Nuclear and Particle Physics

Lecture in SS 2016 at the KFU Graz

Axel Maas

Contents

Chapter 1

Introduction

Nuclear and particle physics are essentially at the forefront of nowadays understanding of physics. Except for the astrophysical sciences it is here where one is at the edge of conceptual knowledge. In contrast, for problems of solid or applied physics we known essentially what are the correct theories, and the focus is today on the study of emergent phenomena or on applications. Within this lecture, however, the emphasis is on the frontier.

Nuclear physics was essentially the paradigmatic example of understanding particle physics. In itself, its basic description is nowadays also well understood. Still, nuclear physics is a perfect example of how particle physics works, and therefore still of significant importance. It is also in itself quite important, as various aspects influence in many different ways our everyday life, from medicine to the burning of the sun.

Particle physics proper, on the other hand, 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 linked to quantum mechanics, as quantum effects dominate the world of the elementary particles. However, as will be seen, it is also linked to astrophysics: The theory of elementary particles is linked tightly to cosmology, and the behavior of many celestial bodies.

It is also this field which will be presented in this lecture. After the treatment of nuclear physics and 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 later in the lecture. After that, several candidates for extensions of the standard model will be introduced, and briefly discussed. Given their enormous numbers, and each of them warranting an own specialized lecture, these expositions can give only the roughest idea. It must furthermore be noted that particle physics is a very dynamical and living field, and the influx of new results, both experimentally and theoretically, is large. Hence, 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 the 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. Still, a few remarks on this topic will be given towards the end of the lecture.

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 later literature studies. Therefore, no general recommendations for this lecture will be given.

However, the latest experimental results for the standard model included in this lecture are from the year 2016, and at the time of writing, 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 for the standard model. 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 roughly 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 an overview talk at a purely theoretical conference. Nonetheless, even in the latter cases, many of the ideas and underlying physics should sound then sound at least familiar to you. Eventually, the aim of this lecture is to prepare you to follow also the future discourse on particle physics, such that it can be incorporated to your teaching activities.

Chapter 2

Nuclear physics

2.1   Natural units

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

$$1 \text{ eV} = 1.60 \times 10^{-19} \text{ J} \tag{2.1}$$
$$1 \text{ fm} = 10^{-15} \text{ fm,} \tag{2.2}$$

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 units 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 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.

2.2   Atoms and nuclei

The periodic table of elements has shown very early on that there is a deep structure behind the types of chemical elements observed in nature. It were than the experiments of Rutherford at the turn of the 19th to the 20th century which showed how this structure came about. Every atom consists out of two components. These are a positively charged nucleus, about 10^5 times smaller than an atom - about 1 fm in size. The other is a number of electrons just compensating for the positively charged nucleus, which fill out essentially the rest of the atom. It was furthermore found that the nuclei come actually in a variety of types even for a fixed chemical element. However, they only differed in their mass, but the chemical properties were not affected. Such chemically identically but differing mass nuclei were called isotopes. The reason for this behavior was found by the fact that the chemical properties are entirely determined by the number of electrons, while the mass is essentially made up, up to a fraction of 0.05% by the nucleus. Since atoms are electrically neutral, the number of electrons must equal the nuclei's positive charge Z. Thus, the mass and the charge of the nuclei must be two independent quantities to explain this observation.

2.3   Structure of nuclei

The reason behind this independence was found out to be that also the nuclei themselves have a substructure. They consists out of two independent particles, which have both roughly the same mass, 1 GeV, about 2000 times that of the electron, but only one is electrically positively charged, the proton, while the other one is not electrically charged, and called neutron. The difference in mass at fixed electric charge could therefore be explained by the differing number of neutrons at a fixed number of protons.

The charge of a single proton is just the same but opposite of the one of the electron. Hence, the minimal composition of a nuclei must be Z protons, and then enough neutrons to obtain the final mass, such that the total count of protons and neutrons is A, and the total mass of the nuclei is thus A times the mass of the two particles, which are called collectively nucleons.

