark of each color, e. g. a proton contains a red and a blue up quark, and a green down quark. In fact, since the total charge of a proton is zero, it is a mixture of any possible combination of color assignments to each three quarks, which are consistent with neutrality and the Pauli principle. Similar, the Δ^{++} now consists of a red up quark, a green up quark, and a blue up quark.

This construction is rather strange at first sight, but it can formulated in a mathematically well-defined way. This will be done below in section 5.16, after collecting the other ingredients of the strong interactions.

One other important ingredient, now that there is a new charge, is what mediates the force between the charges. In electromagnetism it was the massless photons. It is therefore reasonable to assume that there is also a mediator of the force between color charges. These were indeed found, and named gluons. As the photons these are massless[2] bosons with spin one. However, they differ from photons in a very important property. While photons are only mediating electromagnetic force, they are not themselves affected by it, since they carry no electric charge. But gluons carry color charge. In fact there are 8 different charges[3] carried by gluons, and none of these eight are either the quark charges, nor is their any simple relation to the quark charges. Especially, it is impossible to add a single quark charge with any combination of the gluon charges to obtain a neutral object. To achieve this, at least two quarks have to be added to one or more gluons.

Nonetheless, the idea of gluons has been experimentally verified, and they have been identified as the carrier of the strong interaction, binding quarks into color-neutral hadrons. The exchange of mesons to bind nucleons into nuclei can be viewed as a high-order effect of the gluon interaction. This is similar to Van-der-Waals force, though the details are different, as here not a color dipole moment enters, and the details are not yet fully resolved. Still, the effect can be traced back to the gluons.

---

[2]Again, the notion of mass is murky here, and the precise details are only becoming clear in a full non-perturbative treatment of the corresponding quantum field theory. However, for most purposes the notion of massless is sufficient.

[3]One can think of four of them as particles and four as anti-particles, though this is not really how it works, see section 5.16.

Hence, the combination of quarks, gluons, and colors, can explain the structure of all known hadrons, similar to the periodic table. Unfortunately, the strong force binding quarks by gluon exchange is not accessible using perturbation theory, at least when it comes to describing hadrons. Its treatment is therefore highly non-trivial. Because of the color, this underlying theory of hadrons is called chromodynamics, it quantum version quantum chromodynamics, or QCD for brief.

5.6 Chiral symmetry breaking

No word has yet been said about how the masses of the hadrons relate to the masses of the quarks. The mass of the proton is known very precisely to be 938.3 MeV, and the proton to be 939.6 MeV. This implies that the mass difference between up and down quarks must be tiny. The Δ is somewhat heavier, about 1230 MeV. This can be understood as an excited state, and is therefore heavier. Most ground state mesons have a mass of about 600 MeV or more. All this suggest a mass of about 300 MeV for the up quark and down quark, with very little difference[4]. But two mysteries appear. One is that the pions are very light, just about 140 MeV. The second is that any attempt to directly measure the quark masses yield consistently a mass of about 2.3(7) MeV for the up quark, and 4.8(5) MeV for the down quark. Though the difference is consistent, the absolute values are much smaller than the suggested 300 MeV from the nucleon properties.

The resolution of this puzzle is found in a dynamic effect of the strong interactions. To understand it, recall that massless free fermions have a chiral symmetry, see section 4.12. I. e. they can be multiplied by a phase factor $\exp(i\alpha\gamma_5\tau^a)$, where the τ^a are the Pauli matrices for two flavors of quarks. This symmetry is, at first sight, approximately valid for up and down quarks as their intrinsic masses (often called current mass) is very small compared to other hadronic scales, like the hadron masses. Thus, naively it is expected to hold. This chiral symmetry has a number of particular consequences. E. g. it implies that bound states of opposite parity, but otherwise identical content, should have the same mass. But the mass splitting between such bound states for hadrons is large. E. g., the parity partner of the nucleon is called the $N(1535)$ and has, as its name suggests a mass of 1535 MeV, about 50% heavier than the nucleons. This is much larger than expected due to the explicit breaking of the chiral symmetry because of the small current masses.

Thus, chiral symmetry must be much stronger broken than just from the current quark masses. The strong interactions must either spontaneously or explicitly break chiral sym-

[4]Though this difference is crucial for making the proton lighter and thus (more) stable (than the neutron), and therefore chemistry possible.

metry. Experiments show that at sufficiently high energies, many times the hadronic scales, the quarks again behave as if they have only their current masses. Thus, the breaking must be spontaneous. This will be confirmed when writing down the field theory of QCD in section 5.16. Hence, the strong interaction break chiral symmetry spontaneously.

This proceeds in the following way. The gluon interaction create a strong binding of the quarks, which creates a so-called vacuum condensate $\langle \bar{u}u \rangle + \langle \bar{d}d \rangle$[5]. Since the condensate is made from the quarks, any other quark will interact with it. Since the strong force is attractive, otherwise protons, neutrons, or nuclei would not exist, this will slow down any movement of a quark. Thus, quarks gain in this process inertial mass[6], and thus effectively a larger mass. This additional mass is the same for up quarks and down quarks, and approximately 300 MeV. This explains how the heavier mesons and the baryons gain a mass of this size. It does not yet explain two other facts: The lightness of the pions nor why this is no longer the case at high energies. These two questions will be answered in the next two sections.

Before this, some remarks should be added. It is easy to wonder how the universe should be filled with such a condensate, but no effect appears to be visible. Here, one notes how everyday experience can be deceiving. Of course, the effect is visible, since without it, all nucleons would be very much lighter, and so would be we, since almost all our mass is due to nucleons. Thus, whenever we feel our mass, we feel the consequences of this condensate. In fact, this condensate is responsible for essentially 99% of the mass of everything we can see in the universe, i. e. suns, planets, asteroids, and gas, the so-called luminous mass of the universe.

5.7 The Goldstone theorem

The lightness of the pions is a very generic feature of particle physics theories with spontaneous symmetry breaking. It is formulated in Goldstone's theorem. To lay out this theorem, it is helpful to investigate a very simplified model.

Take two scalar particles[7], which can interact with themselves. Such particles can either be described by two scalar fields ϕ_1 and ϕ_2 or with a single complex field $\phi = \phi_1 + i\phi_2$. Such a theory has a classical Lagrangian of

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial^\mu \phi - U(\phi^\dagger \phi), \tag{5.5}$$

[5]The omission of the state generally denotes that the expectation value is taken with respect to the vacuum.

[6]That this also acts as a gravitational mass is not at all clear. Only the equivalence principle implements this as a law of nature, an axiom.

[7]With only one scalar field it is not possible to construct a simple example with a continuous symmetry.

where U describes the potential for the field, which can only come from interactions between the two fields. The simplest (and in known particle physics so far exclusively appearing) possibility to obtain a symmetry for this theory is that the potential depends only the product $\phi^\dagger\phi$.

To obtain the simplest example for a symmetry, take the potential

$$U(\phi^\dagger\phi) = \frac{\mu^2 v^2}{4} - \frac{\mu^2}{2}\phi^+\phi + \frac{\mu^2}{4v^2}(\phi^\dagger\phi)^2.$$

The pre-factors, i. e. coupling constants, as well as the irrelevant constant term have been chosen judiciously such that the result will be looking simple. This potential, as well as the kinetic term, is invariant under the phase rotation $\phi \to \exp(i\alpha)\phi$. Therefore, this theory has a global (U(1)) phase symmetry. It is a bit odd theory, as the quadratic term is usually associated with a mass. This can be seen when determining the classical equation of motion, which read

$$0 = \partial_\mu \frac{\partial \mathcal{L}}{\partial\partial_\mu\phi^\dagger} - \frac{\partial\mathcal{L}}{\partial\phi^\dagger} = \partial^2\phi - \frac{\mu^2}{2}\phi + \frac{\mu^2}{2v^2}\phi^2\phi^\dagger. \tag{5.6}$$

The equation of motion for the field $\phi^\dagger$ is, up to conjugation, identical. Ignoring the interaction term, this is just again the Klein-Gordon equation (5.1), which describes the movement of a free scalar particle, if the sign of the second term would be positive. In this case, it looks like the particle has an imaginary mass, what would be called a tachyon. This is, however, not a problem. The sign of the quartic term is positive. Therefore, the energy is bounded from below classically, and the theory remains stable. It is therefore just an odd term in the potential energy.

To proceed, an interesting question is what the classical lowest energy state is. Since the kinetic term is positive, any spatial or temporal variation would increase the total energy. Hence, the state of lowest energy is necessarily a field ϕ_0 constant throughout space and time, which minimizes the potential (5.6). This constant is found to be $\phi_0 = ve^{i\theta}$, where θ is an arbitrary phase, i. e. all values of θ are a minimum. Thus, the solution manifold is highly degenerate. This is a consequence of the symmetry: Any change in θ can be offset by a symmetry transformation, without changing the physics.

One can proceed by specifying θ further. However, any choice is physically indistinct, and therefore arbitrary. For further calculations keeping θ manifest is awkward, and therefore in the following the explicit choice $\theta = 0$ is made. Since the choice is arbitrary, the symmetry is not violated. However, the symmetry is no longer manifest either. One therefore speaks of hiding the symmetry, or a hidden symmetry.

To make the situation more transparent, the next step is to shift the field ϕ by its vacuum expectation value, i. e. replace

$$\phi(x) \to v + \eta(x) + i\xi(x), \tag{5.7}$$

making now the complex structure manifest with the two real fields η and ξ. In this way, fluctuations of the fields around the classical vacuum can be studied. Inserting this into the Lagrangian (5.5) yields

$$\mathcal{L} = \partial_\mu\eta\partial^\mu\eta + \partial_\mu\xi\partial^\mu\xi - \mu^2\eta^2 + \frac{\mu^2}{v}\eta^3 + \frac{\mu^2}{v}\eta\xi^2 + \frac{\mu^2}{2v^2}\eta^2\xi + \frac{\mu^2}{4v^2}\eta^4 + \frac{\mu^2}{4v^2}\xi^4. \tag{5.8}$$

This Lagrangian shows now a number of very interesting features, which are very generic.

The first is that the two fields η and ξ behave differently. While there is a mass term, now with the correct sign, for η, giving it a mass of μ, there is no mass for ξ. Pictorially, one can think of η as excitations which describe fluctuations out of the minimum, while ξ, which is orthogonal to the direction of the chosen vacuum, moves between the different minima of the potential Since the vacua all have the same energy, this does not cost any energy, and therefore the mode is massless. This is a generic feature of such situations, and is known as Goldstones theorem. In a nutshell, it is the statement that there as many massless particles as there are directions in which the minima are equivalent[8].

The second is that there are now many different interactions between the fields η and ξ. However their couplings, i. e. their pre-factors, are not all different, but completely determined by the original parameters. The reason is that the symmetry is just hidden. To ensure that any symmetry transformation is still valid requires that the various interactions cannot have arbitrary pre-factors, because otherwise it would no longer be invariant under the symmetry transformation

$$v + \eta(x) + i\xi(x) \to e^{i\alpha}(v + \eta(x) + i\xi(x)), \tag{5.9}$$

where it should be noted that the vacuum solution v is also transformed accordingly. For that to work out, it is necessary to keep track when changing from (5.5) to (5.8), which occurrence of v stem from the original coupling constants in (5.5), and which from the shift (5.7), since only the latter are affected.

If at any point a term is added to the Lagrangian, which violates the symmetry, the symmetry becomes explicitly broken. The most obvious way is to add a mass term for the ξ field to the Lagrangian (5.8). Then, the Lagrangian is no longer invariant under the symmetry transformation (5.9). Of course, this can be translated back into the original

[8]The precise formulation is that there are as many massless particles as there are generators of the symmetry group minus the number of generators of symmetry transformations after all possible remaining choices have been made, here this is one minus zero, and thus one massless particles.

Lagrangian (5.5), where it takes the form of an additional quadratic term in the potential (5.6) of type $\Im(\phi)^2$. The effect is essentially that the potential is tilted, and the vacuum state has now a unique solution v, which has no longer an invariance. This gives also a physical explanation for the mass: Since there are no degenerate vacuum solutions anymore, any movement increases the energy.

This now illustrates how pions gain their small mass. The breaking of chiral symmetry would lead to a number of Goldstone bosons. In case of two quark flavors[9], there will be three Goldstone bosons, which are the three pions. These would be massless, if the quarks would be massless. However, because of the small current mass of the quarks, the symmetry is not exact, but rather explicitly broken. This gives the mass to the pions. That the masses of the pions are still large compared to current masses of the quarks is a dynamical effect. Approximately, the pion masses scale linearly with the current masses of the quarks, the so-called Gell-Mann-Oaks-Renner relation, but the pre-factor is large.

Before continuing on, a few words about subtleties and semantics must be said. When going to the quantum theory, quantum effects do what they always do, and they will mix all the degenerate vacuum states, and no vacuum will be preferred. Therefore, the vacuum state will exhibit a perfect symmetry, in contrast to the classical case. But the quantum system also carries the seed of the classical physics within, as it is metastable. If any arbitrarily small external perturbation, e. g. an infinitesimal mass for the ξ from some other physics process, arises, it will immediately have a unique vacuum state, in which the symmetry is no longer realized. Hence, though strictly speaking the system without external influence is perfectly symmetric, the presence of this metastability has lead to the expression that the symmetry is nonetheless spontaneously broken.

In fact, even though the symmetry is exact, a full non-perturbative calculation shows that the system has both an ordinary massive and a massless excitation, and hence the most pertinent feature of the Goldstone theorem are realized even with the symmetry present. Of course, in a perturbative calculation this will not show. Since perturbation theory only permits very small deviations from the vacuum state, the particles will still appear to be tachyons, as the relevant Lagrangian is (5.5). To cure this problem, one can introduce a weak external perturbation to the theory, which prefers a single vacuum, perform perturbation theory around this vacuum, i. e. using the Lagrangian (5.8) instead, and remove the external perturbation at the end[10]. Then, also in perturbation theory the system exhibits a massive and a massless particle. Especially because of this trick, which is extremely useful in the standard model, the exact notion of hidden symmetry is

[9]It can be shown that the number of colors does not matter.

[10]There are some subtleties associated with non-analyticities in this case, but in most cases they can be dealt with.

nowadays very rarely used, and almost always the situation will be denoted by spontaneous symmetry breaking.

5.8 Confinement and asymptotic freedom

This explains the light mass of the pion. It does not explain why the quarks appear to become lighter at high energies. The reason is that the strong interaction work very differently than electromagnetism or gravity. These two forces have the property that the potential[11] between two particles decays as $1/r$, and thus becomes weaker the longer the distance between the two particles is. Similarly, the force between two color-neutral particles, the residual effect of the strong interactions due to meson exchange, even decrease quicker because of the Yukawa potential (5.2) between them.

The situation for the interactions between two colored objects is drastically different. In fact, this was one of the reasons why it took very long to identify the underlying theory for the strong interaction, as its behavior is quite unique: The strong force becomes stronger with distance. An appropriate modeling of the inter-quark force is given by the potential

$$V = -\frac{4}{3}\frac{\alpha_s}{r} + \sigma r, \tag{5.10}$$

i. e. it has two components. There is a short-range Coulomb-like interaction, with the strong fine-structure constant $\alpha_s = g_s^2/(4\pi)$ defined by the strong coupling constant, similar to the electromagnetic force. But there is a second component, which rises linearly with distances, with a constant of proportionality σ. This latter contribution is of purely non-perturbative origin, which also made it very hard to discover its theoretical origin. That we do need such a potential for the strong interaction is an experimental fact. Only that α_s and k are not unrelated can be understood so far, but why this is necessary is just an observation.

This now explains why quarks appear to be lighter at high energies and thus shorter distances. At long distances, the interaction is very strong, which also manifests itself in chiral symmetry breaking and the condensation of quarks There the quarks are thus behave as described, and are quite heavy. However, this interaction becomes weaker at higher energies, i. e. quarks moving through the quark condensate at high enough energies will no longer be slowed down by it. The quarks will thus behave as if the only have their

[11]Note for the following that the notion of potential has no real meaning anymore when particles can be created or destroyed, and it is necessary to pass to quantum field theory for a more precise description. But for the present purpose the analogy is sufficiently good as to give a qualitative correct picture, and is even quantitatively surprisingly good.

current mass. This explains the difference of both masses. This effect that the strong interaction becomes weaker at higher energies is called asymptotic freedom. Of course, the remaining force is still as strong as the electromagnetic one, due to the classical Coulomb potential in (5.10), but this is much weaker than at long distances, and therefore effectively free. Moreover, quantum corrections reduce this interaction further: The running strong coupling decreases at high energies, because the leading coefficient of the β-function β_0 in (3.9) is negative, with a scale of roughly 1 GeV. Hence, at a very high energy, the electromagnetic force becomes actually stronger than the strong one, above roughly 10^{15} GeV.

The potential (5.10) also explains why no individual quarks are seen, at least to some extent. Since the potential energy between two quarks rises indeterminate when increasing the distance, one needs enormous amounts of energy when trying to separate them, much more than available for any earthbound experiment for any experimentally resolvable distance. This energy is stored in the gluonic field between the two quarks. Investigations have shown that this field is highly collimated between the two quarks, it is like a string connecting them both. Since the tension of this string is characterized to lowest order by the parameter σ in (5.10), this parameter is therefore also called string tension. This explains[12] why colored quarks are not observed individually, but are confined in hadrons, and the phenomenon is denoted as confinement.

However, in practical cases the energy limits of experiments is not a real concern. Long before the relevant energy limit is reached, there is so much energy stored in the string between the two quarks that it becomes energetically favored to convert it into hadrons. Hence, when trying to separate the two quarks inside, e. g., a meson by pumping energy into it, at some point it will just split into two mesons. This is called string breaking or screening, and is a 100% efficient mechanism, i. e. there is no remaining quark. The reason is that color, like electric charge, is conserved. Hence, overall the system must remain color-neutral. Thus, there would be at least two colored objects left over, which would again create immediately a string. This is also the mechanism why gluons are not observed. When trying to separate gluons from a hadron, again a string is created, which breaks and yields just additional hadrons.

5.9 Hadronic resonances

A consequence of the potential (5.10) is that for each hadronic state, there can be further excited states. These are the hadronic resonances, and exist for both baryons and mesons.

[12] Again, the true QFT version of this is much more subtle, and not yet fully understood.

E. g. the first excited state of the nucleon is the N(1440), with a mass of 1440 MeV, as the name indicates. In total, about 28, experimentally more or less well established, resonances have been observed for the nucleons, and likewise for the other hadrons.

An interesting observation, which can be understood on basis of the string picture discussed in the previous section, is the presence of so-called Regge trajectories. A Regge trajectory is that the mass squared of the resonances are roughly proportional to its total angular momentum, i. e. the combination of spin and orbital angular momentum. This has been experimentally observed before the inception of QCD, and has led to the first formulation of a string theory as a theory of the hadronic interactions. Due to consistency problems, and the lack of ability to describe other experimental observations, this approach of a quantized string has been dropped later in favor of QCD.

All hadronic resonances are unstable, and decay in various ways. Most quickly are those where the decay proceeds by splitting into hadrons, which usually involves the decay into a lower resonance and some pions. E. g. the N(1440) decays in about 50-75% into a proton or neutron and a pion, with a decay time of roughly $(200\text{ MeV})^{-1}$. Some resonances can kinematically not decay into two hadrons, and they de-excite by emission of photons or other particles. This happens on a much longer time scale. Such states are often considered as excited states in contrast to the resonances. These are under the decay threshold into purely hadronic final states. The much longer decay time is of course due to the fact that the strong force is, indeed, the strongest one, and all processes subjected to it occur very quickly. The only exceptions are those processes close to, but above, the hadronic decay threshold. There kinematic reasons, the absence of a large phase space, statistically suppress decays[13].

Th study of the decays of hadrons has been very useful in the understanding of the strong force. As will be seen later in section 5.12, today decays of hadronic resonances have also become an important window into the study of other processes.

5.10 Glueballs, hybrids, tetraquarks, and pentaquarks

When studying the quark model and the possible color charges, it becomes quickly clear that three quarks or one quark and one anti-quark are not the only possibilities how to create color-neutral objects. Tetraquarks, made from two quarks and two anti-quarks, as well as pentaquarks, made from four quarks and one anti-quark, are equally possible. There is indeed no a-priori reason why such bound states should not exist. However, the

[13] Quantum interference effects can occasionally modify those statements, but they are rather good guidelines for estimates of decays.

are not yet any unequivocally observed tetraquarks or pentaquarks. The reason for this are due to two effects.

The first effect is mixing. E. g. for a tetra-quark, it is almost always possible to construct a meson with the same quantum numbers, i. e. the same spin, parity, charge-parity and electric charge. There is also the possibility to construct equally well a dimeson molecule. One of the most infamous examples is the σ meson[14]. It is a light neutral meson, with quantum numbers compatible with the states $\bar{u}u$, $\bar{d}d$, $\bar{u}u\bar{d}d$, $\bar{u}u\bar{u}u$, $\bar{d}d\bar{d}d$, $\pi^+\pi^-$, and $\pi^0\pi^0$. Since it is a quantum state, it follows the quantum rules, which in particular imply that all states with the same quantum numbers mix. It is therefore a superposition of all these seven states. The question which of these states contribute most is highly non-trivial. It can, in principle, be experimentally measured or theoretical calculated. There is no really reliable way of estimating it. The results found so far indicate that the combination of two pions is most dominant, it is therefore likely a molecule. For most other states the two-quark components appears to be the dominant one. Similarly, almost all baryons turn out to be completely dominated by the three-quark state. There are only few cases left, where there is still some doubt.

The second possibility is to investigate one of those possibilities where the quantum numbers of the tetraquarks cannot be created using a two-quark system. Such cases are rare, but they exist. In principle, it would therefore be sufficient to just observe such a state. Unfortunately, almost all of these states are highly unstable. They are therefore experimentally hard to observe, and only very few candidates have been found so far, though recently a few candidates have been identified, to be discussed later in section 5.12.

This problem becomes more complicated due to the gluons. Though it is not possible to create a color-neutral state from a quark and a gluon, it is possible to combine a quark, an anti-quark and one or more gluons to obtain a colorless state. It is similarly possible to combine three quarks and a number of gluons to obtain a colorless state. Such states are called hybrids. However, the gluons can at most add angular momentum, but no other charges to the state. Therefore, there is always a state with the same quantum numbers, but just made from quarks. Since adding a particle or orbital angular momentum to a state usually increases its mass, these states are unstable against the decay to a state with just the minimum number of quarks. Though these hybrids are therefore formally admixtures to any state, it is essentially always a small one, and therefore hybrids are very hard to identify both experimentally and theoretically.

The last class of states which can come into the mix are bosonic glueballs, which

[14] The official name is $f_0(500)$, though the historical name of σ meson still is commonly used.

combine only gluons to a colorless objects. The usual counting rules of the quark model do not apply to them, but as a rough estimate even the simplest state is made out of four gluons. Such states carry no electric charge, and most of them have the same J^{PC} quantum numbers as mesons, and therefore mix. However, there are some candidates, particularly the so-called f(x) mesons, with x around 1500 MeV, which appear to have a large admixture from glueballs. This is experimentally identified by the possible decays. Since gluons are, in contrast to quarks, electrically neutral, decays into electrically neutral decay products, except for photons, should be favored if there is a large glueball contribution in the state. This has been observed, especially when comparing the partial decay widths of decays to uncharged particles to the one to photons. However, even at best these state ar only partially glueballs.

There are some glueball states which have quantum numbers which cannot be formed by only quarks, at least not in the simple quark model. Unfortunately, all theoretical estimates place theses states at masses of 2.5-3 GeV, and therefore far above any hadronic decay threshold. They are therefore highly unstable, and decay quickly. Hence, there is not yet any experimental evidence for them, though new dedicated searches are ongoing or are prepared.

5.11 Flavor symmetry and strangeness

So far two quark flavors, up and down, have been introduced. As discussed in section 4.11, this implies the presence of an approximate SU(2) flavor symmetry, since the masses of both quarks are small. It was possible to describe all hadrons introduced so far using just these two quarks with the quark model.

However, already before QCD was formulated, hadrons were observed, which do not fit into this picture. The most well-known of them are the kaons $K^{\pm}$, K^0, and $\bar{K}^0$, four mesonic hadrons of masses 494 MeV for the charged ones and 498 MeV for the two uncharged ones. Most remarkably, these new mesons were more stable than those of similar masses made from the two quarks inside the quark model.

The resolution of this mystery is that there are more than the two quark flavors necessary to construct the proton and neutron. These additional quark flavors, which will be discussed now and in the following two sections, do not occur in naturally observed atomic nuclei[15] but can be produced in natural or artificial collisions.