2.4   The strong nuclear force

The existence of nuclei then immediately poses the question how such an object can exist. Same-sign electric charges repel each other, and thus the protons in a nuclei should

immediately move way from each other. Also, the neutrons could not bind, since they are electrically neutral, and gravitation would be far to weak to bind them on the scale of a nuclei. The only solution to this puzzle is the introduction of a new force, which is called the strong nuclear force. The origin of this force will be discussed in much more detail in chapter 7. For the moment, it is sufficient to accept that this force exists, and most of its effect can be described by a potential between two nucleus of the form

$$V(r) \sim \frac{e^{-mr}}{r}, \tag{2.3}$$

where the size of the constant m is about 150 MeV. This potential is quite different from both electromagnetic and gravitational forces due to the factor $\exp(-mr)$. This factor reduces the size of the potential much quicker than any $1/r$ behavior. It is thus said that the force is short-range. This also explains why this force is not felt beyond the limits of a nuclei: Already at a distance of about 10 fm, the size of the largest known nuclei, it is reduced by a factor of about five orders of magnitude compared to the electric force at the same relative distance.

Though this seems to make the nuclear force simpler than the electric and gravitational force, this is only superficially so. As will be seen, to really describe nuclei quantitatively requires to include further terms in the potential.

2.5 Model of the nuclei

Nucleons themselves have a finite size, about 1 fm. Comparing this to the average size of the nuclei of a few fm, this implies that the nucleons are rather densely packed inside the nucleus. Still, because the force (2.3) is so short range, the nucleons are essentially only affected by some average impact of the other nucleons around it, an approximation known as mean-field. Thus, a single nucleon can be considered to be moving alone in an effective potential derived from (2.3). This potential takes the form of the so-called Woods-Saxon potential

$$V(r) \sim \frac{1}{1 + e^{\frac{r-R}{a}}} \tag{2.4}$$

with some characteristic constant R, the average size, and a the range of the force. The structure already indicates that it is related to the potential (2.3) by a kind of series resummation.

Since it turns out that nucleons are fermions, like electrons, they are affected by the Pauli principle. Hence, constructing the state of a nucleus appears to be as simple as for an atom: Just determine the energy levels, like for the hydrogen atom, and then fill

up all levels as long as there are nucleons available. Since protons and neutrons can be distinguished by their charge, there are actually two sets of levels to be filled. The remainder of the presence of the other nucleons can then be accounted for, just like with the Helium atom, by perturbation theory.

Unfortunately, it turns out to be not as simple. In atoms, there are shells, which are exceptionally robust. These are obtained when a given set of (almost) degenerate energy levels have been filled up by electrons, and any further electron would have a considerably lower binding energy. The potential (2.3) predicts also such filled shells, separately for protons and neutrons, but they turn out to be in stark contrast to observations: Shells are observed, but at entirely different numbers of nucleons.

The reason for the discrepancy is that an effect, which is very small for electrons in the atomic shell, becomes much more important for nucleons: The spin-orbit coupling. The total potential is thus

$$V_t(r) = V(r) + V_{ls}(r)\vec{l} \cdot \vec{s},$$

and thus of the same form as for atoms. The function V_{ls} is, in principle, an independent function. However, experimentally it is found that

$$V_{ls} = \frac{1}{r}\frac{dV}{dr}.$$

This relation can actually be understood to be a consequence of the nature of the strong force, and can be derived. This also shows already that both potentials are of similar size, and both must be taken into account.

This modified potential is, as in the atomic case, still creating degenerate energy levels, but reshuffles the number of occupants to create a closed shell. These magic numbers are 2, 8, 20, 28, 50, 82, 126 for those which are currently observed. Higher numbers exist, of course, as well. Thus, this then represents a very good description of the observed special stability of nuclei, especially when it comes to so-called 'doubly magic nuclei', i. e. those where for both protons and neutrons a shell is filled, though not necessarily the same one.

It should be noted that typical binding energies for nuclei are much larger than for atoms, of the order of a few MeV. The binding energy of a nucleon in an infinitely large nucleus would be 16 MeV. In contrast to atoms, this is of the order of a percent of the rest mass of the nucleons, and therefore the binding energy leads to a sizable mass defect

$$\Delta m = m_{\text{Nucleus}} - Zm_p - (A-Z)m_n,$$

that is the difference between the nucleus and the total mass of the constituents is sizable.

A rough estimate of the mass is given by the empirical Weizsäcker's mass formula

$$
\begin{aligned}
m(Z, A) \quad = \quad & Zm_p + (A - Z)m_n - (15.85 \text{ MeV})A + (18.34 \text{ MeV})A^{\frac{2}{3}} + (0.71 \text{ MeV})\frac{Z^2}{A^{\frac{1}{3}}} \\
& + (92.86 \text{ MeV})\frac{\left(Z - \frac{A}{2}\right)^2}{A} \pm (11.46 \text{ MeV})A^{-\frac{1}{2}}
\end{aligned}
$$

The first two terms are just the mass of the constituents. Since the nucleons are densely packed, the volume of a nucleus is proportional to the maximum number of nucleons A. The third term is thus a volume energy. In infinite nuclear matter with the same number of protons and neutrons, this is the only relevant term. The next term is then a surface term, and thus originates from a similar effect as a surface tension. The next term comes from the Coulomb repulsion between the protons. It is thus influenced by the relative number of protons and neutrons. The next term is an asymmetry energy, which is minimized for the same number of protons and neutrons. Finally, the last term contributes with negative sign if the number of protons and neutrons are odd, while it is positive if both are opposite. The last two effects have something to do with the spin-orbital coupling, and the possibility how to also pair protons and neutrons, as well as the fact that protons and neutrons are quite similar for reasons to become more clear later. This formula describes most nuclear masses already to a rather good accuracy, provided they are not too heavy and do not have too different numbers of protons and neutrons. It is noteworthy that the iron isotope with $A = 56$ is the most stable isotope, i. e. its binding energy is largest. This has numerous astrophysical implications. If the number comes out positive, this implies that the nucleus is unstable. This will certainly happen if the imbalance between protons and neutrons is too large - the Coulomb repulsions will then just tear the nucleus apart. This is, e. g. the reason why there is no particle consisting only of, say, two protons. At the same time, the asymmetry energy ensures that the number of protons and neutrons can never become too large. That is the reason, why there is, say, no hydrogen with 10 neutrons. Note, however, that this formula does not describe shell closures, and thus only the rough shape rather than a detailed quantitative estimate of the binding energies.

Though purely phenomenological, this formula is reproduced to good accuracy with the description in terms of quantum mechanics used beforehand. Especially, this underlines that the assumption that the nuclei is essentially a droplet of more or less independent nucleons is a more or less accurate picture. This also explains the big differences compared to the atom, which is rather like a gas of electrons. But in the end, the reason for these differences comes from the facts that the strong nuclear force (2.3) is short-range, but actually has a very large pre-factor, and acts in the same way on both protons and neutrons.

2.6 Nuclear transitions

One property of both practical and fundamental interest is that some nuclei are not necessarily immutable, even when no external influence acts on them. In a sense, this was a new feature when nuclear physics was discovered. In atomic and molecular physics, this was not really seen before. Whenever atoms or molecules dissociated, there was something external at work. However, with hindsight, also in atomic physics situations can be setup where such a spontaneous disintegration can be observed. This happens, if artificially more and more electrons are attached to a single atom, such that it becomes negatively charged. If such a situation occurs in nature, this usually leads immediately to a chemical reaction with other present substances. But if an atom is kept in isolation, this can be avoided. While such a negatively charged ion may still be stable when adding only a few electrons, at some point it becomes unstable, and starts to spontaneously eject, after some time, an electron, usually accompanied by some photons. Since this is a quantum-mechanical process, the time the surplus electrons are kept is not predetermined, but only statements about probabilities can be made.

Another example are lasers. Here, an electron is pushed artificially into a higher level by a pumping light. This excited atoms then decays into its ground state by the emission of a photon, again after some random time.