The quark to obtain the kaons in the quark model has been called the strange quark s. Its current mass is 90-100 MeV, and therefore its full mass about 400-450 MeV. It follows

[15]With the possible exception of the inner core of neutron stars, though this is not yet settled.

then that the kaons, with only 500 MeV mass, are also would-be Goldstone mesons, obtaining their mass from the with the strange quark enlarged chiral symmetry breaking. Since the strange quark's current mass is already not too small compared to usual hadronic scales, the kaons are much closer to their expected mass of about 700 MeV in the quark model obtained from a construction of a u or a d quark and one of the new s quarks.

Indeed, the s quark has an electric charge of $-1/3$, just like the d quark. The charged kaons are therefore the combinations $u\bar{s}$, $\bar{u}s$, and the neutral ones $d\bar{s}$ and $\bar{s}d$, explaining their small mass difference, and their multiplicity. The Goldstone theorem actually predicts that there should be 8 Goldstone bosons. These are six. The other two are the η and η' mesons, which are made from $\bar{s}s$ combinations, and some admixtures from neutral combinations of u and d quarks. They are therefore even heavier, the η having a mass of 550 MeV. Somewhat peculiar, the mass of the η' is much higher, about 960 MeV. The reason is that the η' also receives mass from another source, the so-called axial anomaly. The latter will be discussed below in section 6.11.

Besides the broken chiral symmetry, there is also the now enlarged flavor symmetry. Since it involves up, down, and strange quarks, it is an SU(3) group. Since the quarks have different masses, the group is not unbroken, but reduced to three counting symmetries, i. e. $U(1)^3$. Hence, the individual quark flavors are conserved, but bound states with differing quark content have differing masses. This conservation of quark flavor by the strong interaction is also at the origin of the name strangeness. When the kaons were discovered, the quark model was yet to be established. The kaons, and also baryons, called hyperons, with a single or more strange quarks included, showed a different decay patterns than ordinary hadrons, due to the conservation of strangeness. Thus, they did not fit into the scheme, and were therefore considered strange.

The presence of the strange quark, which has an effective mass of about 400 MeV, and thus close to the masses of the up and down quarks of 300 MeV, adds many more possible combinations to the quark model, which all have very similar masses. Thus, there is a very large number of hadrons with masses between 500 MeV, and roughly 3000 MeV, where the states become too unstable to be still considered as real particles. In fact, the number of states N turns out to rise exponentially with mass, $N \sim \exp(M/T_H)$. This is a so-called Hagedorn spectrum, where T_H is called the Hagedorn temperature. The reason for the name is that naively a system with such a spectrum has the property that at a finite temperature, the Hagedorn temperature T_H, an infinite number of particles is created, and thus the system cannot be heated beyond this point. For the strong interactions, the Hagedorn temperature is about 160 MeV. Of course, it is in practice possible to go beyond this temperature. What happens is that at this point the quark substructure can no longer be ignored, and this effect limits the number of states growth. This has originally lead to the idea that at this temperature a phase transition has to occur, which signals the change from a hadronic system to one were the quark substructure becomes more important. However, the most reliable results to date rather indicate that it is a cross-over. This transition plays nonetheless a role in the development of the early universe, though a rather small one. This will be discussed in detail in section 5.18

The experimental and theoretical study of this transition is an important part of modern particle physics. Nowadays it is not restricted to the question of large temperatures, but also small temperature and large densities. This is the situation which is encountered during supernovas and in the interior of neutron stars and during neutron star mergers. At theses densities, which become larger than the density of nuclei, nucleons can no longer exist without starting to overlap with each other. Similarly to the high-temperature case, such an overlap of nucleons is expected to change the properties of matter dramatically, as the quark substructure can no longer be ignored. However, much less is known about this situation, as it is both experimentally and theoretically much more demanding to investigate. Some hope is provided by the possibility to use gravitational waves to probe the interior of neutron stars during neutron star mergers, but this will crucially depend on the developments in the next few years, whether this turns out to be correct. One of the most interesting questions is, whether strange quarks play a role for neutron stars.

Another aspect were strange quarks play an important role are the quests for tetraquarks and pentaquarks. Since strangeness is a conserved quantum number of the electromagnetic and strong interaction, it is possible to construct states which do not have the quantum numbers of an ordinary state, e. g. a meson with total strangeness of 2, which therefore must contain two strange quarks, and cannot be a simple quark-anti-quark state, or a baryon with strangeness -1, due to a single anti-strange quark, which therefore must be a tetra-quark. Searches for such signatures are intensively pursued.

5.12 Charm, bottom and quarkonia

It appears at first rather surprising that there should be just one other quark, which has the same electric charge as the down quark. This appears unbalanced, and another quark with the electric charge of the up quark appears to be necessary. Indeed, this is correct, and there is also a heavier copy of the up quark, which is called charm[16] quark. However,

[16]The name originates from the fact that it solves several experimental mysteries observed in the weak interactions, to be discussed in chapter 6, and because at that time it appeared to complete the quark picture.

while the strange quark has a rather similar mass, despite its larger current mass, as the light quarks, the charm quark has not. Its current mass is 1275 MeV, and thus similar to its effective mass of roughly 1600 MeV. As a consequence, hadrons involving a charm quark are much heavier than hadrons containing only the lighter quarks. Furthermore, though chiral symmetry is even more enlarged, its so badly explicitly broken that would-be Goldstone bosons involving charm quarks have essentially the same mass as without the breaking of chiral symmetry. The same is true for the flavor symmetry, and the only remnant is that charm is again a conserved quantum number in both the electromagnetic and strong interaction.

This conservation of charm has very interesting consequences. Of course there are hadrons were only one of the quarks is a charm quark, which are called open charm hadrons. The best known ones are the D mesons, with masses of about 1870 MeV mass and having the structure of a single charm quark and either an up or down quark. These are the lightest particles with a charm quark.

But there are also particles, especially mesons, which consists only of charm quarks. In the meson case, where the total charm is zero if they consist out of a charm and an anti-charm quark, these are called hidden charm. The latter states are particularly interesting, because they show a very interesting mass spectrum. In fact, the lightest $\bar{c}c$ states have a mass which is just below threshold for the decay into two hadrons with a charm quark and an anti-charm quark each, the $\bar{D}D$ threshold. They can therefore not decay directly. Of course, the charm and anti-charm quarks can annihilate. But because of how quark and gluon color charges are arranged, such a process is substantial suppressed in QCD compared to the decay with a production of an additional quark-anti-quark pair. Hence, decays occur very slowly. Hence, these hadrons are extremely stable compared to hadrons made from lighter hadrons, where the pions offer a simpler decay channel. Thus, these meta-stable state carry the name of charmonia, which have masses of about 3 GeV, but decay widths of around a few 100 keV.

Because of this fact, the charmonia states turn out to present a very good realization of the possible states permitted by the potential (5.10). Similarly to the hydrogen atom, this potential creates states distinguished by a main quantum number and orbital quantum numbers. The most well-known state is the J/Ψ, at about 3097 MeV with a decay width of 93 keV, which is a state with one unit of angular momentum. However, the ground state of the system is the η_c, with a mass of 2984 MeV and a decay width of 320 keV. That the ground state decays quicker is mainly due to kinematic effects from the angular momentum. Simply put, the ground state is in an s-wave, and thus the wave-functions of the two charm quarks have a large overlap. Thus an annihilation into photons is much more likely than in the case with angular momentum, where the overlap of the wave functions is much smaller. Right now about 8 states are known, which are below the $\bar{D}D$ threshold, the heaviest the so-called $\psi(2S)$, with a mass of 3690 MeV and a decay width of 303 keV.

These charmonia have been very instrumental in understanding the potential (5.10), and thus the strong interactions. The very well-defined spectrum, which provides the opportunity of a true spectroscopy, including many angular momentum states, permits a much cleaner study than in case of the light hadrons, where the ubiquitous decays into pseudo-Goldstone bosons make resonances decay very quickly.

However, not all of the states in this spectrum are easily explained within the framework of the quark model and the potential (5.10). These are the so-called X, Y, and Z mesons, with masses above the $\bar{D}D$ threshold, and some also with open charm. These states do not fit into the spectroscopic scheme, and especially some may have quantum numbers, which are not in agreement with a simple quark-anti-quark system. This is still under experimental and theoretical investigation. However, it already shows that the simple quark model is not also able to explain all hadrons.

With the charm quark, it may appear that everything is complete and symmetric. However, nature did not decide to stop at the charm quark, but added another quark: The bottom (or in older text beauty) quark. It is a heavier copy of the down quark, and has therefore an electric charge of -1/3. Its mass is about three times that of the charm quark, with a current mass of 4.2 GeV. It therefore introduces another quark flavor, but like for the charm quark, this does not play any role for dynamical chiral symmetric breaking, as the explicit breaking is far too large.

Other than the mass and the electric charge, the bottom quark behaves essentially as the charm quarks. Especially, there is a rich spectroscopy with open and hidden beauty[17], the latter also called bottonium in analogy to charmonium. Similar to the case for the charm quark, the lightest meson with open beauty is rather heavy, $B^{\pm}$ and B^0 being 5.3 GeV. As a consequence, the bottonium spectrum has a large number of quasi-stable states, the lightest being the η_b with a mass of 9.4 GeV and a decay width of roughly 10 MeV, the Υ playing the role of the J/ψ with a mass of 9.5 GeV and a width of 54 keV, and then even 15 states to the heaviest $\chi_b(3P)$ with 10.5 GeV observed so far. There are also heavier states, including bottom versions of the X, Y, and Z mesons, which do not fit easily into a simple quark model explanation. Thus, an even richer spectroscopy is possible, though the production of bottonia in so- called beauty farms, requires more resources than for the charmonia.

[17]For the flavor quantum number the term beauty still survives.

Of course, for both bottom quarks and charm quarks there exist also baryons, with one or more of these quarks, also with both charm and bottom quarks. These are rather complicated to produce, but have been observed, though baryons with multiple charm or bottom quarks only very recently. These baryons are not as stable as the mesons, but are still sufficiently stable that they production and decay take place so far apart from each other, a few μm, that both processes can be experimentally resolved. They play therefore an important role to identify the production of charm and especially bottom quarks in high-energy collisions (so called c- and b-tagging).

Together, charmonia and bottonia are usually referred to as quarkonia. Studying theses states are also interesting for other reasons than to understand QCD. Because the J/ψ and Υ are very long-lived, they are very well suited for precision measurements. Furthermore, as will become evident in section 6.5, it appears plausible that new phenomena will be influencing heavy quarks stronger than light quarks. Searching therefore for deviations of the decays of quarkonia, especially bottomonia, from the standard-model predictions has become an important branch in experimental physics, which is also called flavor physics.

5.13 Top

With the introduction of the bottom quark the situation appears again as unbalanced as with the introduction of the strange quark. This is indeed true, and the picture extended by the top (or in old texts truth) quark. This is the last quark, which so far has been found, though there is no a-priori reason to expect that there may not be further quarks, and searching for them is a very active field in experimental physics.

The top quark is a heavy version of the up quark, and thus has an electric charge of 2/3. The fact which is really uprising about it, is its mass of 173 GeV. It is therefore forty times heavier than the bottom quark, and this is the largest jump in masses in the quark mass hierarchy. It is an interesting side remark, and a triumph of theory, that this mass has been established within 10 GeV before the direct observation of the top quark, by measuring other processes sensitive to the top quark very precisely, and then using theory to make a prediction.

The enormous mass of the top quark makes it the heaviest particle detected so far. Due to its large mass, it decays much quicker than the lighter quarks, with a decay width of 2 GeV. This is a consequence of effects to be introduced in section 6. Hence, the formation even of short-lived hadrons with a top quark is almost impossible, and no hadron with a top quark has so far been directly observed. Whether a quasi-stable toponium is possible is not clear, but so far there is no experimental evidence for it. But, due to the mass, it is also not trivial to produce large numbers of top quarks, and thus the study of top quarks is rather complicated. Hence, the top quark is so far more of interest for its own properties, particularly its mass, rather than for its relation to QCD.

5.14 Partons and scattering experiments

Due to confinement, quarks (and gluons) cannot be isolated as individual particles. This even applies to the top quark though this is of little practical relevance as it decays so quickly. This has quite important consequences for the experimental detection of quarks and gluons, as it implies that only hadrons are ever observed.

This also implies that a hadron is, in fact, a complicated structure of the quarks providing its quantum numbers, the so-called valence quarks, and many virtual particles. The latter can be either also quarks, then called sea quarks to distinguish them from the valence quark, and (also virtual) gluons. Collectively, the valence quarks and the virtual particles inside a hadron are called partons, a term originating from a time as it was not yet clear whether there are different particles inside a hadron, or just a single type. Furthermore, since quarks are indistinguishable, there are also processes where a sea quark and a valence quark are exchanged. Thus, valence quarks are not immutable.

The collective of all partons carries the total quantum numbers, like momentum, electric charge, spin etc., of a hadron. Especially, each parton carries a momentum fraction x (called Feynman x) of the total hadron momentum. If there would be only the valence quarks, each of them would have $x = 1/3$. Because of the virtual particles, the average x for each particle is smaller, since there are more particles, though quantum fluctuations permit also values of x larger than 1/3, up to 1, though rarely. Since gluons are massless, they contribute the majority of particles at small x. As a consequence, many properties of hadrons can already reasonable well be approximated when the sea quarks are neglected, the so-called quenched approximation.

5.14.1 Fragmentation functions and jets

One of the most direct consequences of confinement is that the quarks and gluons are not observed. In fact, when they are produced in a high-energy collision they behave, due to asymptotic freedom, for a short time as quasi-free particles. But confinement kicks in after time or distance scales of about or less than a fermi. Since the initial state has been color neutral, the final state is also color neutral. Thus, any produced partons are connected, in the string picture by strings, with all the other colored particles. While the distances are small, the string is short and little energy is required. When, after the collision, the particles start to move apart, energy must be invested into the string, and it breaks eventually. This produces a spray of hadrons. It is said that the partons hadronize.

In the rest frame of the produced quark or gluon, this hadronization will be isotropic, if sufficiently many quarks and gluons are present that the color charge is roughly isotropically distributed. That is a rather good approximation, due to the light gluons, for most cases. But in high-energy collisions the quarks and gluons are usually produced at high energies. Hence, in the laboratory frame the produced hadrons will be highly collimated along the direction of the original quark and gluon. Such a collimated spray of hadrons originating from a single quark or gluon is called a jet.

This jet, in principle, carries all the information of the original quark and gluon, like electric charge, momentum, invariant mass, and flavor. In practice, some particles may still be not emitted very close to the original direction, i. e. not in the jet cone, and there may be other jets in an event. Thus, such information is not always easy to reconstruct. Still, by identifying such a jet, especially if it carries very much energy, it is possible to infer information about the original quark or gluon. E. g. top quarks are identified by the total invariant mass of the jet.

A further information can be obtained from which hadrons are present in the jet. Since quarks and gluons have different flavors, masses, spins, and electric charge, they produce different hadrons. The mathematical description of the probability for a hadron of type i to emerge from a quark or gluon of type j is called a fragmentation function. Such fragmentation functions are very hard to calculate, but can be measured experimentally, to some extent.

5.14.2 Parton distribution functions

Even more important in actual experiments is the inverse question: How probable is it to find inside the projectile hadrons (e. g. the protons at the LHC) a certain quark flavor or a gluon with a given momentum fraction x of the total hadron momentum? The answer to this question are given by the parton distribution functions $P_{hi}(x)$ for a hadron of type h and a parton of type i with momentum fraction x. The practical importance of these functions becomes immediately clear when considering some reactions which lead to a given final state f. The actual cross-section is then a convolution[18]

$$\sigma_{hh\to f} = \sum_{ij} \int dx dy \sigma_{ij\to f} P_{hi}(x) P_{hj}(y) \theta(1-x-y). \tag{5.11}$$

[18]This is somewhat symbolic, and the indices i, j, and f include many further properties, like spins, polarizations etc.

Hence, the actual cross-section gets contribution from all possible parton combinations inside the projectile hadrons.

The calculation of the parton distribution functions, or PDFs for short, is highly complicated. It is, especially for small values of x a non-perturbative question, and has not yet been performed satisfactorily. Hence, they are usually parametrized using experimental input, where theoretical considerations determine the parametrization. The experimental results are most easily obtained when instead of two hadrons electrons and hadrons are collided, as has been done at DESY in Hamburg with electrons on protons.

The importance of the PDFs cannot be underestimated, since they vary substantially with x. Especially, except for the valence quark flavors they drop very quickly with x. As a consequence, it is rather unlikely to find a parton with large energy fraction inside a hadron. Since the cross-section for particle production are usually peaked around the mass of the particle to be produced, see equation (2.4), the effectiveness of producing new particles drops quicker with the energy than would be expected from just the energy dependency of the elementary cross section. Hence, the parton luminosities, i. e. the luminosity folded with the PDFs, is the actual figure of merit for a hadron collider. On the other hand, this means that increasing the energy for a hadron collider can increase the effectiveness faster than the naive energy scaling of the elementary cross-section.

PDFs have also a theoretical importance. Because of asymptotic freedom, it is often possible to obtain perturbatively a rather good approximation to the elementary cross-section $\sigma_{ij\to f}$. The involved PDFs are then known experimental input, and the interesting process can be obtained directly. That is the standard procedure used for most calculations in high-energy experiments. The assumption is that (5.11) is valid, i. e. that the process can be factorized into the PDFs and the elementary interaction. Because of asymptotic freedom, this is a reasonable assumption so far as the fractions x and y yield energy fractions for the corresponding partons much larger than 1 GeV. Since an integration is performed over all x and y, however, this will only work if the product of the PDFs and the cross-section becomes small at small x and y. For many cases of practical relevance, this is still a good approximation. Otherwise, so-called factorization violations are found, and this approach breaks down.

Formally, due to the presence of virtual particles, any elementary particle has to be described in terms of a PDF. E. g., an electron is surrounded by virtual photons. There is hence a PDF describing the probability to observe a photon with an energy fraction x of the total electron momentum inside an electron, $P_{e^-\gamma}(x)$. Of course, also quarks or gluons can be found in this virtual cloud. However, since the electron is not strongly interacting, the PDFs for non-partons can be often calculated to a reasonable accuracy perturbatively.

Still, since especially at low energies also hadrons can be encountered inside an electron or a photon, this is not completely possible. This is not an esoteric effect. E. g. in the interaction of photons with nuclei the PDF $P_{\gamma\rho}(x)$ describing the mutability of protons into ρ mesons plays an important, and experimentally verified, role.

5.14.3 Monte-Carlo generators

As is visible, the experimental description of collisions is usually performed in a three step process, as far as factorization is viable. The first step is to obtain the initial particles for the interesting process. These are described by the various PDFs of the initial projectiles. The second step is to calculate the elementary process, usually performed perturbatively, to a certain order in perturbation theory. For the LHC at this time most such hard processes are calculated to orders between tree-level and NNLO. The third step is then to transform the final state of the elementary reaction into the particles finally observed in experiments. This involves usually first a decay cascade of the particles obtained in the elementary process into electrons, positrons, muons, photons, and other particles stable and weakly interacting on the scale of the experiments or into quarks and gluons. In the latter case also the creation of jets and/or fragmentation must be calculated, to obtain the final state.

While the initial state is rather simple, already the possibly involved intermediate particles before the hard collisions can easily number dozens. Moreover, even at moderate orders of perturbation theory, very many contributions appear. The final state is even more complex, and can involve many final particles, and several jets - up to eight jets and about a dozen quasi-stable particles are currently feasible for experiments.

To deal with this complexity requires computers. For many theories, the outer part of the event, the creation and fragmentation of the particles entering into the elementary hard process, is the same. Hence, this part can be automatized. Such programs, in which only the elementary cross-section is used as an input, are called Monte-Carlo generators. Current examples are publicly available codes like Sherpa, Pythia, Herwig, or Whizard, and can be downloaded e. g. at www.hepforge.org.

Even the elementary cross-sections can be calculated in many cases automatically, especially at tree-level. In these cases only the Lagrangian has to be provided, to permit an automated calculation from the initial to the final state.

5.15 Some algebra and group theory

Up to now, QCD has not been specified. This will be done in the next subsection 5.16. The first question to be answered, however, is how to to implement the color structure discussed in section 5.5 into a theory. The fact that three colors should create a neutral state, but not any other combination without anti-colors, already tells that it is impossible to do so just with real numbers. Vectors can indeed be arranged to have such a property. Since in two dimensions at most two vectors can be linearly independent, the space must be three-dimensional, to have three independent quark colors. But also the eight gluons colors, and the anti-colors must be fitted in.

It turns out that there is a unique mathematical structure which offers all these properties. This is the so-called Lie algebra su(3), and its associated Lie group SU(3). These are special cases of so-called Lie algebras and Lie groups. These are in turn special cases of symmetry groups and algebras, as introduced in section 4.1. In particle physics, some theory of these algebras and groups is inevitable, as it is not only necessary for the strong interactions, and QCD, but also for the weak interactions. In this brief section, the most basic concepts for Lie groups and algebras will be given. Much more group theory than the following is usually not necessary, though quite useful.

In fact, Lie groups are ubiquitous in particle physics. Examples have already been encountered in the context of flavor in section 4.11 and gauge symmetry 3.4, but there the details were not yet necessary.

To understand the concept, return to QED. QED implemented gauge-fields and matter which were related by the gauge-transformation to the u(1) algebra and the U(1) group. Thus, the fields could be regarded as tensor products of a function of space-time times an element of the group. To deal with QCD, it will turn out that it is only necessary to replace this u(1) by a different group, the aforementioned su(3), which is able to faithfully represent the concept of color. To proceed requires more insight into this group structure than for the flavor, and therefore some basic elements of the mathematical foundations of non-Abelian Lie algebras, which will be introduced before formulating QCD.

Note that the fact that such mathematical structures describe particle physics is an assumption, which must be experimentally supported. This was indeed possible for the standard model of particle physics.

The basic element in this representation will be to represent a quantity carrying a gluon-like color charge by the base vectors of a particular algebra. Such base vectors are called generators of the algebra. The type of algebra needed in particle physics are so-called Lie algebras, G. Hence, if there should be N independent gluon-like charges, there must be N independent base vectors τ^a with $a = 1...N$ and $N = \dim G$, and the Lie algebra must therefore be N-dimensional.

The defining property of such an algebra are the commutation relations

$$\left[\tau^a, \tau^b\right] = 2if^{ab}_{\ \ c}\tau^c. \tag{5.12}$$

with the anti-symmetric structure constants f^{abc}. Noe that sometimes the factor of 2 is included in the f^{abc}. This algebra implies already that everything is non-commutative, hence the name of non-Abelian Lie algebra. These structure constants fulfill the Jacobi identity

$$f^{abe}f^{cd}_{\ \ e} + f^{ace}f^{db}_{\ \ e} + f^{ade}f^{bc}_{\ \ e} = 0. \tag{5.13}$$

These base vectors have to be further hermitiean, i. e., $\tau_a = \tau_a^\dagger$. Note that the position at top or bottom (covariant and contravariant) of the indices is of no relevance for the Lie algebras to be encountered in the standard model, but can become important in more general settings: For Lie algebras, the metric is just δ^{ab}.

Such a Lie algebra can be represented, e. g., by a set of finite-dimensional matrices. An example is the su(2) algebra with its three generators, which can be chosen to be the Pauli matrices,

$$\begin{aligned}\tau^1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\\ \tau^2 &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\\ \tau^3 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.\end{aligned}$$

Furthermore, to each algebra one or more groups can be associated by exponentiation, i. e.,

$$\lambda^a = e^{i\tau^a}, \tag{5.14}$$

provides base vectors for the associated group, which are by definition unitary and thus $\lambda_a^{-1} = \lambda_a^\dagger$. For su(2), these are again the Pauli matrices, generating the group SU(2). However, the relation 5.14 is not necessarily a unique relation, and there can be more than one group representation. E. g., for su(2) there are two possible groups related to the algebra by the relation (5.14), the group SU(2) and the group SU(2)/Z_2 $\approx$SO(3), where matrices which differ only be a negative unit matrix are identified with each other.

Because of the exponential relation, a generic group element $\exp(i\alpha_a\tau^a)$ with real numbers α_a can be expanded for infinitesimal α_a as $1 + i\alpha_a\tau^a$. Thus the algebra describes infinitesimal transformations in the group. This will play an important role when introducing gauge transformations for non-Abelian gauge theories.

There is only a denumerable infinite number of groups which an be constructed in this way. One are the N-dimensional special unitary groups with algebra su(N), and the simplest group representation SU(N) of unitary, unimodular matrices. The second set are the symplectic algebras sp($2N$) which are transformations leaving a metric of alternating signature invariant, and thus are even-dimensional. Finally, there are the special orthogonal algebras so(N), known from conventional rotations. Besides these, there are five exceptional algebras g_2, f_4, and e_6, e_7, and e_8. The u(1) algebra of Maxwell theory fits also into this scheme, the u(1) group is the special case of all f^{abc} being zero, and the algebra being one-dimensional. This is equivalent to so(2).