Nuclear decays are quite similar in concept. There are stable nuclei, corresponding to the neutral atoms, and unstable ones, corresponding to highly charged atoms. However, the analogy only carries so far. In nuclear decays, the individual components of the nuclei, the nucleons, may undergo a change as well. This is different from the previous case, as there the electrons and atomic nuclei remain unchanged. However, this change is not the real decay, but rather creates the surplus of particles the additional electrons have been in the previous example, which then triggers the decay. In fact, it is often possible to transform nucleons forward and backward, and only after several cycles the decay occurs.

Thus, nuclear decays are not something mysteriously different, but just something different in details. It is only because the energy scales, and thus time scales, of the strong nuclear force are so different that they are perceived differently. Also, in the previous two cases, some external artificial source was needed to create a situation which could yield a decay. For nuclear decays, the source is equally 'artificial': Usually nuclear reactions in stars create unstable nuclei at some point in the far past. Only because of the exceedingly long life-times of some of these unstable nuclei it appears as if they have been there already all the time. But they also needed to be created initially.

The reason for the decay is just that the decay products have less energy, and the decayed state is thus more favorable. Thus, the situation prior to the decay is just metastable,

albeit the barrier to transformation may be so high that the tunneling times may be larger than the age of the universe, possibly by many orders of magnitude.

Stable nuclei are those that not decay on their own, like the most stable ^{56}Fe nucleus or the hydrogen nucleus made out of a single proton. Others have just so long life-times that for all practical purposes they are stable. Things are often subtle, and other issues, like conversation of angular momentum may play a role as well. In the end, everything will decay, as long as the final configuration has less energy than before and no conservation laws are violated. Thus, to determine whether a nucleus is stable, its total energy is compared to those of the potential decay products, and whatever kinetic energy is required to satisfy momentum conservation as well as angular momentum conservation. It will decay, ultimately, if this energy is less. However, the closer the two energies are, the longer the life-time. The increase in life-time with the decrease in energy difference is essentially exponential, and it is thus possible to quickly reach enormously long life-times. In addition, pre-factors, as non-trivial effects due to the decay kinematics, can effectively decrease the decay even much further.

2.6.1 Decays

Another difference to the atomic cases is that nuclear decays can occur in many different versions. These will be treated here step by step.

The common feature of all the decays is that they are characterized by a decay probability. The determination of this probability is essentially a quantum-mechanical problem. It is based upon the energies of the initial and final states, as well as the potential barrier in between, similar to a double-well potential with two differently deep minima in between. The calculation of this number depends on many details, and will not be detailed here. The end result is always a decay probability for a single nucleus.

Taking some chunk of such nuclei, the probability to observe N not-decayed nuclei at some time t is then given by

$$\frac{dN}{dt} = -\frac{1}{\tau}N,$$

where $1/\tau$ is the decay probability, that is the number of decays per second. This equation is integrated to

$$N(t) = N_0 e^{-\frac{t}{\tau}}, \tag{2.5}$$

and thus nuclear decays follow an exponential behavior, and τ is the half-life and N_0 the initial amount.

If a decay can occur in different ways, numbered by an index i, and there is a (time-

independent) production rate s, the corresponding equation is

$$\frac{dN}{dt} = -\sum_i \frac{1}{\tau_i} N + s,$$

which integrates to

$$N(t) = \frac{s}{\sum_i \frac{1}{\tau_i}} + N_0 e^{-\sum_i \frac{1}{\tau_i} t}. \tag{2.6}$$

Thus, an equilibrium is found at long times, depending on the total decay width

$$\Gamma = \sum_i \frac{1}{\tau_i}$$

which is the dominating quantity to characterize (nuclear) decays.

2.6.1.1 γ decay

Possibly the simplest decay is the γ-decay. The starting point is very similar to the laser. The nucleus is not in a ground-state to begin with, but in an excited energy state of the potential, e. g. the effective Woods-Saxon potential. It can thus decay to its ground state by emission of one (or more) photons, which in this context are also called γ quanta or X-rays, the latter for historical reasons. Since the typical energy levels of nuclei are usually spaced by some hundred of keV up to a few MeV, with exceptions in both directions, the photons have much more energy than their atomic counter-parts, thus their name[1].

The quantum-mechanical description of this type of decays proceeds essentially as with the spontaneous emission of laser photons, just that different energy levels and a different potential are involved. This will therefore not be detailed further.

2.6.1.2 α decay

As noticed, there are particularly stable nuclei if both, the number of protons and the number of neutrons, are magical numbers. The smallest such nucleus is the nucleus of helium, having two protons and two neutrons. It is thus an especially stable nucleus,

---

[1]These are precisely the source for medical applications, both for diagnostics and treatment. There are many other industrial applications, e. g. disinfection or diagnostics of material. There are technically based on the interaction of MeV-range energetic photons, and therefore their use has nothing to do with nuclear physics anymore. Thus, other sources than nuclei decays are nowadays often used, like X-ray tubes. Thus, their application is not specifically nuclear physics, and therefore not elaborated on here. Similar statements apply to all the other decay channels, where the decay products just vary. Their interaction with matter is in all cases mostly due to electromagnetic effects. The only exceptions are rare nuclear reactions at high energies, but these will be discussed more generally in section 2.7.

though not as stable as the most stable nucleus ^{56}Fe. Nonetheless, the consequence is that for many nuclei away from a double or even single magic number, it is energetically favorable to decay into a helium nucleus and a remainder nucleus. This is supported by the electromagnetic repulsion, once the helium nucleus has moved a few fermi away from the mother nucleus. Thus, there is a probability for the helium nucleus to form inside another nucleus, and then tunnel through the barrier erected by the strong interaction. Such a decay is called an α decay, and the emitted helium nucleus an α particle, despite being just an ordinary helium nucleus, for historical reasons.

2.6.1.3 Spontaneous fission and nucleon evaporation

Though the emission of an α particle is by far the most frequent so-called nuclear break-up decay or nuclear fission, it is by no means the only one. In principle, a nucleus can always break up into two lighter nuclei, if it is energetically favorable. Due to the large binding of magic and double-magic nuclei, these are the preferred cases in which to break up. This usually happens only for larger than helium nuclei when the mother nucleus is itself already quite heavy, because the α nucleus has also against many other magic nuclei a very large binding. Thus, this spontaneous fission occurs predominantly at very large atomic numbers A.

However, when a nucleus has exceedingly many protons or neutrons, it can happen that these nucleons are literally evaporated from the nucleus, i. e. just get randomly slightly too much energy, and therefore can escape the potential well, just like an atom can escape from a liquid. The point where this becomes an almost immediate process is called the neutron and proton drip-line.

2.6.1.4 β decay

There is a process which is completely different from the three other types of decays. These three types are mediated by the already known forces, the electromagnetic and the strong force. These forces do not conserve nuclei, but they conserve the identity of nucleons: Protons and neutrons are always unchanged. There is, however, another force, the weak force, where this is no longer the case. The weak force can transform the neutron, which is somewhat heavier than the proton by about 1.5 MeV, into a proton, an electron, and a so-called (anti-)neutrino. The latter particle does almost not interact afterwards, and therefore is for most aspects of nuclear physics irrelevant, except that it can carry energy and angular momentum away from the scene. The details of both this weak force and the neutrinos will be discussed at length later in the lecture in chapter 8.

The important point here is that by this process a nucleus is transformed into one of the same number of nucleons, but one proton more, as well as the electron and neutrino. This is always possible, if the new nucleus is lighter by more than what it at minimum required to create the electron and neutrino and to conserve energy and momentum. This type of decays is called β-decay, and the resulting electron for historical reasons β-particle. A famous example of this kind of decay is the one of tritium, i. e. a hydrogen nucleus with two extra neutrons, into helium-3, i. e. a helium nucleus with one neutron less compared to the ordinary ones.