The two algebras relevant for the standard model are su(2) and su(3). The su(2) algebra has the total-antisymmetric Levi-Civita tensor as structure constant, $f^{abc} = \epsilon^{abc}$ with $\epsilon^{abc} = 1$. The algebra su(3) has as non-vanishing structure constants

$$\begin{aligned} f^{123} &= 1 \\ f^{458} = f^{678} &= \frac{\sqrt{3}}{2} \\ f^{147} = -f^{156} = f^{246} = f^{257} = f^{345} = -f^{367} &= \frac{1}{2}, \end{aligned}$$

and the corresponding ones with permuted indices. There is some arbitrary normalization possible, and the values here are therefore conventional.

From these, also the generators for the eight-dimensional algebra su(3) can be constructed, the so-called Gell-Mann matrices,

$$\begin{aligned} \tau^1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \tau^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \tau^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \tau^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \tau^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \qquad \tau^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ \tau^7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \qquad \tau^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \end{aligned} \tag{5.15}$$

In general, there are $N^2 - 1$ base vectors for su(N), but the dependency for the other algebras is different. For the sake of simplicity, in the following only the expressions for su(N) will be given.

Generators, which are diagonal as matrices, and therefore commute with each other, are said to be in the Cartan sub-algebra or sub-group of the algebra or group, respectively.

For su(2), this is only one generator, for su(3) there are two. This also gives the rank of the group.

These lowest-dimensional realization of the commutation relations is called the fundamental representations of the algebra or group. Since the commutation relations are invariant under unitary transformations, it is possible to select a particular convenient realization. Note, however, that there may be more than one unitarily inequivalent fundamental representation. E. g., for su(3) there are two, called fundamental and anti-fundamental, while there is only one for su(2).

It also possible to give representations of the algebras with higher-dimensional matrices. The next simple one is the so-called adjoint representations with the matrices

$$(A^a)_{ij} = -if^a_{ij},$$

which are three-dimensional for su(2) and eight-dimensional for su(3). There are cases in which the fundamental and the adjoint representation coincide. This can be continued to an infinite number of further representations, which will not be needed here, though they may appear in the context of beyond-the-standard-model physics.

This completes the required group theory necessary to describe the standard model.

5.16 Yang-Mills theory and QCD

It is now possible to return to QCD. It turns out that three quark colors (and three anti-quark colors) as well as eight gluon colors are precisely related to group SU(3), i. e. the group of unitary three-by-three matrices with unit determinant. The three quarks can then be assigned to live in the corresponding three-dimensional complex vector space, the anti-colors in the corresponding hermitiean conjugated space, and the eight gluon colors can be associated with the eight of the nine base matrices, the ninth being the unit matrix. The eight non-trivial base matrices are the Gell-Mann matrices λ_a, given in (5.15).

More mathematically, following and generalizing section 3.4 for QED, the gluon fields live in the corresponding algebra, and are connected to the generators, and thus belong to the adjoint representation. The quarks and anti-quarks belong to the fundamental and anti-fundamental representation, respectively[19].

Is it thus sufficient to just generalize the QED Lagrangian (3.6)? The answer to this is both yes and no. The quarks are now vectors in color space, i. e. the spinors carry an

[19]The actual group in the standard model is furthermore $SU(3)/Z_3$, that is all elements divided out which commute with every other elements, i. e. the center of the group, except the unit matrix. This is of no relevance here.

additional color index like ψ^i. A gauge transformation acts now as a rotation in this color space, and is thus a matrix G as

$$\psi_i' = G_{ij}\psi_j = \left(\exp\left(i\frac{g}{2}\alpha_a\tau_a\right)\right)_{ij}\psi_j,$$

where the α_i are space-time-dependent functions, and g is playing the same role as the electric charge, and is thus a coupling constant. The corresponding covariant derivative to form a gauge-invariant Lagrangian has to take this matrix structure into account, yielding

$$D_\mu^{ij} = \delta^{ij}\partial_\mu - i\frac{g}{2}A_\mu^a(\tau^a)^{ij}, \tag{5.16}$$

where the A_μ^a are now the eight gluon fields. These have to transform under an infinitesimal gauge transformation like

$$A_\mu^{a'} = A_\mu^a + \frac{1}{g}D_\mu^{ab}\alpha^b, \tag{5.17}$$

$$D_\mu^{ab} = \delta^{ab}\partial_\mu - gf^{abc}A_\mu^c. \tag{5.18}$$

A finite transformation takes the form

$$i\tau^a A_\mu^{a'} = G\tau^a A_\mu^a G^{-1} + (\partial_\mu G)G^{-1},$$

were the combination $A_\mu^a\tau^a$ is often abbreviated as just A_μ. In the infinitesimal version (5.17) a second covariant derivative appears (5.18), where additional constants f^{abc} appear. These are just the structure constants of SU(3) from the defining algebra (5.12), and are therefore related to the Gell-Mann matrices as

$$\left[\tau^a, \tau^b\right] = 2if^{abc}\tau^c,$$

forming the algebra su(3).

Replacing the spinors by color vectors of spinors in the QED Lagrangian (3.7), and the flavors by the six quark flavors, is already sufficient to obtained the quark part of the QCD Lagrangian. However, it turns out that just taking the same field-strength tensor as in QED (3.5) with replacing the fields by their generalization with an index is not sufficient. In fact, in this case the Maxwell term (3.4) is not gauge-invariant.

This can be solved by introducing a more complicated field strength tensor

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc}A_\mu^b A_\nu^c. \tag{5.19}$$

It is a straightforward, though tedious, calculation to show that with this replacement the Maxwell-term (3.4) becomes gauge-invariant. However, in contrast to QED, this field-strength tensor in itself is still not gauge-invariant, but $\mathrm{tr}F_{\mu\nu}F^{\mu\nu}$ is.

QCD is therefore described by the Lagrangian

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F_a^{\mu\nu} + \sum_f \bar{\psi}^i_f \gamma_\mu (iD^\mu_{ij} - m_f)\psi^j_f. \tag{5.20}$$

This theory has a number of further remarkable features, which an be read off this Lagrangian.

The first is that the interaction between quarks and gluons is given by a vertex very similar to the QED one, it is just adorned by a Gell-Mann matrix. Since the interaction potential between quarks (5.10) is very different than the Coulomb potential, this implies that there must be substantial radiative corrections. In fact, since there is a qualitative change, these must be of non-perturbative origin.

Second, in QED it is in principle possible that each flavor has its own electric charge. This is not the case here. Due to the appearance of the second term in the field-strength tensor (5.19) the coupling constant in the gauge transformation in the gauge field (5.17) is fixed. But the covariant derivative (5.16) will then only yield the necessary cancellations between the transformations of the gluons and the quarks if the coupling constant appearing in it is the same as for the gluon fields. Thus, every flavor has to couple to the gluons with the same coupling constant g. This feature, which arises from the underlying group structure, is called coupling universality, and has been experimentally verified.

Third, since the field strength tensor (5.19) is not gauge-invariant, so are not the chromoelectric and chromomagnetic fields, and then neither can be the chromo version of the Maxwell equation nor the color current. Hence, color and colored fields are not gauge-invariant, and therefore cannot be observable. Because of confinement, as described in section 5.8, this was not expected. But this makes confinement a necessity. Still, this leaves open the question why confinement operators in the way it does, with the potential (5.10), and not in any other way.

Fourth, the appearance of a quadratic term in the field-strength tensor (5.19) implies that there are cubic and quartic self-interactions of the gluons with each other. This coupling strength, due to the underlying group structure, are uniquely fixed, again a consequence of coupling universality. But this also implies that gluons, in contrast to photons, are not ignorant of each other. A theory with only gluons in it is therefore not a trivial theory, but an interacting one. In fact, it turns out to be a strongly interacting one, exhibiting features like confinement or the formation of glueballs. This reduced version of QCD is called Yang-Mills theory. It is sometimes also denoted as quenched QCD, in the sense that any conversion of gluons in quarks (and thus sea quarks) is fully suppressed, i. e. quenched. It is possible to calculate in this reduced theory also approximately quantities like hadron masses. They are then found to be roughly of the same size as in (full) QCD, indicating that most of the physics in QCD resides in the gluon sector.

Fifth, a careful investigation of the vertices shows that the gluon-self-interaction and the quark-self-interaction are opposite. While quarks tend to reduce the interaction, i. e. they screen color charges, gluons tend to increase them, they anti-screen it. The experimental consequences of this fact actually lead originally to the discovery of the group structure of QCD, as this behavior is almost exclusive to theories with a non-Abelian Lie group as gauge group. Because anti-screening dominates this leads to the difference between QED and QCD, especially the feature of asymptotic freedom discussed in section 5.8. In the terms of section 3.8, the coupling does not become stronger but weaker with energies.

Especially, it is possible to calculate the β function (3.8) also of QCD. To leading perturbative order only the first Taylor coefficients is relevant, which takes for QCD the value

$$\beta_0 = \frac{11}{3}N_c - \frac{3}{2}N_f = 2,$$

where $N_c = 3$ is the number of colors and $N_f = 6$ is the number of flavors. The value is positive, in contrast to QED. Thus, because of the formula for the running coupling to leading order (3.9), the interaction strength in QCD diminishes with increasing energy, a manifestation of asymptotic freedom. The scale Λ turns out to be roughly 1 GeV, and the size of the coupling at about 2 GeV is 0.3, to be compared with the one of QED being about 0.0075. QCD is thus indeed much stronger interacting at every-day energies than QED.

However, at energies of the order of the proton mass, the perturbative coupling diverges, and a Landau pole arises. In this case, this clearly signals the breakdown of perturbation theory. In fact, even a value of 0.3 is already shaky for perturbative calculations, and strong deviations from perturbation theory have been encountered up to 25 GeV. Thus, only if all involved energy scales are in the range of (a few) 10 GeV or more QCD can be treated perturbatively.

However, it is possible to treat QCD at low energies also non-perturbatively. There are various methods available. The most successful one is arguably numerical simulations, which confirms that the QCD Lagrangian (5.20) describes objects like the proton and other hadrons. QCD is thus, to the best of our knowledge, the correct theory of hadrons, and hence of nuclear physics.

To obtain the full Lagrangian of QCD and QED, it is just necessary to add the QED Lagrangian to the QCD Lagrangian (5.20), and modify the covariant derivative of the quarks to include the electromagnetic interaction as

$$D_\mu^{ij} = \delta^{ij}\partial_\mu - igA_\mu^a(\lambda^a)^{ij} - ie_f A^\mu,$$

where now the electric charge depends on the flavor, and is either $2/3$ or $-1/3$ in units of the electron charge. This ratio of the electron to quark charge is actually very well established experimentally, though why it is the case is unclear. However, as will be discussed in section 6.11 it is necessary for the formulation of the standard model in its current form.

5.17 Cosmic radiation

Before adding further elements to the standard model it is worthwhile to discuss two examples where QCD (and partly also QED) plays an important role, which are usually not directly associated with particles physics.

One of them is cosmic radiation. There are many sources in the universe which create high-energetic particles. These are leptons and photons, but also protons. Electrically neutral particles can be most easily tracked back to their origin, as their paths are not changed due to galactic or intergalactic magnetic fields, while protons are affected. These cosmic particles can have energies many orders of magnitude larger than created by any experiment. Energies of up to almost 10^9 TeV have been observed[20]. Possible sources may be active galactic nuclei.

When such a high-energetic particles hits earth's atmosphere, it reacts with the atoms and atomic nuclei. As a consequence, a jet, called in this context a cascade, of hadrons and electrons, muons, and photons is created, which an be observed with observatories on earth. To be able to decide whether the original particle was charged or uncharged requires to understand the evolution of this jet. This is a complicated problem, as it requires to understand how the various particles interact and are produced. But this is decisively important, so that it becomes possible to understand whether the direction of the cascade could have been influenced by magnetic fields, and thus whether its origin is meaningful or not.

Similar to the Monte Carlo generators of section 5.14.3 for particle collisions, this can be usually only done by simulations. This is an important task for astrophysical observation, and our knowledge of the universe.

[20] Such very high-energetic particles are only observed extremely rarely, at rates which can be as low as one per km^2 and year, and are thus not useful to test particle physics.

5.18 The QCD phase diagram

As has already been noted in section 5.11, aspects of QCD play an important role for the physics of neutron stars. In fact, a neutron star can be considered as a gigantic, stable nucleus, though this is an oversimplification. Since the density inside a neutron star increases from the outside to the inside, the physics will change. While the outer part is indeed essentially a nuclear system, the situation in the interior is not clear. Because QCD is so strongly interacting, it is very hard to treat in such an environment, and even numerical simulations fail[21]. It is therefore still an open field what actually occurs near the core of a neutron star, and whether the state of matter there is still just nuclei, or whether other hadrons, including strange ones, play a significant role.

The situation in neutron stars is only a part of a wider topic, the QCD phase diagram, i. e. what is the state of matter at densities of similar or larger size than in nuclei and temperatures of size of hadron masses.

Experimentally, these are very hard to address questions. For neutron stars, possibly gravitational wave astronomy, as remarked in section 5.11, together with the X-ray septum of neutron stars as a function of time, as well as their sizes and masses, will be the only available experimental input for a long time to come.

At much smaller densities, the situation becomes essentially the one of nuclear physics. here, experiments with nuclei, especially collisions, can be used. In this way, it was possible to find that there is a phase transition between a gas of nucleons and (meta-)stable nuclei at roughly the density of nuclei, which permits to form nuclei. This phase separation persist for a few MeV in temperature, up to about 15 MeV, where it ends in a critical end-point. Thus, both phases are not qualitatively distinguished, just like in the case of water.

From the point of view of particle physics, the involved energy scales of nuclei are very small, and the distinction between the liquid and gaseous phases is essentially irrelevant. Thus both phases are not regarded as different. This common phase is denoted as the hadronic or vacuum phase, the latter as for all practical purposes the phase consists of far separated hadrons separated by vacuum.

There are now several interesting directions to move on. Usually, they are signified by

[21]The reason is actually not the strongness of QCD. It is rather a problem which has to do with the distinguishability of particles and anti-particles, and how this can be treated if there are more particles than anti-particles. For some theories, like QCD, but there are also such systems in nuclear physics and solid-state physics, this entails algorithmic problems, the so-called sign-problem, which prevents so far the development of efficient simulation algorithms, as the computation time scales exponentially with the system size. It is a technical problem, and not a physics problem.

temperature and the baryo-chemical potential, i. e. the chemical potential distinguishing baryons and anti-baryons. Thus, zero (baryo-)chemical potential denotes a situation in which there is the same number of baryons and anti-baryons, which includes the possibility of none of either type.

When neutron stars are interesting, the axis with zero temperature and finite baryo-chemical potential is most interesting. Neutron stars do have some temperature, but it is of the order of one MeV, and even during their formation in a supernova explosion or during a merger it does not exceed 10-20 MeV. On hadronic scales, this is essentially negligible. Thus, to good approximation it is a movement only along the baryo-chemical potential direction. Starting from the vacuum, i. e. zero baryo-chemical potential, it can be shown exactly that nothing will happen before reaching a chemical potential of roughly a third of the nucleon mass. This is due to some analyticity properties of the free energy as a function of the masses of the lightest baryon, and the fact that it is made out of three quarks. This feature is called the silver-blaze feature. However, any small temperature will change this, and it is then only approximately true for chemical potentials much larger than the temperature, but smaller than this silver-blaze point.

After this point soon the nuclear liquid-gas transition is encountered. After this experimentally established point, as noted above, no fully reliable results are available. There are some reasons to believe that at least one further phase transition will be encountered, though even this is not sure. It is furthermore unclear whether this will happen at densities sill relevant for a neutron star, or significantly above it. However, model calculations, i. e. calculations using simplified versions of QCD, as well as comparisons to other theories which are similar to QCD, indicate that there could even exist many different phases, some amorphous, and some crystalline, in which besides the nucleons also other hadrons, like pions, kaons, and hyperons, may play a role.

The situation is much better regarding zero chemical[22] potential. This situation is relevant in the early universe, where though all matter in the universe is already present, and there is thus a sizable amount of baryons, the temperature is high enough to thermally produce baryon-anti-baryon pairs, reducing the chemical potential very close to zero.

This situation is good accessible experimentally by high-energy heavy-ion collisions, e. g. at the LHC with up to 2.4 TeV kinetic energy per nucleon for lead nuclei. In such

[22] As in most cases only the baryo-chemical potential is present, the word baryo is dropped. The only other chemical potentials of relevance in most cases are an isospin-chemical potential, which gives the difference between up and down quarks, and is therefore relevant in neutron stars, though it is even then only small because of the close similarity of up and down quark, strange-chemical potential, which determines how many more strange than anti-strange quarks appear, and electro-chemical potential, which indicates the presence of electrons.

an experimental setup, temperatures as high as 600-700 MeV with almost zero chemical potential, despite the 416 nucleons in the original nuclei, can be achieved. Furthermore, this situation poses no serious problems to numerical simulations. Hence, the knowledge of this axis is rather good.

It turns out that the physics depends significantly on the masses of the up and down quarks. Though this is also suspected for the remainder of the phase diagram, it is evident in this case. If both quarks are very heavy, there is a first-order phase transition at a temperature of about 250-300 MeV. As the quark masses become lighter, this temperature decreases, and the transition becomes a rapid cross-over at a temperature[23] of about 150-160 MeV, the aforementioned Hagedorn temperature. This is the situation for up and down quark masses observed in nature, the physical masses. The mass of the strange quark influences the precise values of the temperatures, but does not provide any qualitative influence, and the heavier quarks have even less relevance. If the quark masses are decreased further, the temperature still drops a little bit. More importantly, at some point the cross-over turns again into a phase transition, this time of second order, where it stays until zero quark masses.

What happens can be understood already in a simple picture. Temperature is classically nothing more than the kinetic energy of particles. In a quantum theory, temperature is just energy, which can also be converted to new particles. This will be exponentially suppressed with the mass of the created particles. Hence, the lightest particles will be most copiously produced. In QCD, these are the pions. These particles will have large kinetic energies, and will rapidly and repeatedly collide. Because of the asymptotic freedom of QCD, these scatterings will mainly be dominated by hard partonic scatterings, and thus almost perturbative. Thus, QCD becomes essentially as it behaves at high-energies. Especially, this implies that the effects of chiral symmetry breaking become reduced, and the quarks lose their effective mass, though not their current mass, at the phase transition or cross-over. In fact, in the limit of zero quark mass, the second order transition becomes a symmetry transition where chiral symmetry becomes restored. At the same time, since most collisions are hard and partonic, excited states become very unstable and in most cases it does not matter anymore that quarks are confined into hadrons. They act effectively as if they would no longer be confined. Thus, one speaks also of a deconfined phase, and calls the transition a deconfinement transition. However, since it is a cross-over, it is clear that qualitatively nothing has changed, but quantitatively it is a completely different situation.

[23]There is no unique definition of a cross-over temperature. This is just the temperature were most changes occur.

As a consequence of this dominance of the partonic degrees of freedom and asymptotic freedom actually the high-temperature thermodynamic behavior of the theory is essentially that of a free gas of quarks and gluons, a so-called Stefan-Boltzmann gas. The reason is mainly that the hard processes contribute to the free energy like the temperature to the fourth power, while all other effects contribute like the cube of the temperature, or even less. Hence, they become for thermodynamic bulk properties, i. e. extensive properties, irrelevant. Still, there are certain observables which are sensitive to non-trivial effects. Furthermore, the transition is very slow, and even at a few times the transition temperature even the bulk quantities are not yet fully dominated by the partonic processes. The system then behaves not like a gas, but rather like an almost ideal fluid.

This is the situation encountered in the early universe. While it cools done, it will go through this cross-over. Before that, it is essentially dominated by the quarks and gluons, and only afterwards it starts to be dominated by the hadrons. However, because the transition is a cross-over, it seems that the transition had little quantitative influence on the evolution of the universe.

The situation in the remainder of the phase diagram is not yet clear. It is possible to map out parts of it with heavy-ion collisions at lower energies. Because then less energy is available, less particles are produced, and therefore the baryon chemical potential is larger. Still, the accessible region is that of rather high temperatures, above those characteristic for neutron stars, and likely below the relevant chemical potentials. Also, numerical simulations start again to fail the larger the chemical potential becomes. Hence, the situation becomes less and less clear. What seems to be certain at the current time is that for quite some distance into the chemical potential direction little changes, and the cross-over remains at a temperature only slowly decreasing with increasing chemical potential. There are some speculations about a critical end-point, from which a phase boundary starts, which eventually meets with the chemical potential axis, but this is not yet settled. Other than that, the field is still wide open.

Chapter 6

Weak interactions

The last ingredient of the standard model are the weak interactions, which will again be a gauge interaction, though of quite different character than both the electromagnetic and the strong interaction. The name weak is also somewhat misleading, as the interaction is actually stronger than the electromagnetic one, but it is short-range, as is the strong interaction, though for entirely different reasons.

It is impossible to separate from any description of the weak interactions the recently established Higgs sector, as without it the weak interactions would be very different in character. In fact, without the Higgs sector, it would be very similar to the strong interaction. The Higgs effect introduces the only non-gauge interactions into the standard-model. Depending on personal taste, the interactions of the Higgs are not counted at all as true interactions, as many believe these are just low-energy effective realizations of a yet unknown gauge interaction, a single interaction, or, by counting the number of independent coupling constants, 13 interactions. This will become clearer in the following.

One of of the most interesting facts about the weak interactions is that they live on a quite different scale than electromagnetism and the strong interactions. While electromagnetism acts mainly on the scales of atomic physics, and the strong interactions at the scale of nuclear and hadronic physics, the typical scale of the weak interactions will be found to be of the order of 100 GeV. Since this is the highest known scale in particle physics yet, apart from the gravitational scale, it is widely believed that the study of the weak interaction will be the gateway to understand whether there is any other physics between this scale and gravity. This is the main reason it plays such a central role in modern particle physics.

One of the reasons why the weak interactions, or more commonly denoted as the electroweak sectors for reasons to become clear later, is rather unwieldy is that it is an aggregation of several, intertwined phenomena. Though each of them can be studied on its own, essentially none of them unfolds its full relevance without the others. This makes the electroweak sector often at first somewhat confusing. The different elements appearing will be a new gauge interaction, a mixing between this new interaction and QED, parity violation, the Higgs effect, and flavor changes.

6.1 β decay and parity violation

The weak interaction surfaced for the first time in certain decays, particularly the β-decay of free neutrons and of bound nuclei, which could not be explained with either the electromagnetic force nor the then newly discovered strong force. That this is impossible is directly seen from the quark content of the proton and the neutron: A transmutation of an up to a down quark is not possible in either interaction, because both are flavor-conserving. Furthermore, no hadrons are produced in the process, so this cannot be an exchange reaction. hence, another, third force has to be involved. Due to the small rate and also the very rarely seen consequences of this force in ordinary physics, this force was called weak force.

Though no additional hadrons are produced, other particles are involved. Especially the decay conserves electric charge by adding an electron to the final state, i. e. $n \to p+e^-$. However, for a free neutron this violates spin conservation. Furthermore, it is found that both the proton and the electron exhibit a continuous energy spectrum, and therefore energy conservation would be violated. The resolution of this is that the decay is indeed into three particles, and involves a new particle, the so-called electron-neutrino ν_e, actually an anti-electron-neutrino $\bar{\nu}_e$, $n \to p+e^- +\bar{\nu}_e$. Since the neutrino turns out to be a fermion of spin 1/2, this fixes both spin and energy conservation. The neutrino itself is a quite mysterious particles, and will be discussed further in section 6.2.

The first attempts of a field-theoretical formulation were based on a four-fermion interaction of type

$$G_F(\bar{p}\gamma_\mu n)(\bar{e}\gamma^\mu \nu_e), \tag{6.1}$$

where p, n, e, and ν_e represent the fields of the involved proton, neutron, electron, and the (anti)neutrino and are four-component spinors. The characteristic scale for the weak process was set by the Fermi constant, which is of order 1.14×10^{-5} GeV^{-2}. This is not a renormalizable interaction, and as such should be only the low-energy limit of an underlying renormalizable theory.

Besides the appearance of a new particle, the β-decay shows another quite surprising property: It violates parity. Thus, a β-decay is not operating in the same way as it would in the mirror. Experimentally, this can be observed by embedding the process into a magnetic field. A parity-conserving interaction, like electromagnetism, should yield the same result, irrespective of whether the spins are aligned or anti-aligned with the magnetic field[1], due to parity invariance. However, experimentally a difference is observed, indicating parity violation. More precisely, it was observed that a polarized neutron which decays will emit the electron preferentially in one direction. Therefore, the interaction must couple spin s and momenta, and would therefore have a contribution proportional to sp. However, the momenta of the decay products also depend on the invariant mass, p^2, and thus on a scalar contribution. Since therefore both scalars (scalar products of two vectors or two axial vectors) and pseudoscalars (products of a vector and an axial vector) appear imply that the interaction is not having a definite transformation behavior under parity, and is thus parity violating. In fact, it turned out that it is maximally parity violating.

To give this a more formal version, it is necessary to consider how coupling of fermions transform under parity. The parity transformation of a spinor is obtained by multiplying it with γ_0, the time-Dirac matrix. So, for spinors ψ the parity transformation is given by

$$P\psi = \gamma_0\psi.$$

Since furthermore $\gamma_0\gamma_\mu\gamma_0 = -\gamma_\mu$, the four-fermion coupling (6.1) will indeed transform as a scalar

$$P((\bar{\psi}\gamma_\mu\psi)(\bar{\psi}\gamma^\mu\psi)) = (\bar{\psi}\gamma_0\gamma_\mu\gamma_0\psi)(\bar{\psi}\gamma_0\gamma^\mu\gamma_0\psi) = (\bar{\psi}\gamma_\mu\psi)(\bar{\psi}\gamma^\mu\psi).$$

To obtain a pseudoscalar coupling, one of the vectors would have to be replaced by an axial vector. This can be obtained if there would be a matrix such that $\gamma_0\gamma_5\gamma_0 = \gamma_5$. In fact, such a matrix is given by

$$\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3.$$

This matrix anticommutes with all Dirac matrices,

$$\{\gamma_5, \gamma_\mu\} = 0.$$

As a consequence, the current

$$\bar{\psi}\gamma_5\gamma_\mu\psi,$$

is an axial vector, and can be used to obtain a pseudoscalar coupling.

However, this does not yet determine to which extent the weak interactions should be parity violating, and this can also not be predicted on basis of the standard model to be developed. Experimentally, however, it is found that parity is maximally violated. For massless particles (e. g. for neutrinos to a very good accuracy) this would imply that only

[1] In a practical case, the spin of the originating nucleus plays also a role, and a suitable choice is mandatory to see the effect simply. It was first observed in the β-decay of ^{60}Co.

one of the helicity states would be affected by the weak interactions. The helicity state for a spinor is projected out as

$$\frac{1 \pm \gamma_5}{2}\psi.$$

The sign is determined by whether left-handed or right-handed states should be selected. Experiment finds that only left-handed states are involved, and thus a minus sign is appropriate. Furthermore, the weak interactions are found to violate also the charge conjugation symmetry maximally. Hence, the sign is not reversed for the anti-particle state. Therefore, the correct four-fermion interaction version of the weak interactions would be (appropriately normalized)

$$\frac{G_F}{\sqrt{2}}\left(\bar{\psi}\frac{1-\gamma_5}{2}\gamma_\mu\frac{1-\gamma_5}{2}\psi\right)\left(\bar{\psi}\frac{1-\gamma_5}{2}\gamma^\mu\frac{1-\gamma_5}{2}\psi\right),$$

which therefore exhibits maximal violation of C and P individually, but conserves CP.

Hence, the new interaction is very different from either the strong or electromagnetic force: There are particles only responding to it, the neutrinos, and it violates both parity and flavor conservation. Furthermore, it is found to be very short range, much more than the strong interaction. On the other hand, electrons and neutrinos are both sensitive to it, but are not confined. Thus, the way how this range limitation is operating has to be different from the confinement mechanism of QCD.

6.2 Neutrinos

Before discussing the weak interaction itself. It is worthwhile to introduce the neutrinos first. Some further properties of them will be added later, in section 6.9, when the weak force has been discussed in some more detail.

So far, only one neutrino has been introduced, the electron-neutrino, a fermion of spin 1/2. There are actually two more flavors of neutrinos, the muon-neutrino ν_μ and the τ-neutrino ν_τ, which are again fermions of spin 1/2. As their name suggest, they are partnered with one lepton each, forming the last member of the particle generations, which are conventionally assigned to be the sets $\{u, d, e, \nu_e\}$, $\{c, s, \mu, \nu_\mu\}$, and $\{t, b, \tau, \nu_\tau\}$. These combinations are somewhat arbitrary, and mainly motivated by the masses of the particles. Another motivation is that flavor is, though violated, still approximately conserved for leptons and neutrinos. In then turns out that, if electrons carry an electron flavor number of one and positrons of -1, electro-neutrinos likewise carry an electron flavor number of 1 and anti-electron neutrinos of -1. Hence, e. g. the β-decay, conserves this flavor number. Likewise, there are approximately conserved flavors for muon-like and tauon-like particles.

As a consequence, neutrinos are usually also included in the term leptons, though not always.

When it comes to masses, the neutrinos actually are very surprising. To date, it is only known for sure that their mass is below[2] 2 eV. A direct measurement of the mass of the neutrinos is complicated by the fact that they are only interacting through the weak force, which reduces cross-sections by order of magnitudes compared to the leptons and the quarks. However, it was possible to measure the mass differences of the neutrinos in a way to be explained in section 6.9. But so far, it was only possible to measure differences without knowing which of the possible differences are measured. The two values are about 0.009 eV and 0.05 eV. This implies that at least one of the neutrinos has a mass of 0.05 eV, though it is not clear, which of them. It is furthermore not yet clear, whether the other two to or just one is much lighter. If the electron neutrino would be the heaviest, there are very good chances to measure its mass directly within the next decade. Since the other neutrino species are much harder to produce in larger numbers, this will take much longer if it is is either of the other ones. It is also, in principle, still possible that one of the neutrinos could be massless, though there is yet no strong evidence for this possibility.

6.3 Flavor-changing currents

As already observed, the weak interaction changes the flavor of down quarks to up quarks. Likewise, processes have been observed that changes strange to charm and bottom to top, and vice versa, kinematics always permitting. There are also processes observed which changes, say, charm to up quarks. However, these processes are strongly suppressed, and will be neglected for now. They will be added below in section 6.8.

After neglecting these processes, the flavor changes only occur within one generation. Since in this process also one unit of electric charge is transferred to leptons, these are called flavor-changing charged current, or FCCC for short. In addition, there are also flavor-changing neutral currents, FCNC, but they belong to the class of processes postponed.

This observation strongly suggest that the quark flavors in any generation can be regarded just as two states, a doublet, under the weak interaction, similar to the isospin. Hence, a new symmetry is introduced, the weak isospin symmetry, under which doublets are formed from the two quarks of each generation, and also the electron and the neutrino of each generation. This doublet structure has been already observed with the isospin relation (5.4), and it will be seen later how the electromagnetic charge enters this. Note,

[2]This is a conservative upper bound. There are good indications that it is at least an order of magnitude smaller.

however, that in FCCC also the masses play a role, and for reasons to be discussed later massless particles would not participate. Therefore flavor changes for the leptons occur much more rarely because of the very small neutrino masses, though have been observed nonetheless.

6.4 W and Z bosons

The doublet structure can be regarded as a charge, the weak isospin charge. The weak force couples to this charge. The interaction which mediates the neutron decay appears to be described by the interaction of four fermions, an up quark, a down quark, an electron, and an anti-electron neutrino, as given in (6.1). The strength of this interaction is characterized by a new coupling constant, the Fermi constant G_F, which has units of inverse energy squared, and a value of 1.16×10^{-5} GeV^2, corresponding to an energy scale of about 293 GeV, though usually this energy scale is reduced by a factor of $2^{\frac{1}{4}}$ to 246 GeV for reasons of conventions. At any rate, the energy scale is much larger than any hadronic scale. The only other object encountered so far with a similar scale is the top quark, though this appears at the current time unrelated.

Since weak charges are not confined, another mechanism is required to give the weak interaction a finite range. According to section 5.1, a possibility is that the charge carriers are massive. Another problem with this theory was that such a theory is not renormalizable, in the sense of section 3.7. Hence, it would not be possible to obtain meaningful predictions from it at energies of similar size or larger than 246 GeV, which would already cover any top quark physics. It was therefore early on argued that this cannot be the underlying theory, and at a scale of around 100 GeV there must be a different theory present.

This indeed turns out to be true. In fact, it was suggested that there should exist, similarly to QED, an exchange boson. However, to produce such a scale and range, they would have to be massive with a mass M_W. Assuming a coupling constant g' for the process, the scale would be set by expressions of the type

$$G_F \approx \frac{g'^2}{M_W^2}, \tag{6.2}$$

indicating a mass scale of about 100 GeV for the process, if g' is not too different from one. This already sets the scale of the weak interactions. Furthermore, the appearance of a mass-scale indicates that the weak interactions will have only a limited range, of the order of $1/M_W$, and would be screened beyond this scale. Its range is therefore much shorter than that of any other force.

Over the time other weak effects were found, all characterized by the scale G_F. In particular, after some time the postulated interaction bosons have been observed directly. There are two charged ones, the $W^{\pm}$, and a neutral one Z^0, with masses about 80 and 91 GeV, respectively, and where the superscripts indicate their electric charge. In fact, the weak interaction turns out to have a very similar structure as QED and QCD in the sense that it is a gauge theory, where additional gauge bosons appear. The doublet structure already suggests as the underlying gauge group one with two different charges and three gauge bosons. It finally turns out that the correct gauge group is SU(2). That the $W^{\pm}$ carry electric charge indicate a relation to QED, which will be discussed in section 6.7.

However, masses are not permitted for gauge bosons, or else gauge invariance would be broken. This breaking has severe effects, making the theory non-renormalizable. Therefore, despite the experimental verification of the existence of the weak gauge bosons and their masses, this would not improve the theoretical description. This was only achieved after gauge theories have been combined with the Goldstone theorem described in section 5.7.

To understand that this is the right way, consider a photon A_μ coupled to some external current j in the Lagrangian

$$\mathcal{L} = -\frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) - j^\mu A_\mu.$$

It will be this current, and its generalization, which will be instrumental in obtaining a massive gauge boson and in addition a gauge symmetry, and thus a superficially massive vector boson. However, since these two concepts are usually contradictory, this will lead to a hiding of the gauge symmetry, similarly as the global symmetry has been hidden in section 5.7.

The corresponding equation of motion for the photon is

$$\partial^2 A^\nu - \partial^\nu \partial_\mu A^\mu = j^\nu. \tag{6.3}$$

Furthermore, the field should, somehow, satisfy the condition

$$(\partial^2 + M^2)A^\nu = 0. \tag{6.4}$$

To proceed further, it is useful to choose a particular gauge, in this case the Landau-Lorentz gauge $\partial_\mu A^\mu = 0$. Of course, fixing a gauge is a perfectly acceptable way to perform a calculation, provided the calculation is not violating gauge invariance at any point. Then the results for gauge-invariant quantities, like the number of physical polarizations, will be valid. In the present case, doing a calculation without fixing the gauge is at best tedious. The results can be translated to the general case of an arbitrary gauge, but it should be kept in mind that any relation on gauge-variant quantities, like the gauge field A_μ, will only be valid in this particular gauge. However, statements about gauge-invariant quantities, like, e. g., the field-strength tensor $F_{\mu\nu}$ in this Abelian example, will hold in any gauge.

After imposing this gauge, the equation of motion (6.3) for the spatial components becomes

$$\partial^2 \vec{A} = -\vec{j}. \tag{6.5}$$

To simplify matters further, restrict to a time-independent situation. Then all time-derivatives vanish, and the equation (6.4) takes the form

$$\partial^2 \vec{A} = M^2 \vec{A}. \tag{6.6}$$

From the equations (6.5) and (6.6) it is clear that the current must satisfy the condition

$$\vec{j} = -M^2 \vec{A}$$

in order that the photon (in this gauge) becomes apparently massive and at the same time the gauge invariance is not broken. It is left to find some way of producing the appropriate current. Physically, the origin of such a current being proportional to $\vec{A}$ is due to the response of a medium to the acting electromagnetic fields. E. g., this is realized by the Meissner effect in superconductivity. Therefore, giving a mass to the photon requires a medium which provides a response such that the photon becomes damped and can therefore propagate only over a finite distance. In the electroweak case, the role of this medium will be taken by a condensate of Higgs particles.

This is a screening process as can be seen from the following simple example. Consider the Maxwell-equation

$$\vec{\partial} \times B = \vec{j} = -M^2 \vec{A}.$$

Taking the curl on both sides and using $\vec{\partial} \times \vec{A} = \vec{B}$ yields

$$\partial^2 \vec{B} = M^2 \vec{B}.$$

In a one-dimensional setting this becomes

$$\frac{d^2}{dx^2} B = M^2 B,$$

which is immediately solved by

$$B = \exp(-Mx).$$

Thus the magnetic field is damped on a characteristic length $1/M$, the screening length. The inverse of the screening length, being a mass in natural units, is therefore the characteristic energy scale, or mass, of the damping process. However, by this only an effective mass has been obtained. The other vital ingredient for a massive vector boson is a third polarization state. Similarly, also this other degree of freedom will be provided effectively by the medium, as will be discussed below.

To make the preceding discussion applicable to elementary particle physics, it is useful to write them down covariantly. For this purpose, select the Landau gauge $\partial_\mu A^\mu = 0$. Then the Maxwell equation takes the form

$$\partial^2 A_\mu = j_\mu.$$

The necessary requirement that, at least approximately, the current has a component proportional to A_μ is then directly obtained by considering the current for a charged, scalar field, which has the form

$$j_\mu = iq(\phi^+ \partial_\mu \phi - (\partial_\mu \phi^+)\phi) - 2q^2 A_\mu |\phi|^2,$$

which can be derived as the Noether current from the corresponding Lagrangian

$$\begin{aligned} \mathcal{L} &= D_\mu \phi^+ D^\mu \phi \\ D_\mu &= \partial_\mu + iqA_\mu. \end{aligned}$$

Thus, the Maxwell equation for A_μ becomes

$$\partial^2 A_\mu = iq(\phi^+ \partial_\mu \phi - (\partial_\mu \phi^+)\phi) - 2q^2 A_\mu |\phi|^2.$$

This will have precisely the required form, if the modulus of the field, $|\phi| = v$ can be arranged to have a non-zero value in the vacuum. However, this will only be possible, if its expectation value obeys

$$\langle 0|\phi|0\rangle \neq 0 \tag{6.7}$$

in the selected gauge. In this case, the photon would acquire an effective mass of $(2q^2v^2)^{1/2}$. And by this, the photon is screened out, making the effect short-ranged. How to implement this will be discussed next.

6.5 The Higgs effect and the Higgs boson

Motivated by the previous discussion, and by the considerations in section 5.7 it appears reasonable to start with a scalar field with self-interactions, call it Higgs field, and couple it to a gauge field. When coupling the Higgs to a gauge field, symmetry breaking, now called the Higgs (or more precisely Englert-Brout-Higgs, and several other names) effect becomes more complicated, but at the same time also more interesting, than in the case without gauge fields.

For simplicity, start with an Abelian gauge theory, coupled to a single, complex scalar, the so-called Abelian Higgs model, before going to the full electroweak and non-Abelian case. This theory has the Lagrangian

$$\mathcal{L} = \frac{1}{2}((\partial_\mu + iqA_\mu)\phi)^+(\partial^\mu + iqA^\mu)\phi - \frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) + \frac{1}{2}\mu^2\phi^+\phi - \frac{1}{2}\frac{\mu^2}{f^2}(\phi^+\phi)^2. \quad (6.8)$$

Note that the potential terms are not modified by the presence of the gauge-field. Therefore, the extrema have still the same form and values as in the previous case, at least classically. However, it cannot be excluded that the quartic $\phi^+\phi A_\mu A^\mu$ term strongly distorts the potential. This does not appear to be the case in the electroweak interaction, and this possibility will therefore be ignored.

To make the consequences of the Higgs effect transparent, it is useful to rewrite the scalar field as[3]

$$\phi(x) = \left(\frac{f}{\sqrt{2}} + \rho(x)\right)\exp(i\alpha(x)).$$

This is just another reparametrization for the scalar field, compared to η and ξ previously. It is such that at $\rho = 0$ this field configuration will be a classical minimum of the potential for any value of the phase α. Inserting this parametrization into the Lagrangian (6.8) yields

$$\mathcal{L} = \frac{1}{2}\partial_\mu\rho\partial^\mu\rho + \frac{1}{2}\left(\frac{f}{\sqrt{2}} + \rho\right)^2\partial_\mu\alpha\partial^\mu\alpha + qA^\mu\left(\frac{f}{\sqrt{2}} + \rho\right)^2\partial_\mu\alpha + \frac{q^2}{4}A_\mu A^\mu\left(\frac{f}{\sqrt{2}} + \rho\right)^2 - \frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) + \frac{1}{2}\mu^2\left(\frac{f}{\sqrt{2}} + \rho\right)^2 - \frac{1}{2}\frac{\mu^2}{f^2}\left(\frac{f}{\sqrt{2}} + \rho\right)^4$$

This is an interesting structure, where the interaction pattern of the photon with the radial and angular part are more readily observable.

Now, it is possible to make the deliberate gauge choice

$$\partial_\mu A^\mu = -\frac{1}{q}\partial^2\alpha. \quad (6.9)$$

[3]Note that if the space-time manifold is not simply connected and/or contains holes, it becomes important that α is only defined modulo 2π. For flat Minkowski (or Euclidean) space, this is of no importance. However, it can be important, e. g., in finite temperature calculations. It is definitely important in ordinary quantum mechanics, where, e. g., the Aharanov-Bohm effect and flux quantization depend on this.

This is always possible. It is implemented by first going to Landau gauge and then perform the gauge transformation

$$\begin{aligned} A_\mu &\rightarrow A_\mu + \frac{1}{q}\partial_\mu\alpha \\ \phi &\rightarrow \exp(-i\alpha)\phi. \end{aligned}$$

This gauge choice has two consequences. The first is that it makes the scalar field real everywhere. Therefore, the possibility of selecting the vacuum expectation value of ϕ to be real is a gauge choice. Any other possibilities, e. g. purely imaginary, would be an equally well justified gauge choice. This also implies that the actual value of the vacuum expectation value f of ϕ is a gauge-dependent quantity, and will vanish when not fixing a gauge. The Lagrangian then takes the form[4]

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}\partial_\mu\rho\partial^\mu\rho + \frac{q^2}{4}A_\mu A^\mu\left(\frac{f}{\sqrt{2}}+\rho\right)^2 \\ &-\frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) + \frac{1}{2}\mu^2\left(\frac{f}{\sqrt{2}}+\rho\right)^2 - \frac{1}{2}\frac{\mu^2}{f^2}\left(\frac{f}{\sqrt{2}}+\rho\right)^4 \end{aligned} \quad (6.10)$$

i. e., the second term has now exactly the form of a screening term, and yields an effective mass $qf/4$ for the photon field. Furthermore, if ρ could be neglected, the Lagrangian would be just

$$\mathcal{L} = \frac{q^2f^2}{8}A_\mu A^\mu - \frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu),$$

i. e., the one of a massive gauge field. Together with the gauge condition for A_μ, which links the longitudinal part of the gauge boson to the now explicitly absent degree of freedom α, this implies that the field A_μ acts now indeed as a massive spin-1 field. Furthermore, this Lagrangian is no longer gauge-invariant. This is not a problem, as it was obtained by a gauge choice. Thus, it is said that the gauge symmetry is hidden. Its consequences are still manifest, e. g., the mass for the gauge boson is not a free parameter, but given by the other parameters in the theory. By measuring such relations, it is in principle possible to determine whether a theory has a hidden symmetry or not.

Colloquial, the hiding of a symmetry is also referred to as the breaking of the symmetry in analogy with the case of a global symmetry. However, a theorem, Elitzur's theorem, actually forbids this to be literally true.

The Goldstone theorem of section 5.7, actually guarantees that a mass will be provided to each gauge boson associated with one of the hidden directions at the classical level. This

[4]Upon quantization, additional terms are introduced due to the gauge-fixing. At the classically level studied here, they can be ignored.

hidden direction is here the phase, which is massless. This masslessness of the Goldstone boson is actually guaranteed by the Goldstone theorem. Hence, a massless Goldstone boson is effectively providing a mass to a gauge boson by becoming its third component via gauge-rotation, and vanishes by this from the spectrum. It is the tri-linear couplings which provide the explicit mixing terms delivering the additional degree of freedom for the gauge boson at the level of the Lagrangian. Hence, the number of degrees of freedom is preserved in the process: In the beginning there were two scalar and two vector degrees of freedom, now there is just one scalar degree of freedom, but three vector degrees of freedom. The gauge-transformation made nothing more than to shift one of the dynamic degrees of freedom from one field to the other. This was possible due to the fact that both the scalar and the photon are transforming non-trivially under gauge-transformations.

This can, and will be, generalized below for other gauges. In general, it turns out that for a covariant gauge there are indeed six degrees of freedom, four of the vector field, and two from the scalars. Only after calculating a process it will turn out that certain degrees of freedom cancel out, yielding just a system which appears like having a massive vector particle and a single scalar.

Note that though the original scalar field ϕ was charged as a complex field, the radial excitation as the remaining degree of freedom is actually no longer charged: The coupling structure appearing in (6.10) is not the one expected for a charged field.

However, the choice (6.9), which is called the unitary gauge, is extremely intransparent and cumbersome for most actual calculations. The reason is that the quantization of the the theory in this gauge requires the introduction of an infinite number of further terms into the Lagrangian, though with fixed coefficients.

A more convenient possibility, though at the cost of having unphysical degrees of freedom which only cancel at the end, are 't Hooft gauges. To define this gauge once more the decomposition

$$\phi = (\chi, f + \eta)$$

for the scalar field is useful. The Lagrangian then takes the form, up to constant terms,

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{(gf)^2}{2}A_\mu A^\mu + \frac{1}{2}\partial_\mu\eta\partial^\mu\eta - \mu^2\eta^2 + \frac{1}{2}\partial_\mu\chi\partial^\mu\chi - gfA^\mu\partial_\mu\chi \\ &\quad +2igA^\mu(\chi\partial_\mu\chi + \eta\partial_\mu\eta) - 2g(gf)\eta A_\mu A^\mu - g^2(\chi^2+\eta^2)A_\mu A^\mu - \frac{1}{2}\frac{\mu^2}{f^2}(\chi^2+\eta^2)^2.\end{aligned} \tag{6.11}$$

There are a number of interesting observations. Again, there is an effective mass for A_μ, due to the four-field interaction term. Secondly, only the η field has a conventional mass term, the mass-term for the χ field has canceled with the two-field-two-condensate term from the quartic piece of the potential. Finally, there will be terms of type $fA^\mu\partial_\mu\chi$. This implies that a photon can change into a χ while moving, with a strength proportional to the condensate f, and thus mixing between one of the Higgs degrees of freedom and the photon occurs. Therefore, the photon and this scalar, the would-be Goldstone boson, will mix. Note finally that though many more interaction terms have appeared, none of them has any new free parameter, as a consequence of the now hidden symmetry. The only quantity which looks like a new quantity is the condensate value f, though it is classically uniquely determined by the shape of the potential. In a quantum calculation, it can also be determined, but not in perturbation theory, where it remains a fit parameter.

The gauge condition to be used for quantizing this Lagrangian is then given by

$$\partial_\mu A^\mu = qf\xi\chi, \tag{6.12}$$

the so-called renormalizable or 't Hooft gauge, where ξ is an arbitrary parameter, a so-called gauge parameter. This gauge choice removes at the quantum level the mixing between the photon and the Higgs. The resulting theory has then the massive Higgs field, a massless Goldstone field, and a massive photon field. The surplus degrees of freedom are, as stated, not actually there, but will cancel at the quantum level.

The form of the Lagrangian (6.11) already indicates how the weak gauge bosons will receive their mass. The necessary Higgs boson has by now indeed been discovered, and its mass is found to be about 125 GeV.

6.6 Parity violation and fermion masses

So far, the construction seems to reproduce very well and elegantly the observed phenomenology of the electroweak interactions.

However, there is one serious flaw. The Dirac equation for fermions, like leptons and quarks, has the form

$$0 = (i\gamma^\mu D_\mu - m)\psi = \left(i\gamma^\mu D_\mu + \frac{1-\gamma_5}{2}m + \frac{1+\gamma_5}{2}m\right)\psi.$$

The covariant derivative is a vector under a weak isospin gauge transformation, and so is the spinor $(1-\gamma_5)/2\psi$. However, the spinor $(1+\gamma_5)/2\psi$ is a singlet under such a gauge transformation. Hence, not all terms in the Dirac equation transform covariantly, and therefore weak isospin cannot be a symmetry for massive fermions. Another way of observing this is that the mass term for fermions in the Lagrangian can be written as

$$\mathcal{L}_{m\psi} = m(\psi_L\psi_R - \psi_R\psi_L),$$

and therefore cannot transform as a gauge singlet.

However, massless fermions can be accommodated in the theory. Since the observed quarks and (at least almost) all leptons have a mass, it is therefore necessary to find a different mechanism which provides the fermions with a mass without spoiling the isoweak gauge invariance.