Less common is also an inverse type of β-decay. In this case, a proton is converted into a neutron and a positron, a particle like the electron but with positive sign, is ejected, together with a neutrino. This appears at first to be impossible, as the neutron is heavier. However, this is a quantum mechanical effect, where it is only relevant that the total state's energy is conserved. This can be achieved, if the final binding energy is lower, and the amount of binding energy won can be used to supply the proton with the necessary amount of energy. This process is somewhat involved, and will be discussed later once more, after the corresponding concepts have been introduced in the next chapter.

Another possibility are so-called electron capture processes. These are possible, if the nucleus is inside an atom. It can then be advantageous, in a very similar way as before, to capture an electron from the hull, and combine it with a proton to get a neutron. Again, as long as the final state has lower binding energy, this is possible.

Note that the electron (and positron) can also interact with the other electrons in the hull, if present. This may lead to various reactions, which can result in the emission of further electrons and/or of γ rays. E. g., an electron capture usually captures one of the innermost electrons. This space is then occupied by an outer electron. This transition yields the usual transition radiation, as for an excited atom.

2.6.2   Stability islands

A quite interesting question is the competition between stabilizing effects, e. g. due to shell completions, and the energy gain due to decays into more stable nuclei. Since the level spacings for magic shells increase the more nucleons are present, at some point all nuclei become unstable. This is reached at lead, ^{208}Pb, to be precise. After that, no stable nuclei are known. However, the next magic shell is further out than so far accessible. Since calculations of decay rates are difficult for such heavy elements, a logical possibility is that there are at, or around, magic numbers, especially doubly-magic numbers, again one, or more, stable or almost stable nuclei. These would be called islands of stability. Though not observed on earth, their production could be so rare, that they just have not been seen

directly. This question is yet unsolved. However, it is known that at some point nuclei will again become stable, in the form of neutron stars to be discussed later. But strictly speaking, this is rather a new state of matter than just a big nucleus, and especially one where gravity plays a non-negligible role.

2.7 Nuclear reactions

The nuclear reactions described so far are those which occur for a nucleus in isolation, or at most embedded in its atomic shells. There is an enormous amount of nuclear reactions possible, once multiple nuclei come into the game. Such reactions proceed by the collisions of two (or more) participants, which then either form a single product, or split again in two or more reaction products, which may be quite different from the initial ones.

Such nuclear reactions play singular important roles. Probably the most important is the burning of the sun, and thus the possibility for life on earth. This is entirely due to nuclear reactions, and nuclear physics can described these processes at least qualitatively, if not yet in all details quantitatively perfectly. Nuclear reactions play also a central role in nuclear reactors, but are also of central importance in the creation of isotopes for the purpose of medical and industrial uses.

An important aside is that nuclear reactions depend quite strongly on the initial conditions, especially the amount of energies carried by the involved particles. Especially, for some amount of energy a process may occur very often, while it is extremely rare for other parameters. Especially for processes occurring in astrophysical bodies the required parameters may be very hard to realize in an experiment, and thus often results are indirect.

2.7.1 Fission

The best known case of nuclear reaction is nuclear fission. In general, nuclear fission is essentially that a nucleus breaks up. This can happen in some cases as a decay, but also if the nucleus is interacting with some other particle. A simple example is that of ^{7}Be, which when shot at with a proton will decay into two helium nuclei. In this case, for a brief moment, a compound nucleus may be formed, ^{8}Be, which however is highly unstable. At higher energies, the interaction may also create more fragments, e. g. a helium nucleus, a tritium nucleus, and the proton.

Fission targets do not need to be stable nuclei themselves. E. g. the fission of an uranium nucleus after being hit by a neutron is probably the most well-known example of nuclear fission. It is also the practically most important one, as it is the one employed in

most nuclear power plants. This is also a process where the parameter dependence plays an important role, as this will occur efficiently only for neutrons and uranium of very small relative speed. The speed is about the same as those obtained from thermal motion, and therefore the neutrons, which are usually the ones shot at the uranium nuclei, are also called thermal neutrons.

2.7.2   Fusion

Fission is encountered as α-decay or spontaneous fission already as a property of a single nucleus. More interesting is a process which cannot occur on its own: Fusion.