A possibility to do so is by invoking the Higgs-effect also for the fermions and not only for the weak gauge bosons. By adding an interaction

$$\mathcal{L}_h = g_f\phi_k\bar{\psi}_i\left(\alpha^k_{ij}\frac{1-\gamma_5}{2}+\beta^k_{ij}\frac{1+\gamma_5}{2}\right)\psi_j + \left(g_f\phi_k\bar{\psi}_i\left(\alpha^k_{ij}\frac{1-\gamma_5}{2}+\beta^k_{ij}\frac{1+\gamma_5}{2}\right)\psi_j\right)^+,$$

this is possible. The constant matrices α and β have to be chosen such that the terms become gauge-invariant. Calculating their precise form is tedious, but straightforward. If, in this interaction, the Higgs field acquires a vacuum expectation value, $\phi = f+$quantum fluctuations, this term becomes an effective mass term for the fermions, and it is trivially gauge-invariant. Alongside with it comes then an interaction of Yukawa-type of the fermions with the Higgs particle. However, the interaction strength is not a free parameter of the theory, since the coupling constants are uniquely related to the tree-level mass m_f of the fermions by

$$g_f = \frac{\sqrt{2}m_f}{f}.$$

But the 12 coupling constants for the three generations of quarks and leptons are not further constrained by the theory, introducing a large number of additional parameters in the theory. Though phenomenologically successful, this is the reason why many feel this type of description of the electroweak sector is insufficient. However, even if it would be incorrect after all, it is an acceptable description at energies accessible so far, and thus has become part of the standard model.

6.7 Weak isospin, hypercharge, and electroweak unification

It is now clear that a theoretical description of the weak interactions requires a gauge theory, as well as some matter field(s) to hide it and provide a mass to the weak gauge bosons. The fact that there are three distinct gauge bosons, two Ws and one Z indicates that the gauge theory has to be more complex than just the U(1) gauge theory of QED. It will thus be a non-Abelian gauge theory. Furthermore, since two of them are charged the connection to the electromagnetic interactions will not be trivial, and there will be some kind of mixing. All of these aspects will be taken into account in the following.

6.7.1 Constructing the gauge group

Phenomenological, as discussed in section 6.3, the weak interactions provides transitions of two types. One is a charge-changing reaction, which acts between two gauge eigenstates. In case of the leptons, these charge eigenstates are almost (up to a violation of the order of the neutrino masses) exactly a doublet - e. g., the electron and its associated electron neutrino. Therefore, the gauge group of the weak interaction should provide a doublet representation. In case of the quarks this is less obvious, but also they furnish a doublet structure. Hence, an appropriate gauge group for the weak interaction will contain at least SU(2). Since only the left-handed particles are affected, this group is often denoted as $\mathrm{SU}(2)_L$, but this index will be dropped here. Therefore, there are three doublets, generations, of leptons and quarks, respectively in the standard model. However, the mass eigenstates mix all three generations, as will be discussed in detail below.

Since the electric charge of the members of the doublets differ by one unit, the off-diagonal gauge bosons, the W, must carry a charge of one unit. Furthermore, such a gauge group has three generators. The third must therefore be uncharged, as it mediates interactions without exchanging members of a doublet.

The quantum number distinguishing the two eigenstates of a doublet is called the third component of the weak isospin t, and will be denoted by t_3 or I_W^3. Therefore, the gauge group of the weak interactions is called the weak isospin group.

However, the weak gauge bosons are charged. Therefore, ordinary electromagnetic interactions have to be included somehow. Ordinary electromagnetism has as gauge group the Abelian U(1). The natural ansatz for the gauge group of the electroweak interactions is thus the gauge group SU(2)×U(1)[5]. With this second factor-group comes a further quantum number, which is called the hypercharge y. The ordinary electromagnetic charge is then given by

$$eQ = e\left(t_3 + \frac{y}{2}\right). \tag{6.13}$$

Thus, the ordinary electromagnetic interaction must be somehow mediated by a mixture of the neutral weak gauge boson and the gauge boson of the U(1). This is dictated by observation: It is not possible to adjust otherwise the quantum numbers of the particles such that experiments are reproduced. The hypercharge of all left-handed leptons is -1, while the one of left-handed quarks is $y = +1/3$.

Right-handed particles are neutral under the weak interaction. In contrast to the $t = 1/2$ doublets of the left-handed particle, they belong to a singlet, $t = 0$. All in all, the following assignment of quantum numbers for charge, not mass, eigenstates will be

[5]Actually, the correct choice is $\mathrm{SU}(2)/\mathrm{Z}_2\times\mathrm{U}(1)$, although this difference is very rarely of relevance.

necessary to reproduce the experimental findings:

- Left-handed neutrinos: $t = 1/2$, $t_3 = 1/2$, $y = -1$ ($Q = 0$)
- Left-handed leptons: $t = 1/2$, $t_3 = -1/2$, $y = -1$ ($Q = -1$)
- Right-handed neutrinos: $t = 0$, $t_3 = 0$, $y = 0$ ($Q = 0$)
- Right-handed leptons: $t = 0$, $t_3 = 0$, $y = -2$ ($Q = -1$)
- Left-handed up-type (u, c, t) quarks: $t = 1/2$, $t_3 = 1/2$, $y = 1/3$ ($Q = 2/3$)
- Left-handed down-type (d, s, b) quarks: $t = 1/2$, $t_3 = -1/2$, $y = 1/3$ ($Q = -1/3$)
- Right-handed up-type quarks: $t = 0$, $t_3 = 0$, $y = 4/3$ ($Q = 2/3$)
- Right-handed down-type quarks: $t = 0$, $t_3 = 0$, $y = -2/3$ ($Q = -1/3$)
- W^+: $t = 1$, $t_3 = 1$, $y = 0$ ($Q = 1$)
- W^-: $t = 1$, $t_3 = -1$, $y = 0$ ($Q = -1$)
- Z: $t = 1$, $t_3 = 0$, $y = 0$ ($Q = 0$)
- γ: $t = 0$, $t_3 = 0$, $y = 0$ ($Q = 0$)
- Gluon: $t = 0$, $t_3 = 0$, $y = 0$ ($Q = 0$)
- Higgs: a complex doublet, $t = 1/2$ with weak hypercharge $y = 1$. This implies zero charge for the $t_3 = -1/2$ component, and positive charge for the $t_3 = 1/2$ component and negative charge for its complex conjugate

This concludes the list of charge assignments for the standard model particles. The Higgs case will be special, and will be detailed in great length below.

Since at the present time the photon field and the Z boson are not yet readily identified, it is necessary to keep the gauge boson fields for the SU(2) and U(1) group differently, and these will be denoted by W and B respectively. The corresponding pure gauge part of the electroweak Lagrangian will therefore be

$$
\begin{aligned}
\mathcal{L}_g &= -\frac{1}{4}G^a_{\mu\nu}G_a^{\mu\nu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\
F_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\
G^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g f^{abc} W^b_\mu W^c_\nu,
\end{aligned}
$$

where g is the weak isospin gauge coupling. f^{abc} are the structure constants of the weak isospin gauge group, which is just the SU(2) gauge group.

Coupling matter fields to these gauge fields proceeds using the ordinary covariant derivative, which takes the form

$$D_\mu = \partial_\mu + \frac{ig}{2}\tau_a W_\mu^a + \frac{ig'y}{2}B_\mu,$$

where g' is the hypercharge coupling constant, which is modified by the empirical factor y. Note that y is not constrained by the gauge symmetries, and its value is purely determined by experiment. Thus, why it takes the rational values it has is an unresolved question to date. However, this fact would come about naturally, if the weak gauge group would originate from a different gauge group at higher energies, say SU(2)×U(1)⊂SU(3), which is hidden to the extent that all other fields charged under this larger gauge group are effectively so heavy that they cannot be observed with current experiments.

The matrices τ_a are determined by the representation of SU(2) in which the matter fields are in. For a doublet, these will be the Pauli-matrices. For the adjoint representation, these would be given by the structure constants, $\tau^a_{bc} = f^{abc}$, and so on. For fermions, of course, this covariant derivative is contracted with the Dirac matrices γ_μ. Precisely, to couple only to the left-handed spinors, it will be contracted with $\gamma_\mu(1-\gamma_5)/2$ for the W_μ^a term and with γ_μ for the kinetic and hypercharge term. By this, the phenomenological couplings are recovered in the low-energy limit, as the exchange of a massive gauge boson then becomes proportional to $1/M^2$, thus recovering the Fermi-coupling g^2/M^2. How this mass disappears in the case of the Abelian gauge group will be discussed next.

6.7.2 Hiding the electroweak symmetry

To have a viable theory of the electroweak sector it is necessary to hide the symmetry such that three gauge bosons become massive, and one uncharged one remains massless. Though this can be of course arranged in any gauge, it is most simple to perform this in the unitary gauge. Since three fields have to be massive, this will require three pseudo-Goldstone bosons. Also, since empirically two of them have to be charged, as the $W^\pm$ bosons are charged, the simplest realization is by coupling a complex doublet scalar field, the Higgs field, to the electroweak gauge theory

$$\begin{aligned} \mathcal{L}_h &= \mathcal{L}_g + (D_\mu\phi)^+ D^\mu\phi + V(\phi\phi^+) \\ \phi &= \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1 + i\phi_2 \\ \phi_3 + i\phi_4\end{pmatrix} \end{aligned} \qquad (6.14)$$

where the field ϕ is thus in the fundamental representation of the gauge group. Its hypercharge will be 1, an assignment which will be necessary below to obtain a massless photon. The potential V can only depend on the gauge-invariant combination $\phi^+\phi$, and thus can, to be renormalizable, only contain a mass-term and a quartic self-interaction. The mass-term must be again of the wrong sign (imaginary mass), such that there exists a possibility for the ϕ field to acquire a (gauge-dependent) vacuum-expectation value.

To work in unitary gauge it is best to rewrite the Higgs field in the form

$$\phi = e^{\frac{i\tau^a\alpha_a}{2}}\begin{pmatrix}0\\ \rho\end{pmatrix}.$$

There are now the three α^a fields and the ρ field. Performing a gauge transformation such that the phase becomes exactly canceled, and setting $\rho = f + \eta$ with f constant makes the situation similar to the one in the Abelian Higgs model. Note that by a global gauge transformation the component with non-vanishing expectation value can be selected still at will.

The mass will be made evident by investigating the equation of motion for the gauge bosons

$$\begin{aligned}\partial^2 W_\mu^a - \partial_\mu(\partial^\nu W_\nu^a) &= j_\mu^{a\phi} + j_\mu^{aW}\\ \partial^2 B_\mu - \partial_\mu(\partial^\nu B_\nu) &= j_\mu^y,\end{aligned}$$

where j^W contains contributions which involve only the Ws, while j^ϕ contains all remaining terms. Only the latter will be relevant below, and therfeore for it the subscript ϕ will be dropped. Thus, it is necessary to determine the weak isospin current j_μ^a and the hypercharge current j_μ^y. In contrast to the Abelian case this current will now also contain contributions from the self-interactions of the W gauge bosons. However, at tree-level these can be ignored.

The contribution of the Higgs field to the currents[6] are

$$\begin{aligned}j_\mu^a &= \frac{ig}{2}(\phi^+\tau^a D_\mu\phi - (D_\mu\phi)^+\tau^a\phi)\\ j_\mu^y &= \frac{ig'}{2}(\phi^+ D_\mu\phi - (D_\mu\phi)^+\phi).\end{aligned}$$

The appearance of τ^a makes manifest that the current j_μ^a is a weak isovector current, while their absence signifies that the hypercharge current j_μ^y is a weak isoscalar current.

[6]Of course, all fields with non-zero weak isospin will contribute to this current. But only the Higgs-field will do so classically.

Expanding the covariant derivatives yields[7]

$$
\begin{aligned}
j^a_\mu &= \frac{ig}{2}(\phi^+\tau^a\partial_\mu\phi - (\partial_\mu\phi)^+\tau^a\phi) - \frac{g^2}{2}\phi^+\phi W^a_\mu - \frac{gg'y}{2}\phi^+\tau^a\phi B_\mu \qquad (6.15)\\
j^y_\mu &= \frac{ig'}{2}(\phi^+\partial_\mu\phi - (\partial_\mu\phi)^+\phi) - \frac{gg'y}{2}\phi^+\tau_a W^a_\mu\phi - \frac{(g'y)^2}{2}\phi^+\phi B_\mu.
\end{aligned}
$$

It should be noted that the two equations are not only coupled by the Higgs field, but by both, the W^a_μ and B_μ gauge fields. Selecting now the Higgs field to be of the form

$$
\phi = \begin{pmatrix} 0 \\ \frac{f}{\sqrt{2}} \end{pmatrix} + \text{quantum fluctuations},
$$

where the quantum fluctuations vanish at tree-level will provide the vacuum expectation value

$$
\langle 0|\phi|0\rangle = \begin{pmatrix} 0 \\ \frac{f}{\sqrt{2}} \end{pmatrix}.
$$

This is then as in the Abelian case, and indeed sufficient to provide a mass for the gauge bosons.

Ignore for a moment the hypercharge contribution. Then the term linear in W^a_μ on the right-hand side of (6.15) provides a term

$$
-M^2 W^a_\mu = -\left(\frac{gf}{2}\right)^2 W^a_\mu,
$$

and thus a mass $M_W = gf/2$ is provided. Thus, the equation of motion for the W^a_μ field takes the form

$$
(\partial^2 + M_W^2)W^a_\mu - \partial_\mu(\partial^\nu W^a_\nu) = j^a_\mu(f, \phi, W^a_\mu, ...),
$$

where the mass-term has been removed from the current. Thus, formally this is the equation of motion for a massive gauge boson, interacting with itself and other fields due to the current j^a_μ. As has been seen in the Abelian case, the additional degree of freedom has been selected by the choice of gauge, and is provided by the phase of the Higgs field.

Next it is necessary to check what happens when including also the B_μ part once more. Since the $t_3 = -1/2$ component of the Higgs field delivers the mass, this component should not be electrically charged. According to the relation (6.13) thus the assignment of the hypercharge $y = 1$ for the Higgs field is a-posterior justified. In the full current of the Higgs field, besides the contribution $\phi^+\phi W^a_\mu$ also the contribution $gg'y/2B_\mu\phi^+\tau^a\phi$ appears. Since the matrices τ^1 and τ^2 are off-diagonal it follows that

$$
(0, f)\tau^{1,2}\begin{pmatrix} 0 \\ f \end{pmatrix} = 0.
$$

[7] Using $\{\tau^a, \tau^b\} = 2\delta^{ab}$.

Thus only in the third component of the current a contribution due to the vacuum expectation value of the Higgs field appears. As a consequence, despite the appearance of the additional B_μ gauge boson the result for the mass (at tree-level) for the off-diagonal $W_\mu^{1,2}$ bosons is not changed. However, this is not the case for the third component. The relevant part of the equation of motion reads

$$\partial^2 W_\mu^3 - \partial_\mu(\partial^\nu W_\nu^3) = -\left(\frac{g^2 f^2}{2}\right)^2 W_\mu^3 + \frac{gf}{2}\frac{g'f}{2}B_\mu.$$

Consequently, also the relevant part of the equation of motion for the hypercharge gauge boson is of similar structure

$$\partial^2 B_\mu - \partial_\mu(\partial^\nu B_\nu) = -\left(\frac{g'f}{2}\right)^2 B_\mu + \frac{gf}{2}\frac{g'f}{2}W_\mu^3.$$

Hence, even in the vacuum the equations are coupled, and thus both gauge bosons mix. These equations can be decoupled by changing variables as

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

which are the fields given the name of the photon A_μ and the Z boson Z_μ. The mixing parameter θ_W is the (Glashow-)Weinberg angle θ_W, and is given entirely in terms of the coupling constants g and g' as

$$\begin{aligned} \tan\theta_W &= \frac{g'}{g} \\ \cos\theta_W &= \frac{g}{\sqrt{g^2+g'^2}} \\ \sin\theta_W &= \frac{g'}{\sqrt{g^2+g'^2}}. \end{aligned}$$

Using the inverse transformations

$$\begin{aligned} W_\mu^3 &= \frac{A_\mu \cos\theta_W + Z_\mu \sin\theta_W}{\cos\theta_W \sin\theta_W} \\ B_\mu &= \frac{A_\mu \sin\theta_W - Z_\mu \cos\theta_W}{\cos\theta_W \sin\theta_W} \end{aligned}$$

it is possible to recast the equations of motion. The equation of motion for the photon takes the form

$$\partial^2 A_\mu - \partial_\mu(\partial^\nu A_\nu) = \sin\theta_W j_\mu^3.$$

Here, j_μ^3 contains all contributions which are not providing a mass to the gauge bosons. Thus, the photon is effectively massless. The price paid is that it now has also direct interactions with the weak gauge bosons. On the other hand, the equation for the Z boson takes the form

$$\partial^2 Z_\mu - \partial_\mu \partial^\nu Z_\nu = -\frac{f^2(g^2+g'^2)}{4} Z_\mu + \cos\theta_W j^3_\mu .$$

Thus, the Z boson acquires a mass larger than the $W^{1,2}$ boson. In particular the relation is

$$M_Z \cos\theta_W = M_W .$$

According to the best measurements this effect is, however, only of order 15%. At the same time the self-interaction of the Z boson with the other weak bosons is reduced compare to the one of the original W^3_μ boson.

Of course, this changes also the form of the coupling to the neutral fields for matter. In particular, the symmetry of SU(2) is no longer manifest, and the $W^{1,2}$ cannot be treated on the same footing as the neutral bosons. E. g., the neutral part of the coupling in the covariant derivative now takes the form

$$D^{(N)}_\mu = \partial_\mu + ig\sin\theta_W A_\mu \left(t_3 + \frac{y}{2}\right) + i\frac{g}{\cos\theta_W} Z_\mu \left(t_3 \cos^2\theta_W - \frac{y}{2}\sin^2\theta_W\right).$$

Using the relation (6.13), it is possible to identify the conventional electric charge as

$$e = g\sin\theta_W ,$$

i. e., the observed electric charge is smaller than the hypercharge. It should be noted that this also modifies the character of the interaction. While the interaction with the photon is purely vectorial, and the one with the $W^\pm$ bosons remains left-handed (axial-vector), the interaction with the Z boson is now a mixture of both, and the mixing is parametrized by the Weinberg angle.

Note that the masslessness of the photon is directly related to the fact that the corresponding component of the Higgs-field has no vacuum expectation value,

$$\left(\frac{y}{2} + \frac{\tau^3}{2}\right) \langle 0|\phi|0\rangle = 0,$$

and thus the vacuum is invariant under a gauge transformation involving a gauge transformation of A_μ,

$$\langle 0|\phi'|0\rangle = \langle 0|\exp(i\alpha(y/2+\tau_3))\phi|0\rangle .$$

Thus, the original SU(2)×U(1) gauge group is hidden, and only a particular combination of the subgroup U(1) of SU(2) and the factor U(1) is not hidden, but a manifest gauge symmetry of the system, and thus this U(1) subgroup is the stability group of the electroweak gauge group SU(2)×U(1). It is said, by an abuse of language, that the gauge group SU(2)×U(1) has been broken down to U(1). Since this gauge symmetry is manifest, the associated gauge boson, the photon A_μ, must be massless. If, instead, one would calculate without the change of basis, none of the gauge symmetries would be manifest. However, the mixing of the B_μ and the W_μ^3 would ensure that at the end of the calculation everything would come out as expected from a manifest electromagnetic symmetry.

Also, this analysis is specific to the unitary gauge. In other gauges the situation may be significantly different formally. Only when determining gauge-invariant observables, like scattering cross-sections or the masses of gauge invariant bound-states, like positronium, everything will be the same once more.

6.8 CP violation and the CKM matrix

In any strong or electromagnetic process the quark (and lepton) flavor is conserved. E. g., the strangeness content is not changing. This is not true for weak processes. It is found that they violate the flavor number of both, quarks and leptons. In case of the leptons, this effect is suppressed by the small neutrinos masses involved, but in case of quarks this is a significant effect.

Considering the weak decays of a neutron compared to that of a strange Λ, it is found that the relative strengths can be expressed.as

$$\begin{aligned} g_{\Delta S=0} &= g' \cos\theta_C \\ g_{\Delta S=1} &= g' \sin\theta_C \end{aligned}$$

where g' is a universal strength parameter for the weak interactions, its coupling constant. The angle parametrizing the decay is called the Cabibbo angle. A similar relation also holds in the leptonic sector for the muon quantum number

$$\begin{aligned} g_{\Delta\mu=0} &= g' \cos\theta_C^L \\ g_{\Delta\mu=1} &= g' \sin\theta_C^L, \end{aligned}$$

where, however, $\sin\theta_C^L$ is almost one, while in the quark sector $\sin\theta_C$ is about 0.22. Corresponding observations are also made for other flavors.

This result implies that the mass eigenstates of the matter particles are not at the same time also weak eigenstates, but they mix. Hence, on top of the P-violating and C-violating factors of $(1-\gamma_5)/2$, it is necessary to include something into the interaction which provides this mixing. This can be done by introducing a flavor-dependent unitary coupling matrix

$$G' = \frac{g'}{\sqrt{2}} \begin{pmatrix} 0 & \cos\theta_C^{(L)} & 0 \\ \cos\theta_C^{(L)} & 0 & \sin\theta_C^{(L)} \\ 0 & -\sin\theta_C^{(L)} & 0 \end{pmatrix}.$$

This is equivalent to just use a doublet

$$\begin{pmatrix} u & d\cos\theta_C + s\sin\theta_C \end{pmatrix},$$

e. g., in the quark sector. Such a doublet structure can be associated with (weak) charges $Q_u = 1/2$ and $Q_{ds} = -1/2$. This is just the weak isospin.

Hence, the flavor (and mass) eigenstates of the quarks are effectively rotated by a unitary matrix. For two generations, this matrix, the Cabibbo matrix is uniquely given by

$$V_C = \begin{pmatrix} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{pmatrix},$$

with again the Cabibbo angle θ_C, with a value of about $\sin\theta_C \approx 0.22$. For three generations, there exist no unique parametrization of the mixing matrix, which is called the Cabibbo-Kobayashi-Maskawa (CKM) matrix. The interaction is thus parametrized as

$$(\bar{u}_L, \bar{c}_L, \bar{t}_L)^T \gamma^\mu W_\mu^+ V_{\text{CKM}} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} + \text{h.c.}$$

$$V_{\text{CKM}} = V_L^u V_L^{d\dagger} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}.$$

The absolute values of the entries have been measured independently, and are given by

$$\begin{matrix} |V_{ud}| = 0.9743(3) & |V_{us}| = 0.2252(9) & |V_{ub}| = 4.2(5)\times 10^{-3} \\ |V_{cd}| = 0.23(2) & |V_{cs}| = 1.01(3) & |V_{cb}| = 41(1)\times 10^{-3} \\ |V_{td}| = 8.4(6)\times 10^{-3} & |V_{ts}| = 43(3)\times 10^{-3} & |V_{tb}| = 0.89(7) \end{matrix}.$$

Thus, the CKM matrix is strongly diagonal-dominant, especially towards larger quark masses, and within errors compatible with a unitary matrix.

To make the unitarity more explicit, it is common to recast the CKM matrix into the following form

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & e^{-i\delta_{13}}s_{13} \\ -s_{12}c_{23} - e^{i\delta_{13}}c_{12}s_{23}s_{13} & c_{12}c_{23} - e^{i\delta_{13}}s_{12}s_{23}s_{13} & s_{23}c_{13} \\ s_{12}s_{23} - e^{i\delta_{13}}c_{12}c_{23}s_{13} & -c_{12}s_{23} - e^{i\delta_{13}}s_{12}c_{23}s_{13} & c_{23}c_{13} \end{pmatrix}. \quad (6.16)$$

$$c_{ij} = \cos\theta_{ij}$$

$$s_{ij} = \sin\theta_{ij}$$

To have only 4 free parameters (θ_{12}, θ_{13}, θ_{23}, and δ_{13}) requires not only unitarity, but also to exploit some freedom in redefining the absolute phases of the quark fields. Testing whether this matrix is indeed unitary by measuring the nine components individual is currently recognized as a rather sensitive test for physics beyond the standard model. The condition for unitarity can be recast into a form that a sum of three functions of the angles has to be π, forming a triangle. Thus, testing for the unitarity, and thus standard-model compatibility of CP violations, is often termed measuring the unitarity triangle.

The presence of this matrix also gives rise to the possibility that not only C and P are violated separately, but that also the compound symmetry CP is violated (and therefore also T by virtue of the CPT-theorem). That occurs to lowest order in perturbation theory by a box-diagram exchanging the quark flavor of two quarks by the exchange of two W bosons.

That such a process violates CP can be seen as follows. The process just described is equivalent to the oscillation of, e. g., a $d\bar{s}$ bound state into a $s\bar{d}$ bound state, i. e., a neutral kaon K^0 into its anti-particle $\bar{K}^0$. The C and P quantum numbers of both particles are P$=-1$, C$=1$ and P$=1$, C$=1$, respectively, and thus CP$=-1$ and CP$=1$. Thus, any such transition violates CP. Performing the calculation of the corresponding diagram yields that it is proportional to the quantity

$$\begin{aligned}\chi &= \sin\theta_{12}\sin\theta_{23}\sin\theta_{13}\cos\theta_{12}\cos\theta_{23}\cos^2\theta_{13}\sin\delta_{13}\\ &\quad\times(m_t^2-m_c^2)(m_t^2-m_u^2)(m_c^2-m_u^2)(m_b^2-m_s^2)(m_b^2-m_d^2)(m_s^2-m_d^2).\end{aligned}$$

Thus, such a process, and thus CP violation is possible if there is mixing at all (all θ_{ij} non-zero) with a non-trivial phase δ_{13}, and the masses of the quarks with fixed charge are all different. They may be degenerate with ones of different charge, however. Since such oscillations are experimentally observed, this already implies the existence of a non-trivial quark-mixing matrix. The value of δ_{13} is hard to determine, but is roughly 1.2.

Within the standard model, there is no explanation of this mixing, however, and thus these are only parameters.

6.9 Neutrino oscillations and the PMNS matrix

A consequence of the mixing by the CKM matrix is that it is possible to start at some point with a particle of any flavor, and after a while it has transformed into a particle of a different flavor. The reason for this is a quantum effect, and proceeds essentially through emission and reabsorption of a virtual $W^{\pm}$. In quantum mechanical terms for a two-body system, this can be easily deduced. Take a Hamiltonian H as

$$H = \begin{pmatrix} H_0 & \Delta \\ \Delta & H_0 \end{pmatrix}.$$

This leads to a time-evolution operator

$$U(t) = e^{-iH_0 t} \begin{pmatrix} \cos(\Delta t) & -i\sin(\Delta t) \\ -i\sin(\Delta t) & \cos(\Delta t) \end{pmatrix}.$$

Under time evolution an initial pure state $(1, 0)$ will therefore acquire a lower component, if Δ is non-zero. On the other hand, if the composition of a pure state after an elapsed time is measured, it is possible to obtain the size of Δ. The probability P to find a particle in the state $(0, 1)$ after a time t is given by

$$P = \sin^2(\Delta t). \tag{6.17}$$

In the standard model, the corresponding expression for the transition probability involves the mass difference between the two states. To lowest order it is given for the two-flavor case by

$$P = \sin^2\left(\frac{\Delta_m^2 L}{4E}\right)\sin^2(2\theta),$$

where Δ_m is the mass difference between both states, E is the energy of the original particle, L is the distance traveled, and θ is the corresponding Cabbibo angle. If the probability, the energy and the distance is known for several cases, both the Cabbibo angle and the mass difference can be obtained. Of course, both states have to have the same conserved quantum numbers, like electrical charge.

The same calculation can be performed in the more relevant three-flavor case. The obtained transition probability from a flavor f_i to a flavor f_j is more lengthy, and given by

$$P(f_i \to f_j, L, E) = \sum_k |V_{jk}|^2 |V_{ik}|^2 + 2\sum_{k>l} |V_{jk}V_{ik}^* V_{il} V_{jl}^*| \cos\left(\frac{\Delta_{m_{kl}^2}}{2E} L - \arg(V_{jk}V_{ik}^* V_{il} V_{jl}^*)\right)$$

This sum shows that the process is sensitive to all matrix elements, including the CP-violating phase, of the CKM matrix, while it cannot determine the sign of the mass differences.

For the quarks, the first observation of CKM effects were due to decays. Such oscillations have also been found later, and studied to determine the matrix elements more precisely. E. g., it is possible with a certain probability that a particle oscillates from a short-lived state to a long-lived state. This is the case for the so-called K-short K_S and K-long K_L kaons, mixed bound states of an (anti-)strange and an (anti-)down quark. This has been experimentally observed, but the distance L is of laboratory size, about 15 m for K_L and 15 cm for K_S, giving a Δ_m of about 3.5×10^{-12} MeV. However, in this case the effect is rather used for a precision determination of the mixing angle, since the mass can be accurately determined using other means.

In the lepton sector, these oscillations were the original hint for non-zero neutrino masses. As it is visible, such oscillations only occurs if the involved particles have a mass. Thus, only with massive neutrinos there was chance to see the effect. But there is no direct determination of their mass, and the best results so far is an upper limit on the order of 2 eV from the β-decay of tritium. Observing these oscillations then indicated the presence of neutrino masses, and that there is a CKM-like matrix also in the lepton sector, as also otherwise there would be no mixing - for a unit matrix the above given formula reduces to the first term only.

With this, it is possible to determine the mass difference of neutrinos. It is found that $|\Delta_{m_{12}}| = 0.0087$ eV and $|\Delta_{m_{23}}| = 0.049$ eV. As only the squares can be determined, it is so far no possible to establish which neutrino is the heaviest, and if one of them is massless. Still, the mass difference of 0.05 eV indicates that with an increase in sensitivity by a factor of 40 it can be determined in decay experiments whether the electron neutrino is the heaviest. If it is, the mass hierarchy is called inverse, as the first flavor, corresponding to the electron, is lightest. Otherwise, it would be called a normal hierarchy. As a side remark, these tiny mass differences imply that the oscillation lengths L are typically macroscopically, and of the order of several hundred kilometers and more for one oscillation.

Aside with these mass differences, it was also possible to determine some of the elements of the CKM matrix in the lepton sector, which is called Pontecorvo-Maki-Nakagawa-Sakata, or PMNS, matrix. In terms of the parametrization (6.16), the values of the known three angels are

$$\begin{aligned}\sin^2\theta_{12} &= 0.31(2)\\ \sin^2\theta_{23} &= 0.4(1)\\ \sin^2\theta_{13} &= 0.025(4)\end{aligned}$$

while the value of δ_{13} was not yet measured, and it is thus not clear whether there is also CP violation in the lepton sector. However, the values of the other angles imply that with the sensitivity of current experiments, it should be possible to get a reasonable estimate of δ_{13} until the mid 2020s.

The values of the three real angles are quite astonishing. They imply that, irrespective of the value of δ_{13}, the PMNS matrix is not strongly diagonal dominant. Thus, the lepton flavors mix much stronger than the corresponding quark flavors, and only the very small values of the neutrino masses reduce the oscillation probabilities so strongly that the effect is essentially not seen for electrons, muons, and tauons, and even requires the very weakly interacting neutrinos, which can move over long distances, to observe it at all.

Again, at the current time there is no understanding for why this is the case, nor why the quark sector and the lepton sector are so extremely different also in this respect, and not only for the hierarchy of masses.

6.10 The weak interactions as a gauge theory

Combining everything, the electroweak sector of the standard model, the Glashow-Salam-Weinberg theory is of the following type: It is based on a gauge theory with a gauge group G, which is SU(2)×U(1), weak isospin and hypercharge. There are left-handed and right-handed fermions included, which are differently charged under the gauge group G. These are doublets for the left-handed quarks and leptons, and singlets for the right-handed quarks and leptons for the weak isospin. There is also the scalar Higgs field which is again a doublet.

To formulate the Lagrangian of the electroweak sector of the standard model requires hence a number of steps. First of all, the vacuum expectation value for the Higgs field is chosen to be $f/\sqrt{2}$. Its direction is chosen such that it is manifest electrically neutral. That is provided by the requirement

$$Q\phi = 0 = \left(\frac{\tau_3}{2} + \frac{y_\phi}{2}\right)\phi = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = 0.$$

The field is then split as

$$\phi = \begin{pmatrix} \phi^+ \\ \frac{1}{\sqrt{2}}(f + \eta + i\chi) \end{pmatrix},$$

with $\phi^- = (\phi^+)^+$, i. e., its hermitiean conjugate. The fields ϕ^+ and ϕ^- carry integer electric charge plus and minus, while the fields η and χ are neutral. Since a non-vanishing value of f leaves only a U(1) symmetry manifest, $\phi^\pm$ and χ are the would-be Goldstone bosons. This leads to the properties of the Higgs fields and vector bosons as discussed previously.

The fermions appear as left-handed doublets in three generations

$$L_i^L = \begin{pmatrix} \nu_i^L \\ l_i^L \end{pmatrix}$$
$$Q_i^L = \begin{pmatrix} u_i^L \\ d_i^L \end{pmatrix},$$

where i counts the generations, l are the leptons e, μ and τ, ν the corresponding neutrinos ν_e, ν_μ, ν_τ, u the up-type quarks u, c, t, and d the down-type quarks d, s, b. Correspondingly exist the right-handed singlet fields l_i^R, ν_i^R, u_i^R, and d_i^R. Using this basis the Yukawa interaction part reads

$$\mathcal{L}_Y = -\bar{L}_i^L G_{ij}^{lr} l_j^R \phi^r + \bar{L}_i^L G_{ij}^{\nu r} \nu_j^R \phi^r - \bar{Q}_i^L G_{ij}^{ur} u_j^R \phi^r + \bar{Q}_i^L G_{ij}^{dr} d_j^R \phi^r + h.c..$$

The matrices G are connected to mass matrices by

$$M_{ij}^l = \frac{1}{\sqrt{2}} G_{ij}^l f \qquad M_{ij}^\nu = \frac{1}{\sqrt{2}} G_{ij}^\nu f \qquad M_{ij}^u = \frac{1}{\sqrt{2}} G_{ij}^u f \qquad M_{ij}^d = \frac{1}{\sqrt{2}} G_{ij}^d f.$$

It is then possible to transform the fermion fields[8] f into eigenstates of these mass-matrices, and thus mass eigenstates, by a unitary transformation

$$f_i^{fL} = U_{ik}^{fL} f_k^L \qquad (6.18)$$
$$f_i^{fR} = U_{ik}^{fR} f_k^R,$$

for left-handed and right-handed fermions respectively, and f numbers the fermion species l, ν, u, and d and i the generation. The fermion masses are therefore

$$m_{fi} = \frac{1}{\sqrt{2}} \sum_{km} U_{ik}^{fL} G_{km}^f (U^{fR})_{mi}^+ f.$$

In this basis the fermions are no longer charge eigenstates of the weak interaction, and thus the matrices U correspond to the CKM matrices. In fact, in neutral interactions which are not changing flavors always combinations of type $U^{fL}(U^{fL})^+$ appear, and thus they are not affected. For flavor-changing (non-neutral) currents the matrices

$$V^q = U^{uL}(U^{dL})^+ \qquad (6.19)$$
$$V^l = U^{\nu L}(U^{lL})^+$$

[8] The vacuum expectation value f of the Higgs field, carrying no indices, should not be confused with the fermion fields f_i^f, which carry various indices, f denoting the fermion class.

remain, providing the flavor mixing. Finally, the electric charge is given by

$$e = \sqrt{4\pi\alpha} = g' \sin\theta_W = g\cos\theta_W,$$

with the standard value $\alpha \approx 1/137$.

Putting everything together, the lengthy Lagrangian for the electroweak standard model emerges:

$$\begin{aligned}
\mathcal{L} &= \bar{f}_i^{rs}(i\gamma^\mu\partial_\mu - m_f)f_i - eQ_r\bar{f}_i^{rs}\gamma^\mu f_i^{rs}A_\mu \\
&+ \frac{e}{\sin\theta_W\cos\theta_W}(I_{Wr}^3\bar{f}_i^{rL}\gamma^\mu f_i^{rL} - \sin^2\theta_W Q_r\bar{f}_i^{rs}\gamma^\mu f_i^{rs})Z_\mu \\
&+ \frac{e}{\sqrt{2}\sin\theta_W}(\bar{f}_i^{rL}\gamma^\mu V_{ij}^r f_j^{rL}W_\mu^+ + \bar{f}_i^{rL}\gamma^\mu (V^r)_{ij}^+ f_j^{rL}W_\mu^-) \\
&- \frac{1}{4}|\partial_\mu A_\nu - \partial_\nu A_\mu - ie(W_\mu^- W_\nu^+ - W_\nu^- W_\mu^+)|^2 \\
&- \frac{1}{4}\left|\partial_\mu Z_\nu - \partial_\nu Z_\mu + ie\frac{\cos\theta_W}{\sin\theta_W}(W_\mu^- W_\nu^+ - W_\nu^- W_\mu^+)\right|^2 \\
&- \frac{1}{2}\left|\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+ - ie(W_\mu^+ A_\nu - W_\nu^+ A_\mu) + ie\frac{\cos\theta_W}{\sin\theta_W}(W_\mu^+ Z_\nu - W_\nu^+ Z_\mu)\right|^2 \\
&+ \frac{1}{2}\left|\partial_\mu(\eta + i\chi) - i\frac{e}{\sin\theta_W}W_\mu^-\phi^+ + iM_Z Z_\mu + \frac{ie}{2\cos\theta_W\sin\theta_W}Z_\mu(\eta + i\chi)\right|^2 \\
&+ |\partial_\mu\phi^+ + ieA_\mu\phi^+ - ie\frac{\cos^2\theta_W - \sin^2\theta_W}{2\cos\theta_W\sin\theta_W}Z_\mu\phi^+ - iM_W W_\mu^+ - \frac{ie}{2\sin\theta_W}W_\mu^+(\eta + i\chi)|^2 \\
&- f^2\eta^2 - \frac{ef^2}{\sin\theta_W M_W}\eta(\phi^-\phi^+ + \frac{1}{2}|\eta + i\chi|^2) \\
&- \frac{e^2 f^2}{4\sin^2\theta_W M_W^2}(\phi^-\phi^+ + \frac{1}{2}|\eta + i\chi|^2)^2 \\
&- \frac{em_{ri}}{2\sin\theta_W M_W}(\bar{f}_i^{rs}f_i\eta - 2I_{Wr}^3 i\bar{f}_i^{rs}\gamma_5 f_i^{rs}\chi) \\
&+ \frac{e}{\sqrt{2}\sin\theta_W}\frac{m_{ri}}{M_W}(\bar{f}_i^{rR}V_{ij}^r f_j^{rL}\phi^+ + \bar{f}_i^{rL}(V^r)_{ij}^+ f_j^{rR}\phi^-) \\
&+ \frac{e}{\sqrt{2}\sin\theta_W}\frac{m_{ri}}{M_W}(\bar{f}_i^{rL}V_{ij}^r f_j^{rR}\phi^+ + \bar{f}_i^{rR}(V^r)_{ij}^+ f_j^{rL}\phi^-),
\end{aligned} \tag{6.20}$$

where a sum over fermion species r is understood and I_W^3 is the corresponding weak isospin quantum number.

Note that this Lagrangian is invariant under the infinitesimal gauge transformation

$$\begin{aligned}
A_\mu &\to A_\mu + \partial_\mu\theta^A + ie(W_\mu^+\theta^- - W_\mu^-\theta^+) \qquad (6.21)\\
Z_\mu &\to Z_\mu + \partial_\mu\theta^Z - ie\frac{\cos\theta_W}{\sin\theta_W}(W_\mu^+\theta^- - W_\mu^-\theta^+)\\
W_\mu^\pm &\to W_\mu^\pm + \partial_\mu\theta^\pm \mp i\frac{e}{\sin\theta_W}(W_\mu^\pm(\sin\theta_W\theta^A - \cos\theta_W\theta^Z) - (\sin\theta_W u - \cos\theta_W Z_\mu)\theta^\pm)\\
\eta &\to \eta + \frac{e}{2\sin\theta_W\cos\theta_W}\chi\theta^Z + i\frac{e}{2\sin\theta_W}(\phi^+\theta^- - \phi^-\theta^+)\\
\chi &\to \chi - \frac{e}{2\sin\theta_W\cos\theta_W}(v+\eta)\theta^Z + \frac{e}{2\sin\theta_W}(\phi^+\theta^- + \phi^-\theta^+)\\
\phi^\pm &\to \phi^\pm \mp ie\phi^\pm\left(\theta^A + \frac{\sin^2\theta_W - \cos^2\theta_W}{2\cos\theta_W\sin\theta_W}\theta^Z\right) \pm \frac{ie}{2\sin\theta_W}(f+\eta\pm i\chi)\theta^\pm\\
f_i^{f\pm L} &\to f_i^{f\pm L} - ie\left(Q_i^\pm\theta^A + \frac{\sin\theta_W}{\cos\theta_W}\left(Q_i^f \mp \frac{1}{2\sin^2\theta_W}\right)\theta^Z\right)f_i^{\pm L} + \frac{ie}{\sqrt{2}\sin\theta_W}\theta^\pm V_{ij}^\pm f_j^{\pm L}\\
f_i^{fR} &\to -ieQ_i^f\left(\theta^A + \frac{\sin\theta_W}{\cos\theta_W}\theta^Z\right)f_i^{fR}. \qquad (6.22)
\end{aligned}$$

The $\pm$ index for the left-handed fermion fields counts the isospin directions. The infinitesimal gauge functions θ^α are determined from the underlying weak isospin θ^i and hypercharge θ^Y gauge transformations by

$$\begin{aligned}
\theta^\pm &= \frac{1}{g'}(\theta^1 \mp i\theta^2)\\
\theta^A &= \frac{1}{g}\cos\theta_W\theta^Y - \frac{1}{g'}\sin\theta_W\theta^3\\
\theta^Z &= \frac{1}{g'}\cos\theta_W\theta^3 + \frac{1}{g}\sin\theta_W\theta^Y.
\end{aligned}$$

It is now straightforward to upgrade the Lagrangian of the electroweak sector of the standard model (6.20) to the full Lagrangian of the standard model by adding the one for the strong interactions

$$\begin{aligned}
\mathcal{L}_s &= -\frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu} - g''\bar{f}_i^{rs}\gamma^\mu\omega^a f_i^{rs}G_\mu^a\\
G_{\mu\nu}^a &= \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g'' h^{abc}G_\mu^b G_\nu^c,
\end{aligned}$$

where the generators ω^a and the structure constants h^{abc} belong to the gauge group of the strong interactions, the so-called color group SU(3)[9], and g'' is the corresponding coupling constants. G_μ^a are the gauge fields of the gluons, and the fermions now have also an (implicit) vector structure in the strong-space, making them three-dimensional color vectors or singlets, for quarks and leptons, respectively.

[9] Actually SU(3)/Z_3, similar to the case of the weak isospin group.

Note that the number of Higgs fields is larger than the minimal number required: Only two (real) degrees of Higgs are necessary to write down a consistent theory of the weak interactions. However, in this case only one of the weak gauge bosons would become massive. Hence, experiments requires four degrees of freedom. This introduces an additional global symmetry between the four Higgs degrees of freedom, similar to the flavor symmetries of the fermions. It is called a custodial symmetry, as in absence of other interactions it would lead to mass-degenerate W and Z bosons. However, QED breaks this symmetry explicitly, and leads, as has been shown above, to the slight mass difference between W and Z bosons.

6.11 Anomalies

As noted in section 4.9, it is possible that a symmetry is broken by the quantization. Such a breaking is called an anomalous breaking, mainly for the reason that this was not expected to occur in the beginning. Today, it is clear that such a breaking is deeply linked into the quantization process, and is in most cases also related to the necessity for renormalization.

It is now possible to consider an anomaly for a global or a local symmetry. Anomalies for global symmetries do not pose any inherent problems. Indeed, the axial part $\mathrm{U}_A(1)$ of the chiral symmetry described in section 4.12 is broken anomalously in the standard model. This breaking is directly observable consequences. In that particular case, it leads to a much larger partial decay width of the π^0 into two photons, as it would be without this anomaly. Also the large mass of the η' meson, as indicated in section 5.11 can be shown to originate from the same anomaly.

The situation is much more severe for an anomalous breaking of a local symmetry. If this occurs, this implies that while the classical theory is gauge-invariant, the quantum version would not be so. Hence, the quantum version would depend on the gauge chosen classically, and observables would become gauge-dependent. This would not yield a consistent theory, and would therefore be discarded as a theory of nature. Thus, only theories without local anomalies are considered for particle physics.

Considering theories of the type the standard model, i. e. a number of gauge fields with fermions and scalars, it turns out that the possibility of gauge anomalies is tied to the presence of fermions. Theories with only gauge fields and scalars do not develop anomalies. Furthermore, it can be shown that only theories with parity violations can be affected by local anomalies. Finally, anomalies can only occur for certain gauge groups[10], and not for

[10] In total, the Lie groups U(1), SU($N > 2$), Sp(N), and O($2 < N < 6$) are affected.

arbitrary ones, and only for particular assignments of charges to the fermions. However, this affected combination of charge assignments and gauge groups includes the one of the standard model. Thus, in principle, the standard model would be anomalous.

Fortunately, there is an escape route left. Though indeed such anomalies occur, it is possible that if there are more than one anomaly, the consequences of these anomalies cancel each other. If such an anomaly cancellation occurs, the theory becomes anomaly-free, and is again well-defined. Given the gauge groups and charge assignments of the standard model, the condition that all anomalies are canceled and the standard model is consistent is

$$\sum_f Q_f = N_g\left((0-1)+N_c\left(\frac{2}{3}-\frac{1}{3}\right)\right) = N_g\left(-1+\frac{N_c}{3}\right) = 0,$$

where the sum is over all fermion flavors, quarks and leptons alike, n_g is the number of generations, Q_f is the electric charge of each flavor f, and N_c is the number of colors. As can be seen, the anomalies indeed cancel, and they do so for each generation individually. However, the cancellation requires a particular ratio of the quark electric charges and the lepton electric charges as well as the number of colors. Since the anomalies originate from the parity-violating interaction, all three forces are involved in guaranteeing the anomaly-freedom of the standard model.

This fact has led to many speculations. The most favored one is that this indicates an underlying structure, which provides a common origin to all three force of the standard model. Such theories are known as grand-unifying theories (GUT). Such theories have one underlying gauge group, which would be spontaneously broken at high energies, and therefore appear like three separate forces at low energies. This reasoning is supported by the fact that the running gauge couplings of all three forces all approach a similar size at about 10^{15} GeV, the GUT scale, which suggests a common origin at this energy scale.

A second explanation is that only a theory like the standard model, which such a fine balance, has all the features necessary to permit sentient life to arise to recognize this feature. In a universe with different laws of nature, where this kind of balance is not met nobody would be there to observe it. This is called the anthropomorphic principle. Though it cannot be dispelled easily, it is not as compelling an argument, as it does not explain why such a universe should exist at all. This problem is often circumvented by the requirement that actually many universes with all kinds of laws of nature exist, and we just happen to be in one where we could exist.

The third possibility is, of course, that all of this is just coincidence, and nature is just this way.

At any rate, future experiments will decided, which of these three options is realized. In fact, experiments have already ruled out the simplest candidates for grand-unified theories, though many remain.

6.12 Landau-poles, triviality, and naturalness

As has been noted, the three gauge couplings of the standard model depend on energy. The strong and weak coupling diminish at high energies, and both interactions are asymptotically free. However, both have very different energy behaviors: The strong interaction becomes stronger towards the infrared, while the weak interaction reaches a maximum, and then diminishes also towards small energies, mostly due to the screening by the Higgs effect. The electromagnetic coupling, however, rises towards large energies.

Assume for a moment that the three coupling unify at some point, as discussed in the previous section 6.11, and that the resulting GUT is asymptotically free. Hence, the high-energy behavior of the gauge interactions becomes well-defined. Then there are still 13 non-gauge interactions in the standard model left: The four-Higgs self-interaction, and the 12 Yukawa couplings between the Higgs and the fermions[11]. As all these interactions are not of a gauge type, they are all not asymptotically free. Hence, all of them rise with energy, though even the largest one, the top-Higgs-Yukawa coupling, very slowly. As a consequence, perturbation theory fails at some scale due to the emergence of Landau poles in these couplings[12].

This rise is strongly dependent on the mass of the Higgs. If the Higgs mass increases, the couplings start to rise quicker. For a Higgs mass of around 1 TeV, perturbation theory will already fail at 1 TeV, as then the four-Higgs coupling develops the Landau pole at the same energy. At the same time a Higgs of this mass will have a decay width of the same order as its mass. This shows that the physics in the Higgs sector is very sensitive to the values of the parameters, in fact quadratically. This is in contrast to the remainder of the standard model. Both the fermion and gauge properties are only logarithmically sensitive to the parameters.

To achieve that the Higgs mass is of the same order as the W and Z mass requires that the parameters in the standard model are very finely tuned, to some ten orders of magnitude. This is the so-called fine-tuning problem. In fact, if random values for the

[11]It is assumed for simplicity here that all neutrinos are massive, though the experimental data are still compatible with one of them being massless.

[12]Occasionally one finds in the literature the argument of perturbativity, i. e. that any decent theory describing high energies must have weak coupling, and be perturbatively treatable. However, there is no reason for this assumption except technical ones. Hence, such statements express a hope rather than any reasonable physics.

couplings would be taken, the Higgs mass was would usually end up at masses orders of magnitudes larger than the W and Z mass. Why this is not the case is called the hierarchy problem: Why is there no large hierarchy between the electroweak scale and the Higgs mass scale? This is sometimes rephrased as the naturalness problem: Without any prior information, it would naively be expected that all parameters in a theory are of the same order of magnitude. This is not the case in the standard model. Why do the values deviate so strongly from their 'natural' value?