Fusion is a process where two nuclei are brought together, and form a single nucleus afterwards. This requires to first overcome the Coulomb barrier due to the electromagnetic repulsion between the two initial nuclei. This can be achieved by propelling one at the other, though the energy must be suitably chosen to avoid a fission process.

The probably simplest, but at the same time most important, fusion process is the fusion of two deuterium nuclei into a ^{4}He nucleus. This is a process which is occurring within suns abundantly. In the same manner, all heavier elements have been created out of essentially only hydrogen during both cosmological and stellar evolution. However, a fusion process may yield also unstable nuclei, which can then decay. It is also possible to fuse more than two nuclei at the same time. E. g., the process $3\,^4\text{He}\rightarrow\,^{12}\text{C}$ is of fundamental importance for the solar burning. However, even under optimal condition the efficiency of such n-body processes is so small that it is almost impossible to realize them in experiments.

A fusion process may also be directly accompanied by a fission process. An example is the fusion of two hydrogen nuclei, protons, to a deuterium nuclei, the first and most important step of stellar burning as well as the early nuclear cosmological evolution, the nucleosynthesis. In this case, the protons are brought together, in stars just by thermal motion. Since two protons do not form a stable system, it could be expected that this will not produce a stable result. However, the amount of energy available usually permits to perform a β^+-decay, i. e. one of the protons is transformed into a neutron under emission of a positron and a neutrino, thus forming stable deuterium nucleus, ^{2}H.

2.7.3   Exchange reactions

Another possibility are exchange reactions. E. g. $^2\text{H}+\,^3\text{He}\rightarrow\text{p}+\,^4\text{He}$ plays a role in stellar reactions. Here, the exchanged particle is a neutron, and both nuclei still exist afterwards, though modified.

Such exchange reactions can also be mediated by other particles, and sometimes much more complicated objects than a simple nucleon. It is also possible that almost all of the nucleons are afterwards concentrated in one of the initial particles. This is an important process to create heavy nuclei, and used to search for the islands of stability.

2.7.4   Decay chains, networks, and chain reactions

Many decays actually do not go to a stable nucleus, but to other unstable ones. In such a case, the relative decay times and abundances play an important role. If the starting nucleus decays very quickly, then the situation moves to the first nucleus, which is comparatively stable, with respect to all ancestors states. It then provides daughter states with a rate slow compared to the available amount according to (2.5) and thus the production rate of the daughter states will be for a long time rather constant. Thus, the first generation daughter states will form an equilibrium according to (2.6). Since usually as long as in such a decay chain the life-times increase, otherwise the intermediate states can be more or less skipped, such an equilibrium is formed on the level of all metastable states.

Such decay chains play an important role in forming the relative abundance of unstable elements observed. The primary production mechanism is mostly due to supernovas[2], where the initial states are created in various fusion processes. They decay to metastable states, which then form the decay chains described.

Decay networks arise if states can either be decaying in various channels, or fusion processes supplement new ancestor nuclei. Based on the individual structures, such decay networks can become very complicated. They play central roles in the burning of stars, where fusion supplies nuclei, which then can or cannot decay. The most extreme cases of such decay networks arise in supernovas, where the gravitational collapse of a star provides enough energy to drive many otherwise impossible fusion processes, and thus filling many shortlived ancestor states to initiate decay networks.

That is also an example of a chain reaction. Such a chain reaction occurs in the following situation: There is some nucleus A, which when fused with another nucleus B decays, and produces in this course again one or more nuclei of type B. If there is some amount of nucleus A packed sufficiently tight, critical, that the number of decays produces sufficiently many nuclei of type B that the fusion process can undergo efficiently, there may be a rapid buildup of decays, which in turn create more and more nuclei of type B until the reservoir of nuclei of type A becomes so strongly depleted that the chance of a B nucleus hitting an A nucleus becomes again sub-critical, and the chain reaction stops.

---

[2]There are indications that the collision of neutron stars may also contribute a significant amount.