There is another problem connected with these questions. If only the Higgs sector is regarded, the theory is called a ϕ^4 theory, as it has the same structure as the Lagrangian (5.5), except that there are four scalar fields. It is found that in such a theory quantum effects always drive the self-interaction term in the potential to zero, i. e. the quantum theory is non-interacting, even if the classical theory is interacting. Theories in which this occurs are called trivial. Such a triviality can be removed, if the theory is equipped with an intrinsic energy cutoff, which introduces an additional parameter, but restricts the validity of the theory to below a certain energy level. It is then just a low-energy-effective theory of some underlying theory.

It is not clear whether this triviality problem persists when the Higgs sector is coupled to the rest of the standard model. If it does, the necessary cut-off for the standard model will again depend sensitively on the mass of the Higgs boson. Practically, this is not a serious issue, as with the present mass of the Higgs of roughly 125 GeV this triviality cut-off can be easily as large as 10^{19} GeV, far beyond the current reach of experiments and observations. Still, it remains as a doubt on how fundamental the standard model, and especially the Higgs particle, really is, in accordance with the problems introduced by the necessity of renormalization as described in section 3.7.

6.13 Baryon-number violation

Our current understanding of the origin of the universe indicates that it emerged from a big bang, a space-time singularity where everything started as pure energy. Why is then not an equal amount of matter and anti-matter present today, but there is a preference for matter? CP violation explains that there is indeed a preference for matter over anti-matter, but the apparent conservation of lepton and baryon number seems to indicate that this is only true for mesons and other states which do not carry either these quantum numbers. This impression is wrong, as, in fact, there is a process violating baryon (and lepton) number conservation in the standard model. Unfortunately, both this process and CP violation turn out to be quantitatively too weak to explain with our current understanding of the early evolution of the universe the quantitative level of the asymmetry between matter and anti-matter. Thus, the current belief is that so far undiscovered physics is responsible for the missing (large) amount.

It is nonetheless instructive to understand how baryon number violation comes about in the standard model. Lepton number violation proceeds in the same way, but is even more suppressed, due to the much smaller masses.

The basic ingredient is once more a classical field configuration, this time of Yang-Mills theory. Define the field strength tensors $F_{\mu\nu} = \tau^a F^a_{\mu\nu}$, with τ^a the Pauli matrices - here the weak interactions with the gauge group SU(2) is the interesting one. To proceed further, it is useful to make the formal replacement $it \to t$, which can be undone at the end. This is an analytic continuation from Minkowski space-time to Euclidean space-time, as now all components of the metric have the same sign.

The Jacobi identity (5.13) together with the non-Abelian version of the homogeneous Maxwell equation of QED,

$$\partial_\rho F_{\mu\nu} + \partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} = 0,$$

yield the Bianchi identity

$$\begin{aligned} D_\rho F_{\mu\nu} + D_\mu F_{\nu\rho} + D_\nu F_{\rho\mu} &= 0, \\ D_\mu F_{\nu\rho} &= D^{ij}_\mu \tau^a_{jk} F^a_{\nu\rho}. \end{aligned}$$

Together, they imply

$$\begin{aligned} 0 &= \frac{1}{2}\epsilon_\sigma{}^{\rho\mu\nu} D_\rho F_{\mu\nu} = D_\rho \tilde{F}_{\sigma\rho} \\ \tilde{F}_{\sigma\rho} &= \frac{1}{2}\epsilon_{\sigma\rho}{}^{\mu\nu} F_{\mu\nu}, \end{aligned}$$

where $\tilde{F}$ is called the dual field-strength tensor. The correctness of these statements follow from the anti-symmetry of the field-strength tensor and the Jacobi identity (5.13). However, the (inhomogeneous) Maxwell equation

$$D_\mu F^{\mu\nu} = 0, \tag{6.23}$$

are therefore trivially solved by (anti-)self-dual solutions

$$F_{\mu\nu} = \pm \tilde{F}_{\mu\nu}, \tag{6.24}$$

as these convert the equation (6.23) into the trivial Bianchi identity. The self-duality equations (6.24) have the advantage of being only first-order differential equations instead of second-order differential equations, and are therefore easier to solve.

Furthermore, classical solutions have to have a finite amount of energy, and therefore their behavior at large distances is constrained. Especially, since the Lagrangian can be written as the sum of the squares of the electric and magnetic field strength, both these fields must vanish. This can only occur if the potential becomes at large distances gauge-equivalent to the vacuum, i. e. it has the form

$$A_\mu^a \tau^a = A_\mu = ig(x)\partial_\mu g^{-1}(x),$$

where $g = g^a \tau_a$ is an arbitrary function. Since this is the second part of the gauge transformation (5.17) in matrix notation, this is just a version of the vacuum $A_\mu = 0$. Since all choices are gauge-equivalent, any choice will do. One possibility which turns out to be technically convenient is

$$\begin{aligned} g(x) &= \frac{x^\mu \tau_\mu}{|x|} \\ \tau_\mu &= (1, i\tau^a). \end{aligned} \tag{6.25}$$

The simplest extension of this is a multiplication with a function $f(x^2)$ which becomes 1 at large distances,

$$\begin{aligned} A_\mu^a \tau^a &= if(x^2)g(x)\partial_\mu g^{-1}(x) = 2f(x^2)\tau_{\mu\nu}\frac{x^\nu}{x^2} \\ \tau_{\mu\nu} &= \frac{1}{4i}(\tau_\mu \bar{\tau}_\nu - \tau_\nu \bar{\tau}_\mu) \\ \bar{\tau}_\mu &= (1, -i\tau^a). \end{aligned}$$

The matrices $\tau_{\mu\nu}$ are called 't Hooft symbols. Thus, this ansatz mixes non-trivially the weak isospin and space-time.

Plugging this in into the self-duality equation (6.24) yields a first-order differential equation for $f(x^2)$,

$$x^2 \frac{df}{dx} - f(1-f)) = 0.$$

The solution to this equation, which can be obtained by separation of variables, is

$$f(x^2) = \frac{x^2}{x^2 + \lambda^2},$$

where λ is an integration constant. The function indeed goes to one at large distances, as required. The structure described by this field configuration is now localized in space-time at the origin, and extended over a range of size λ. Such a localized event in space and time is called an instanton. Solving with the equation with the other sign in the self-duality equation (6.24) yields a similar result, though with some small differences, and is called an anti-instanton.

Going back to Minkowski space-time, the field configuration will have a singularity at $x^2 = -\lambda^2$. This is called a sphaleron, and hence a violent event in space-time. Note that the gauge-field strength does not explicitly appear in the calculation. This result can therefore not be obtained perturbatively, and the presence of instantons is a non-perturbative effect.

Seeing now that these indeed create baryon number violation is unfortunately technically very complicated, so here only the most important steps will be sketched. One highly non-trivial, but very fundamental, insight needed is that instantons turn out to be very much connected with the anomalies of section 6.11. In fact, any instanton induces via the global anomaly interactions between fermions. Especially, it can connect fermions of different types, as long as all are charged and all are affected by the anomaly. Especially, this yields that instantons effectively create an interaction which involves, besides the gauge fields also three quarks and one lepton, and thus permits baryon number violation. However, this integration still conserves fermion number, and as a consequence the change in the baryon number must be offset by the same change in lepton number. Still, this implies that reducing the baryon number by one can be offset by a change in lepton number by one, which implies that a proton can be converted, with the involvement of gauge fields, into a positron. Hence, baryon (and lepton number) are not conserved in the standard model.

The effect is, however, small, and suppressed exponentially by $\exp(-c/\alpha_W)$, where c is a number, and α_W is the weak isospin fine-structure constant. Since the latter is small, the suppression is huge, and the life-time of the proton in the standard model exceeds the current upper experimental upper limit for the decay in any channel of 2.1×10^{28} years by many orders of magnitude. Hence, the baryon number violation in the standard model is not able to explain the fact that so much more baryons than anti-baryons exist. However, this is a statement about the current state of the universe, and especially its temperature, and it may change in earlier times. Especially, it can be shown that the effect becomes exponentially enhanced with temperature. Still, the temepratures necessary to make this a sucfficiently effective process have been available for too short times in the early universe.

6.14 The early universe

The weak interactions, and particularily the Higgs effect, play also another important role in the early universe. Just as the magnetization in a magnet can be removed by heating it up, so can the Higgs condensate melt at high temperatures, making the symmetry manifest once more. This process is different in nature from the effectively manifest symmetry at large energies, which is a quantitative effect as all masses become negligible. However, there is not necessarily a phase transition associated with the melting of the condensate. In fact, it is in general not a symmetry restoration, as in case of a global symmetry to be discussed here as well. This is due to the fact that the symmetry is just hidden, not broken. As such, there is no local gauge-invariant order parameter associated with it. Only gauge-dependent order parameters can be local, but in their case the temperature where the symmetry becomes manifest once more is in general gauge-dependent. Only a transition which would be indicated by a non-local gauge-invariant order parameter could in principle mark a true phase transition.

Again, it is quite useful to first study the case of a global symmetry, then of an Abelian local symmetry before going to the electroweak theory.

6.14.1 Global symmetry

A useful starting point is given by the Lagrangian (5.8)

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}\partial_\mu\eta\partial^\mu\eta - \frac{1}{2}(6\lambda f^2-\mu^2)\eta^2 + \frac{1}{2}\partial_\mu\chi\partial^\mu\chi - \frac{1}{2}(2\lambda f^2-\mu^2)\chi^2 \\ &\quad -\sqrt{2}\lambda f\eta(\eta^2+\chi^2) - \frac{1}{4}\lambda(\eta^2+\chi^2)^2 - \mu^2 f^2 + \lambda f^4. \end{aligned} \tag{6.26}$$

In this case the explicit zero-energy contribution is kept for reasons that will become apparent shortly, but will be essentially the same as when treating non-relativistic Bose-Einstein condensation. Only terms linear in the fields have been dropped, as they will not contribute in the following. The situation is similar as before, but now the condensate f has not been specified by the minimization of the classical potential, but is kept as a free quantity, which will take its value dynamically.

To investigate the thermodynamic behavior it is useful to analyze the thermodynamic potential Ω in analogy to the non-relativistic case as

$$\Omega(T,f) = -P(T,f) = -T\ln\frac{Z}{V},$$

where P is the pressure, T the temperature, and V the volume. Z is the partition sum. For the following purposes, it is sufficient to use the so-called mean-field approximation. In this case, the interaction terms are first expanded around f, and higher-order terms which involve more than two fields are neglected. In the Lagrangian (6.26), this has already been done.

Without going into the technical details, the thermodynamic potential can be evaluated directly. It reads

$$\begin{aligned}\Omega(T,f) &= -\mu^2 f^2+\lambda f^4 \\ &+\int\frac{d^3p}{(2\pi)^3}\left(\frac{\omega_1^2+\omega_2^2}{2}+T\left(\ln\left(1-e^{-\frac{\omega_1}{T}}\right)+\ln\left(1-e^{-\frac{\omega_2}{T}}\right)\right)\right) \\ \omega_1 &= \sqrt{6\lambda f^2-\mu^2+p^2}=\sqrt{m_\eta^2+p^2} \\ \omega_2 &= \sqrt{2\lambda f^2-\mu^2+p^2}=\sqrt{m_\chi^2+p^2}.\end{aligned} \tag{6.27}$$

The frequencies ω consists of the momenta and the masses of the particles after hiding the symmetry, which is dependent on the value of the condensate f. There are three contributions. The first outside the integral is the classical contribution. The second are the first two terms inside the integral. They are the contributions from quantum fluctuations. The third term represents thermal fluctuations.

To recover the results from section 5.7, the second term must be neglected and the zero-temperature limit taken. This yields

$$\Omega(0,f)=-\mu^2f^2+\lambda f^4.$$

As in section 5.7, this potential has a minimum at non-zero f, $f=\mu^2/(2\lambda)$. Inserting this into the Lagrangian (6.26) makes it equivalent to (5.8), with a massive Higgs boson and a massless Goldstone boson.

Something new happens at finite temperature. At small temperature, little other happens than that it is possible to excite Higgs bosons or Goldstone bosons, which then form a thermal bath of non-interacting bosons, and the total pressure is just the sum of their respective pressures. However, the value of f will become temperature-dependent: At each temperature it will take the value which minimizes the thermodynamic potential.

When going to higher temperatures, it is useful to make a high-temperature expansion for the thermodynamic potential. High temperature requires here T to be larger than the scale of the zero-temperature case, which is given by the condensate, which is of order $\mu/\sqrt{\lambda}$. In this case, it is possible to obtain an expansion for Ω. The leading terms up to order $\mathcal{O}(1)$ are given by

$$\Omega(T,f)=\lambda f^4+\left(\frac{1}{3}\lambda T^2-\mu^2\right)f^2-\frac{\pi^2}{45}T^4-\frac{\mu^2T^2}{12}. \tag{6.28}$$

Note that at very large temperatures only the term $\pi^2T^4/45$ is relevant, which is precisely the one of a free non-interacting gas of two boson species, a Stefan-Boltzmann-like behavior.

This results exhibits one interesting feature. The term of order f^2 has a temperature-dependent coefficient, which changes sign at[13] $T_c^2 = 3\mu^2/\lambda$. As a consequence, the shape of the thermodynamic potential as a function of f changes. Below T_c, it has a minimum away from zero, as at zero temperature. With increasing temperature, this minimum moves to smaller and smaller values, and arrives at zero at T_c. Hence, at T_c, the value of f changes from a non-zero to a zero value, and the symmetry becomes manifest once more. Above T_c, the minimum stays at zero, and for all higher temperatures the symmetry is manifest.

Replacing f with its temperature-dependent value in (6.28) yields the expressions

$$\begin{aligned}\Omega_{T<T_c} &= \tfrac{\mu^2T^2}{12} - \left(\tfrac{\pi^2}{45} + \tfrac{\lambda}{36}\right)T^4 \overset{T=T_c}{=} -\frac{\pi^2\mu^2}{5\lambda^2}\\ \Omega_{T>T_c} &= \tfrac{\mu^4}{4\lambda} - \tfrac{\pi^2}{45}T^4 - \tfrac{\mu^2T^2}{12} \overset{T=T_c}{=} -\frac{\pi^2\mu^2}{5\lambda^2},\end{aligned}$$

which coincide at T_c. Also their first derivatives with respect to the temperature are equal at T_c

$$\begin{aligned}\frac{d\Omega_{T<T_c}}{dT} &= -(8\pi^2T^2 + 10\lambda T^2 - 15\mu^2)\tfrac{T}{90} \overset{T=T_c}{=} -\frac{8\pi^2+5\lambda}{\sqrt{300}}\sqrt{\frac{\mu^2}{\lambda}}\\ \frac{d\Omega_{T>T_c}}{dT} &= -(8\pi^2T^2 + 15\mu^2)\tfrac{T}{90} \overset{T=T_c}{=} -\frac{8\pi^2+5\lambda}{\sqrt{300}}\sqrt{\frac{\mu^2}{\lambda}},\end{aligned}$$

but their second derivatives are not

$$\begin{aligned}\frac{d^2\Omega_{T<T_c}}{dT^2} &= \tfrac{\mu^2}{6} - (4\pi^2 + 5\lambda)\tfrac{T^2}{15} \overset{T=T_c}{=} -\frac{(25\lambda + 24\pi^2)\,\mu^2}{30\lambda}\\ \frac{d^2\Omega_{T>T_c}}{dT^2} &= -\tfrac{8\pi^2T^2+5\mu^2}{30} \overset{T=T_c}{=} -\frac{(5\lambda + 24\pi^2)\,\mu^2}{30\lambda}.\end{aligned}$$

Thus, a phase transition of second order occurs at T_c.

As stressed previously repeatedly, it is possible that quantum effects could modify the pattern considerably or even melt the condensate. It is therefore instructive to investigate the leading quantum corrections to the previous discussion.

This is also necessary for another reason. If the symmetry becomes manifest once more at large temperatures, the mass of Higgs-like excitations become tachyonic, indicating a flaw of the theory. That can be seen directly by reading off the condensate-dependent masses of the excitations,

$$\begin{aligned}m_\eta^2 &= 6\lambda f^2 - \mu^2 = -\mu^2\theta(T-T_c) + (2\mu^2 - \lambda T^2)\theta(T_c - T)\\ m_\chi^2 &= 2\lambda f^2 - \mu^2 = -\mu^2\theta(T-T_c) - \frac{\lambda T^2}{3}\theta(T_c - T).\end{aligned}$$

[13] Note that strictly speaking using the high-temperature expansion at this temperature is doubtful. For the purpose here it will be kept since it makes the mechanisms more evident then the rather technical calculations necessary beyond the high-temperature expansion. The qualitative outcome, however, is not altered, at least within the first few orders of perturbation theory.

Furthermore, also the Goldstone theorem is violated, as the mass of the Goldstone boson χ is no longer zero[14]. Both problems are fixed by quantum corrections, demonstrating the importance of quantum fluctuations even in the high-temperature phase.

In the expression for the free energy (6.27) the zero-point energy, and thus the quantum fluctuations have been neglected. Including them yields the result

$$\begin{aligned}\Omega(T,f) &= -\frac{\pi^2}{45}T^4-\frac{\mu^2T^2}{12}-\frac{(m_\eta^3+m_\chi^3)T}{12}+\frac{\mu^4}{32\pi^2}\ln\frac{8\pi^2T^2e^{-2\gamma+\frac{3}{2}}}{\mu^2}\\ &\quad -\mu^2f^2\left(1+\frac{\delta\xi}{\lambda}+\frac{\lambda}{4\pi^2}\ln\frac{8\pi^2T^2e^{-2\gamma+1}}{\mu^2}-\frac{\lambda T^2}{3\mu^2}\right)\\ &\quad +\lambda f^4\left(1+\frac{\delta\xi}{\lambda}+\frac{5\lambda}{8\pi^2}\ln\frac{8\pi^2T^2e^{-2\gamma+1}}{\mu^2}\right).\end{aligned}$$

Herein is γ the Euler number. The term $\delta\xi$ is a quantum correction, which can be fixed, e. g. by measuring T_c or by other experimental input, as can be seen from the new expression for T_c where f vanishes,

$$T_c^2=\frac{3\mu^2}{\lambda}\left(1+\frac{\delta\xi}{\lambda}+\frac{\lambda}{4\pi^2}\ln\frac{24\pi^2e^{-2\gamma+1}}{\lambda}\right).$$

Here, only the qualitative result is interesting, and its precise numerical value of no importance.

To obtain the corrections for the masses, it is necessary to calculate the corresponding quantum corrections. Without going into the details, the complete masses to this order are

$$\begin{aligned}m_\eta^2 &= 2\mu^2\left(1-\frac{\lambda T^2}{3\mu^2}\right)\theta(T_c-T)+\frac{1}{3}\lambda\left(T^2-\frac{3\mu^2}{\lambda}\right)\theta(T-T_c)\\ m_\chi^2 &= \frac{\lambda}{3}\left(T^2-\frac{3\mu^2}{\lambda}\right)\theta(T-T_c).\end{aligned}$$

These results yield a number of interesting observation. First, since T_c is larger[15] than $3\mu^2/\lambda$, the mass of the Higgs is always positive, stabilizing the system. Secondly, in this case the mass of the Goldstone boson is always zero below the phase transition temperature, in agreement with the Goldstone theorem. Above the phase transition, the masses of both particle degenerate, and the symmetry is manifest once more also in the spectrum.

[14] In a full quantum treatment, the role of the Goldstone boson could be played at finite temperature by some composite excitation instead. However, at the mean-field level no such excitations are available, and thus the Goldstone theorem is violated.

[15] It is not obvious that $\delta\xi$ cannot be negative and large, thus making the improved estimate for T_c smaller than before. However, it turns out not to be the case at this order for any renormalization prescription.

These properties are generic for symmetries hiding by a condensate which thaws with increasing temperature. Also that the mean-field approximation is in general insufficient is a lesson which should be kept duly in mind. Of course, at the present time much more sophisticated methods are available to treat this problem, though they are in general very complicated.

6.14.2 Abelian case

As was visible in the previous case of the global symmetry, quantum fluctuations are important to obtain a consistent result. Again going through the mean-field calculations (introducing only a mean-field for the Higgs fields η condensate, but neither for the pseudo-Goldstone boson χ nor for the photon), it is possible to obtain once more a high-temperature expansion for the free energy. The full expression at mean-field level then reads

$$\Omega(f,T) = \lambda f^4 + \left(\left(\frac{\lambda}{3} + \frac{e^2}{4}\right)T^2 - \mu^2\right)f^2 - \frac{\mu^2 T^2}{12} - \frac{2\pi^2}{45}T^4,$$

where e is now the electromagnetic coupling constant. First, the Stefan-Boltzmann contribution proportional to T^4 is now twice as large, as the photon also contributes two scalar degrees of freedom. Furthermore, the leading high-temperature behavior is not altered otherwise by the presence of the interaction. At the level of the mean-field approximation there is still only a non-interacting gas of two scalars and two photon polarizations left. However, the term crucial for the phase transition, the one of order T^2, is affected by the interactions. Thus, the interactions have an influence on the phase transition. At mean-field level they just shift the phase transition temperature.

Beyond mean-field, their impact is more relevant. At one-loop level, the phase structure becomes dependent on the relative size of λ and e.

As long as λ is larger[16] than e, the situation is found to be in agreement as when the photon field would be absent. In particular, the condensate melts at a (now also e-dependent) critical temperature, and above the phase transition both the scalars and the photon become massless again.

If λ is smaller than e, the critical temperature becomes lower than the effective mass of the photon. As a consequence, the second-order phase transition may in fact become first order. Even more drastic, if λ is below $3e^4/(32\pi^2)$, the phase transition temperature decreases to zero, and spontaneous breaking is not occurring at all. In that case, the stronger photon-scalar interactions stabilize the vacuum, and no condensate can form.

[16] If e and λ are approximately equal, even higher order corrections can have a qualitative impact.

An important insight is found in the case of $\mu = 0$. At first sight no condensation is possible. However, the interactions mediated by the photon may still be sufficiently strong and attractive enough that the scalars condense, yielding the same physics as if μ would not be zero, and λ would be sufficiently large. In this case, the condensation is a genuine quantum effect, as only due to (one-)loop quantum corrections spontaneous condensation of the Higgs occurs.

6.14.3 The electroweak case

Also in the electroweak case it is useful to use the Lagrangian in the 't Hooft gauge, given by (6.20). Though straight-forward, the calculation in the present case is much more cumbersome than in the previous cases. For the present purpose, the masses of the fermions (and thus the CKM matrices) can be neglected. The imprecision of this is comparable to the effects of going to the next order. The final result at mean-field level is

$$\Omega(f,T) = -\frac{427}{360}\pi^2 T^4 - \frac{f^2}{2}\left(\mu^2 - \frac{T^2}{4}\left(2\lambda + \frac{3g^2}{4} + \frac{g'^2}{4}\right)\right) + \frac{\lambda f^4}{4} + \frac{\mu^4}{4\lambda} - \frac{\mu^2 T^2}{6}.$$

Extremalizing this expression with respect to f yields the critical temperature for the electroweak standard model in this approximation as

$$T_c^2 = \frac{4\mu^2}{2\lambda + \frac{3g^2}{4} + \frac{g'^2}{4}}.$$

The temperature dependence of the condensate f and the pressure then read, respectively,

$$\begin{aligned} f^2 &= \frac{\mu^2}{\lambda}\left(1 - \frac{T^2}{T_c^2}\right)\theta(T_c - T) \\ P &= \begin{cases} \frac{427\pi^2 T^4}{360} + \frac{\mu^2}{4\lambda}\left(1 - \frac{T^2}{T_c^2}\right)^2 + \frac{\mu^2 f^2 T^2}{6} - \frac{\mu^4}{4\lambda}, & T \leq T_c \\ \frac{427\pi^2 T^4}{360} + \frac{\mu^2 T^2}{6} - \frac{\mu^4}{4\lambda}, & T \geq T_c \end{cases}. \end{aligned}$$

The corresponding phase transition is at mean-field level thus of second order, as the second derivative of the pressure exhibits a discontinuity. For a Higgs mass of 100 GeV the critical temperature is about 200 GeV. For the actual value of the Higgs mass of 126 GeV, it is only slightly higher. Thus the transition temperature is of the same order as the Higgs condensate, and about three orders of magnitude larger than the corresponding temperature in QCD.

Note that all problems with consistency of the mean-field approach pertain also to the full electroweak standard model. Therefore, for a consistent treatment at least leading order corrections have to be included. As previously, they do not change the phase transition temperature, but may change its order.

The relevance of such a temperature is only given in the early universe, or perhaps during a collapse to a black hole. Here, the more certain case of the early universe will be treated.

To assess the relevance for the early universe it is necessary to add equations which describe its development. For the present purpose simplified versions of the Einstein equations are sufficient. Adding energy conservation gives

$$\left(\frac{dR}{dt}\right)^2 = \frac{8\pi G}{3}\epsilon R^2 \quad (6.29)$$

$$\frac{d(\epsilon R^3)}{dR} = -3PR^2, \quad (6.30)$$

where G is Newton's constant, ϵ is the energy density, R is the scale factor, essentially given by the Ricci curvature scalar, t the proper time, and P is the pressure. To close the equations, an equation of state is necessary, which is given as a function of the pressure by thermodynamic relations as

$$\epsilon = -P + T\frac{\partial P}{\partial T}.$$

Since the sicknesses of the mean-field approximation are not too problematic for this estimate, it is sufficient to use it for obtaining the corresponding energy density as

$$\epsilon = \begin{cases} \left(\frac{1281\pi^2}{360} + \frac{\mu^4}{4\lambda T_c^4}\right)T^4 + \left(1 - \frac{3\mu^2}{\lambda T_c^2}\right)\frac{\mu^2 T^2}{6}, & T \leq T_c \\ \frac{1281\pi^2}{360}T^4 + \frac{\mu^2 T^2}{6} + \frac{\mu^4}{4\lambda}, & T \geq T_c \end{cases}.$$

Rewriting equation (6.30) in terms of the temperature yields the ordinary differential equation

$$R\frac{d\epsilon}{dT} + 3\frac{dR}{dT} = -3P.$$

Imposing as a boundary condition that R should be one at the phase transition yields

$$R^3 = \begin{cases} \frac{T_c}{T}\frac{T_c^2 - b^2}{T^2 - b^2}, & T \leq T_c \\ \frac{T_c}{T}\frac{T_c^2 + a^2}{T^2 + a^2}, & T \geq T_c \end{cases}$$

$$a^2 = \frac{30\mu^2}{427\pi^2}$$

$$b^2 = \frac{a^2 T_c^2(1-r)}{a^2 + rT_c^2}$$

$$r = \frac{4\lambda}{6\lambda + \frac{9g^2}{4} + \frac{3g'^2}{4}}.$$

For a Higgs mass about 100 GeV the characteristic parameters are $r = 0.22$, $a = 6$ GeV and $b = 10$ GeV. Thus, the dominant behavior is that R behaves like T_c^3/T^3, up to some small modifications close to the phase transition, and thus drops essentially in the electroweak domain.

An interesting consequence is obtained if R^3 is multiplied by the entropy

$$s = \frac{\partial P}{\partial T} \sim T^3.$$

An elementary calculations yields thus that sR^3 is constant. Since s is in units inverse length cubed, this is just the statement that entropy is conserved since R only describes the expansion of a unit length over time. Hence, the electroweak interactions at mean-field level conserve entropy. In particular, this is a consequence of the second order phase transition, which is not permitting latent heat or supercooling. Finally, inserting the numbers shows that R increases somewhat slower around the phase transition. Thus, the expansion of the universe slows down during the electroweak phase transition.

Of course, all of this is just an estimate. That it can never be fully correct is seen by the fact that R diverges at the finite temperature $T = b$, much above the QCD phase transition (and nowadays) temperature. This is an artifact of the high-temperature expansion involved. To obtain the correct behavior down to the QCD phase transition would require more detailed calculations.

Chapter 7

Beyond the standard model

At the end of 2009 the largest particle physics experiment so far has been started, the LHC at CERN. With proton-proton collisions at a center-of-mass energy of up to 14 TeV, possibly upgraded to 33 TeV by the end of the 2020ies, there are two major objectives. One was to complete the current picture of the standard model of particle physics by finding the Higgs boson. This has been accomplished, even though its properties remain still to be established with satisfactory precision, to make sure that it is indeed the Higgs boson predicted by the standard model.

The second objective is to search for new physics beyond the standard model. For various reasons it is believed that there will be new phenomena appearing in particle physics at a scale of 1 TeV, and thus within reach of the LHC. Though this is not guaranteed, there is motivation for it.

In fact, there are a number of reasons to believe that there exists physics beyond the standard model. These reasons can be categorized as being from within the standard model, by the existence of gravity, and by observations which do not fit into the standard model. Essentially all of the latter category are from astronomy, and there are currently essentially no observations in terrestrial experiments which are reproducible and do not satisfactorily agree with the standard model, and none which disagree with any reasonable statistical accuracy.

Of course, it should always be kept in mind that the standard model has never been completely solved. Though what has been solved, in particular using perturbation theory, agrees excellently with measurements, it is a highly non-linear theory. It cannot a-priori be excluded that some of the reasons to be listed here are actually completely within the standard model, once it is possible to solve it exactly.

Many of the observations to be listed can be explained easily, but not necessarily, by new physics at a scale of 1 TeV. However, it cannot be exclude that there is no new phenomena necessary for any of them up to a scale of 10^{15} GeV, or possibly up to the Planck scale of 10^{19} GeV, in which case the energy domain between the standard model and this scale is known as the great dessert.

7.1 Inconsistencies of the standard model

There are a number of factual and perceived flaws of the standard model, which make it likely that it cannot be the ultimate theory.

The one most striking reason is the need of the renormalization, as described in section 3.7: It is not possible to determine within the standard model processes at arbitrary high energies. The corresponding calculations break down eventually, and yield infinities. Though we have learned how to absorb this lack of knowledge in a few parameters, the renormalization constants, it is clear that there are things the theory cannot describe. Thus it seems plausible that at some energy scale these infinities are resolved by new processes, which are unknown so far. In this sense, the standard model is often referred to as an low-energy effective theory of the corresponding high-energy theory.

This theory can in fact also not be a (conventional) quantum field theory, as this flaw is a characteristic of such theories. Though theories exist which reduce the severity of the problem, supersymmetry at the forefront of them, it appears that it is not possible to completely resolve it, though this cannot be excluded. Thus, it is commonly believed that the high-energy theory is structurally different from the standard model, like a string theory. This is also the case when it comes to gravity, as discussed in the next section.

There are a number of aesthetic flaws of the standard model as well. First, there are about thirty different free parameters of the theory, varying by at least ten orders of magnitude and some of them are absurdly fine-tuned without any internal necessity. There is no possibility to understand their size or nature within the standard model, and this is unsatisfactory. Even if their origin alone could be understood, their relative size is a mystery as well. This is particularly true in case of the Higgs and the electroweak sector in general. There is no reason for the Higgs to have a mass which is small compared to the scale of the theory from which the standard model emerges. In particular, no symmetry protects the Higgs mass from large radiative corrections due to the underlying theory, which could make it much more massive, and therefore inconsistent with experimental data, than all the other standard model particles. Why this is not so is called the hierarchy problem, despite the fact that it could just be accidentally so, and not a flaw of the theory. Even if this scale should be of the order of a few tens of TeV, there is still a factor of possibly 100 involved, which is not as dramatic as if the scale would be, say, 10^{15} GeV. Therefore, this case is also called the little hierarchy problem.

There is another strikingly odd thing with these parameters. The charges of the leptons and quarks need not be the same just because of QED - in contrast to the weak or strong charge, actually. They could differ, and in particular do not need to have the ratio of small integer numbers as they do - 1 to 2/3 or 1/3. This is due to the fact that the gauge group of QED is Abelian. However, if they would not match within the experimental precision of more than ten orders of magnitude, and also the number of charged particles under the strong and weak force were not perfectly balanced in each generation, then actually the standard model would not work, and neither would physics with neutral atoms. This is due to the development of a quantum anomaly, i. e., an effect solely related to quantizing the theory which would make it inconsistent, see section 6.11. Only with the generation structure of quarks and leptons with the assigned charges of the standard model, this can be avoided. This is ultimately mysterious, and no explanation exists for this in the standard model, except that it is necessary for it to work, which is unsatisfactory.

There is also absolutely no reason inside the standard model, why there should be more than one family, since the aforementioned cancellation works within each family independently and is complete. However, at least three families are necessary to have inside the standard model CP violating processes, i. e. processes which favor matter over anti-matter, see section 6.8. As discussed later, such processes are necessary for the observation that the world around us is made from matter. But there is no reason, why there should only be three families, and not four, five, or more. And if there should be a fourth family around, why would its neutrino be so heavy, more than 100 GeV, compared to the other ones, as can already be inferred from existing experimental data.

Another particle, however, could still be beyond the standard model: The Higgs boson. As the properties of the observed candidate are not yet beyond doubt, there is no certainty, however expected and likely, that it really is the Higgs. If it is not, the whole conceptual structure of the electroweak sector of the standard model immediately collapses, and is necessarily replaced by something else. Therefore, despite the fact that a relatively light Higgs boson is in perfect agreement with all available data, this is not necessarily the case in nature.

As an aside, there is also a very fundamental question concerning the Higgs sector. At the current time, it is not yet clear whether there can exist, even in the limited sense of a renormalizable quantum field theory, a meaningful theory of an interacting scalar field. This is the so-called triviality problem. So far, it is only essentially clear that the only consistent four-dimensional theory describing only a spin zero boson is one without any interactions. Whether this can be changed by adding additional fields, as in the standard model, is an open question. However, since this problem can be postponed to energy scales as high as 10^{15} GeV, or possibly even higher, in particular for a Higgs as light as the observed one, this question is not necessarily of practical relevance.

Finally, when extrapolating the running gauge couplings for a not-to-massive Higgs to an energy scale of about 10^{15} GeV, their values almost meet, suggesting that at this scale a kind of unification would be possible. However, they do not meet exactly, and this is somewhat puzzling as well. Why should this seem to appear almost, but not perfectly so?

7.2 Gravity

7.2.1 Problems with quantization

One obviously, and suspiciously, lacking element of the standard model is gravity. Up to now, no fully consistent quantization of gravity has been obtained. Usually the problem is that a canonical quantized theory of gravity is not perturbatively renormalizable. This is visible when writing down the Einstein-Hilbert-Lagrangian of gravity, based on the metric $g_{\mu\nu}$, $\mathcal{L}_{\mathrm{EH}}$,

$$\begin{aligned} \mathcal{L}_{\mathrm{EH}} &= \omega\left(\frac{1}{2\kappa}R - \frac{1}{\kappa}\Lambda + \mathcal{L}_M\right) \qquad (7.1)\\ \omega &= \sqrt{-g} \\ R^{\lambda}_{\rho\mu\nu} &= \partial_\mu\Gamma^{\lambda}_{\nu\rho} - \partial_\nu\Gamma^{\lambda}_{\nu\rho} + \Gamma^{\lambda}_{\mu\sigma}\Gamma^{\sigma}_{\nu\rho} - \Gamma^{\lambda}_{\nu\rho}\Gamma^{\sigma}_{\mu\rho} \\ \Gamma^{\lambda}_{\mu\nu} &= \frac{1}{2}g^{\lambda\sigma}(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}) \end{aligned}$$

From this, it can be inferred that the coupling constant involved, κ or equivalently Newton's constant, is dimensionful, as is the cosmological constant Λ. Any matter is described by the matter Lagrangian $\mathcal{L}_M$. This also yields the Euler-Lagrange equations of gravity, the Einstein equations

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + g_{\mu\nu}\Lambda = -\kappa T_{\mu\nu} \qquad (7.2)$$

$$T_{\mu\nu} = -\eta_{\mu\nu}\mathcal{L}_M + 2\frac{\delta\mathcal{L}_M}{\delta g^{\mu\nu}}(g_{\mu\nu} = \eta_{\mu\nu}) \qquad (7.3)$$

Superficial (perturbative) calculations for (7.1) show that the theory is perturbatively non-renormalizable. As a consequence, an infinite hierarchy of independent parameters, all to be fixed by experiment, would be necessary to perform perturbative calculations, spoiling any predictivity of the theory. In pure gravity, these problems occur at NNLO, for matter coupled to gravity already at the leading order of radiative corrections. In particular, this implies that the theory is perturbatively not reliable beyond the scale $\sqrt{\kappa}$. Though this may be an artifact of perturbation theory, this has led to developments like supergravity based on local supersymmetry or loop quantum gravity.

Irrespective of the details, the lack of gravity is an obvious flaw of the standard model. Along with this lack comes also a riddle. The natural quantum scale of gravity is given by the Planck scale

$$M_P = \frac{1}{\sqrt{G_N}} \approx 1.22 \times 10^{19}\ \mathrm{GeV}.$$

This is 17 orders of magnitude larger than the natural scale of the electroweak interactions, and 19 orders of magnitude larger than the one of QCD. The origin of this mismatch is yet unsolved.

One of the more popular proposals how to explain this, discussed in section **??**, is that this is only an apparent mismatch: The scales of gravity and the standard model are the same, but gravity is able to propagate also in additional dimensions not accessible by the standard model. The mismatch comes from the ratio of the total volumes, the bulk, and the apparent four dimensional volume, which is thus only a boundary, a so-called brane.

7.2.2 Asymptotic safety

Reiterating, the problem with the renormalizability of quantum gravity is a purely perturbative statement, since only perturbative arguments have been used to establish it. Thus, the possibility remains that the theory is not having such a problem, it is said to be asymptotically safe, and the problem is a mere artifact of perturbation theory. In this case, when performing a proper, non-perturbative calculation, no such problems would be involved. In fact, this includes the possibility that κ imposes just an intrinsic cutoff of physics, and that this is simply the highest attainable energy, similarly as the speed of light is the maximum velocity. As a consequence, the divergences encountered in particle physics then only results from taking the improper limit $\kappa \to \infty$.

This or a similar concept of asymptotic safety can be illustrated by the use of a running coupling, this time the one of quantum gravity. The naive perturbative picture implies that the running gravitational coupling increases without bounds if the energy is increased, similarly to the case of QCD if the energy is decreased. Since the theory is non-linearly coupled, an increasing coupling will back-couple to itself, and therefore may limit its own growth, leading to a saturation at large energies, and thus becomes finite. This makes the theory then perfectly stable and well-behaved. However, such a non-linear back-coupling cannot be captured naturally by perturbation theory, which is a small-field expansion, and thus linear in nature. It thus fails in the same way as it fails at small energies for QCD.

Non-perturbative methods, like renormalization-group methods or numerical simulations, have provided indications that indeed such a thing may happen in quantum gravity, though this requires confirmation.

As an aside, it has also been proposed that a similar solution may resolve both the hierarchy problem and the triviality problem of the Higgs sector of the standard model, when applied to the combination of Higgs self-coupling and the Yukawa couplings, and possibly the gauge couplings.

7.3 Observations from particle physics experiments

There are two generic types of particle physics experiments to search for physics beyond the standard model, both based on the direct interaction of elementary particles. One are those at very high energies, where the sheer violence of the interactions are expected to produce new particles, which can then be measured. The others are very precise low-energy measurements, where very small deviations from the standard model are attempted to be detected. Neither of these methods has provided so far any statistically and systematically robust observation of a deviation from the standard model. Indeed, it happened quite often that a promising effect vanishes when the statistical accuracy is increased. Also, it has happened that certain effects have only been observed in some, but not all of conceptually similar experiments. In these cases, it can again be a statistical effect, or there is always the possibilities that some, at first glance, minor difference between the experiments can fake such an effect at one experiment, or can shadow it at another. So far, the experience was mostly that in such situation a signal was faked, but this then usually involves a very tedious and long search for the cause. Also, the calculation of the theoretically expected values is often involved, and it has happened that by improving the calculations, a previously observed mismatch is actually just an underestimation of the theoretical error.

At the time of writing, while new results are coming in frequently, there are very few remarkable results which should be mentioned, and which await further scrutiny. On the other hand, several hundreds of measurements have not yet indicated any deviations from known physics. One should then be especially wary: If enough measurements are performed, there is always a statistical probability that some of them show deviations, the so-called look-elsewhere effect, and a number of such deviations are expected.

One observation from precision experiments is that the magnetic moment of the muon is not quite the expected one, but shows a very small deviation. However, the theoretical computations are difficult, as virtual corrections from QCD enter, and it is therefore not clear whether this problem is just a theoretical one. Nonetheless, new experiments will be performed to at least obtain a better experimental value, while the theoretical calculations are continously improved.

The next is that two experiments have in agreement reported results that cannot be explained with just the three ordinary neutrinos. However, these results are rather indirect, and at least another independent experiment will be required for confirmation.

All of these observations are currently investigated further. Given the amount of statistics needed, it may take quite some time for a final answer about the reality of any of these deviations.

7.4 Astronomical observations

During the recent decades a number of cosmological observations have been made, which cannot be reconciled with the standard model. These will be discussed here.

7.4.1 Direct observations

Some come from space-borne experiments, like the FERMI-Lat satellite and the AMS-2 spectrometer aboard the ISS. Both show an excess of anti-positrons over the expected yield in cosmic rays. If true, both would hint at an indirect detection of some new matter particles, perhaps dark matter, discussed in the next section. However, it can not yet be excluded that the rise in anti-positrons maybe due to not yet known or poorly understood conventional astrophysical sources.

Also, there are some hints from white dwarf cooling rates which also may support indirect effects from additional matter particles, as they cool too efficiently as possible with just the available particles. But once again the effect may still be purely conventional.

7.4.2 Dark matter

One of the most striking observations is that the movement of galaxies, in particular how matter rotates around the center of galaxies, cannot be described just by the luminous matter seen in them and general relativity. That is actually a quite old problem, and known since the early 1930s. Also gravitational lensing, the properties of hot intergalactic clouds in galaxy clusters, the evolution of galaxy clusters and the properties of the large-scale structures in the universe all support this finding. In fact, most of the mass must be in the form of “invisible” dark matter. This matter is concentrated in the halo of galaxies, as analyses of the rotation curves and colliding galaxies show. This matter cannot just be due to non-self-luminous objects like planets, brown dwarfs, cold matter clouds, or black holes, as the necessary density of such objects would turn up in a cloaking of extragalactic light and of light from globular clusters. This matter is therefore not made out of any conventional objects, in particular, it is non-baryonic. Furthermore, it is gravitational but not electromagnetically active. It also shows different fluid dynamics (observed in the case of colliding galaxies) as ordinary luminous matter. Also, the dark matter cannot be strongly interacting, as it otherwise would turn up as bound in nuclei.

Thus this matter has to have particular properties. The only particle in the standard model which could have provided it would have been a massive neutrino. However, though the neutrinos do have mass, the upper limits on their mass is so low, and the flux of cosmic neutrinos too small, to make up even a considerable fraction of the dark matter. This can be seen by a simple estimate. If the neutrinos have mass and would fill the galaxy up to the maximum possible by Fermi-statistics, their density would be

$$n_\nu = \frac{p_F^3}{\pi^2} \tag{7.4}$$

with the Fermi momentum p_F in the non-relativistic case given by $m_\nu v_\nu$. Since neutrinos have to be bound gravitationally to the galaxy, their speed is linked via the Virial theorem to their potential energy

$$v_\nu^2 = \frac{G_N M_{\text{galaxy}}}{R}, \tag{7.5}$$

with Newton's constant G_N and R the radius of the galaxy. Putting in the known numbers, and using furthermore that the observational results imply that n_ν, the total number of neutrinos approximated to be inside a sphere of the size of the galaxy, must give a total mass larger than the one of the galaxy leads to the bound

$$m_\nu > 100\text{eV}\left(\frac{0.001c}{3v_\nu}\right)^{\frac{1}{4}}\left(\frac{1\text{ kpc}}{R}\right)^{\frac{1}{2}},$$

yielding even for a neutrino at the speed of light a lower bound for the mass of about 3 eV, which is excluded by direct measurements in tritium decays.

Therefore, a different type of particles is necessary to fill this gap. In fact, many candidate theories for physics beyond the standard model offer candidates for such particles. But none has been detected so far, despite several dedicated experimental searches for dark matter. These experiments are enormously complicated by the problem of distinguishing the signal from background, in particular natural radioactivity and cosmic rays. The source of this matter stays therefore mysterious.

But not only the existence of dark matter, also its properties are surprising. The observations are best explained by dark matter which is in thermal equilibrium. But how this should be achieved if it is really so weakly interacting is unclear.

On the other hand, the idea of gravitational bound dark matter is also problematic. In particular, there is no reason why it should neither form celestial dark bodies, which should be observable by passing in front of luminous matter by gravitational lensing, or why it should not be bound partly in the planets of our solar system. Only if its temperature is so high that binding is prohibited this would be in agreement, but then the question remains why it is so hot, and what is the origin of the enormous amount of energy stored in the dark matter.

It should be noted that there are also attempts to explain these observations by a departure of gravity from its classical behavior also at long distances. Though parametrizations exist of such a modification which are compatible with observational data, no clean explanation or necessity for such a modification in classical general relativity has been established. This proposal is also challenged by recent observations of colliding galaxies which show that the center-of-mass of the total matter and the center of luminous matter move differently, which precludes any simple modification of the laws of gravity, and is much more in-line with the existence of dark matter.

Another alternative is that the problem of asymptotic infrared safeness of quantum gravity may be related to the apparent existence of dark matter.

7.4.3 Inflation

A second problem is the apparent smoothness of the universe around us, while having at the same time small highly non-smooth patches, like galaxies, clusters, super clusters, walls and voids. In the standard model of cosmological evolution this can only be obtained by a rapid phase of expansion (by a factor $\sim e^{60}$) of the early universe, at temperatures much larger than the standard model scale, but much less than the gravity scale. None of the standard model physics can explain this, nor act as an agitator for it. In particular, it is also very complicated to find a model which at the same time explains the appearance of inflation and also its end.

However, the predictions of inflation have been very well confirmed by the investigation of the cosmic microwave background radiation, including non-trivial features and up to rather high precision.

7.4.4 Curvature, cosmic expansion, and dark energy

Another problem is the apparent flatness of the universe. Over large scales, the angle sum of a triangle is observed to be indeed π. This is obtained from the cosmic microwave background radiation, in particular the position of the quadrupole moment[1], but also that the large-scale structure in the universe could not have been formed in the observed way otherwise. For a universe, which is governed by Einstein's equation of general relativity, this can only occur if there is a certain amount of energy inside it. Even including the unknown dark matter, the amount of registered mass can provide at best about 30% of the required amount to be in agreement with this observation. The other contribution, amounting to about 70%, of what origin it may ever be, is called dark energy.

A second part of the puzzle is that the cosmic expansion is found to be accelerating. This is found from distant supernova data, which are only consistent if the universe expands accelerated today. In particular, other explanations are very hard to reconcile with the data, as it behaves non-monotonous with distance, in contrast to any kind of light-screening from any known physical process. Furthermore, the large-scale structures of the universe indicate this expansion, but also that the universe would be too young (about 10.000.000.000 years) for its oldest stars (about 12-13.000.000.000 years) if this would not be the case. For such a flat universe such an acceleration within the framework of general relativity requires a non-zero cosmological constant Λ, which appears in the Einstein equations (7.2). This constant could also provide the remaining 70% of the mass to close the universe, and is in fact a vacuum energy. Such a constant is covariantly conserved, since both $T_{\mu\nu}$ and the first two terms together are independently in general relativity, and thus indeed constant. However, the known (quantum) effects contributing to such a constant provide a much too large value for Λ, about 10^{40} times too large. These include quantities like the chiral and gluon condensates[2]. These are of order GeV, and in addition would have the wrong sign. What leads to the necessary enormous suppression is unclear.

7.4.5 Matter-antimatter asymmetry

In the standard model, matter and antimatter are not perfectly symmetric. Due to the CP violations of the electroweak force, matter is preferred over antimatter, i. e., decays produce more matter than antimatter, and also baryon and lepton number are not conserved quantities. However, this process is dominantly non-perturbative. The most striking evidence that this is a very weak effect is the half-life of the proton, which is larger than 10^{34} years. Indeed, only at very high-temperature can the effect become relevant.

[1] The homogeneity of the universe leads to a vanishing of the monopole moment and the dipole moment originates from the observers relative speed to the background.

[2] There are some subtleties involved here with renormalization in a quantum theory, and the true source of the problem is that the energy scale of gravity and the standard model are so vastly different. This is ignored here for the sake of the argument.

After the big-bang, the produced very hot and dense matter was formed essentially from a system of rapidly decaying and recombining particles. When the system cooled down, the stable bound states remained in this process, leading first to stable nucleons and leptons in the baryogenesis, and afterwards to stable nuclei and atoms in the nucleosynthesis. Only over this time matter could have become dominant over antimatter, leading to the stable universe observed today. But the electroweak effects would not have been strong enough for the available time to produce the almost perfect asymmetry of matter vs. antimatter observed today, by a factor of about 10^{19}. Thus, a further mechanism must exist which provides matter dominance today.

There is a profound connection to inflation. It can be shown that inflation would not have been efficient enough, if the number of baryons would have been conserved in the process. In particular, the almost-baryon-number conserving electroweak interactions would have permitted only an inflationary growth of e^{4-5} instead of e^{60}.

The possibility that this violation is sufficient to create pockets of matter at least as large as our horizon, but not on larger scales has been tested, and found to yield only pockets of matter much smaller than our horizon.

A further obstacle to a standard-model-conform breaking of matter-antimatter symmetry is the necessity for a first order phase transition. This is required since in a equilibrium (or almost equilibrium like at a higher-order transition), the equilibration of matter vs. anti-matter counters the necessary breaking. However, with a Higgs of about 125 GeV mass, this will not occur within the standard model.