This is the process used in conventional nuclear power plants. Here, the amount of reactions is moderated by engulfing the nuclei of type A, usually uranium, in a material which efficiently stops nuclei of type B, here neutrons. If this moderator is reduced, the number of reactions increases. A nuclear meltdown occurs as a chain reaction if too much moderator is removed. There are several modern processes which work sub-critical, so that such a risk does not occur. Nuclear weapons are based on intentionally starting a chain reaction.

Chapter 3

Particle Physics

3.1 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. This was discussed in chapter 2.

In the middle of the 20th century it was then found that the 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.

3.2 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 any (massive) 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 37 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. Nuclei are an example for these. Of course, composite objects can consist out of composite objects, like 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. This also illustrates that the subject of particle physics shifted throughout time. Atomic and nuclear physics could have been considered elementary particle physics, though this term had not been coined then, about a hundred years ago.

3.3   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.

3.4   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.

The simplest way in which anti-particles appear can be gleaned from the relativistic energy-momentum relation $E^2 - \vec{p}^2 = m^2$. Taking a particle at rest, $\vec{p} = 0$, yields $E^2 = m^2$. However, this result permits besides the usual solution $E = m$ also the solution $E = -m$. Such a negative energy solution makes no sense in ordinary quantum mechanics. In quantum-field theory, this second solution can be associated with the anti-particle. Thus, even this simplest case already harbors particles and anti-particles.

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.

3.5   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 interactions 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; theoretically, 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 again of the famous Einstein equation $E^2 = m^2$, 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, and thus more rest mass than originally available as 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 other particles. 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, as well as from the uncertainty principle proper.

Chapter 4

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 alluded 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.

4.1   Transitions

The basic concept behind scattering experiments is that it is possible to perform a transition from one (initial) state $|\alpha\rangle$ to another (final) state $|\beta\rangle$, characterized by a transition matrix element $\langle\beta|\alpha\rangle$. Scattering itself is just a tool with which this transition is conve-

niently realized. It is thus worthwhile to briefly repeat the basic ideas behind transition and scattering in quantum mechanics.

Usually, particle physics theory can be separated in a free part and an interaction part. The former can be solved exactly, while the full theory is usually accessible using approximations. One of the most successful approximations is perturbation theory. Since almost all non-perturbative approaches are generalizations of perturbative approaches, it is always a good starting point to first understand the perturbative one, and then move on to the more complicated non-perturbative treatment. Hence, assume that α and β differ only slightly, in a sense to be made more precise below.

It can then be shown, to lowest order in a perturbative expansion, that the probability of a transition from a state α to some close-by (in terms of energy) eigenstate N of the Hamilton is then given by

$$P_{N\alpha} = 4|\langle N|H_{\text{int}}|\alpha\rangle|^2 \frac{\sin^2 \frac{(E_N - E_\alpha)t}{2}}{(E_N - E_\alpha)^2},$$

where E_α is the energy of the state α and E_N is the energy of this state, and t is the time the system is permitted to evolve.

Hence, at any fixed time, transitions to states with vastly different energies are strongly suppressed, which is consistent with the assumption of a small perturbation. In fact, for times T much shorter than half a period, essentially only states with energy difference $\Delta E \approx 2\pi/T$ will be populated. Hence, it can be assumed that the matrix element is essentially independent[1] of N, and it can be replaced by $|\langle \beta|H_{\text{int}}|\alpha\rangle|^2$. Thus, the difference between α and β must be such that they have similar energies.

Summing over N will provide the total transition probability. Since usual particle physics experiments will be on unbound particles, which are scattered, the energy levels can be assumed to be a continuum. Hence, the sum turns into an integral, weighted by the density of states, yielding as total transition probability

$$P(T) = 2\pi T|\langle \beta|H_{\text{int}}|\alpha\rangle|^2 \left.\frac{dn}{dE}\right|_{E=E_\alpha \approx E_\beta} \tag{4.1}$$

where the last factor is the density of states at the energy, also known as phase space $\rho(E)$. This formula is also known as Fermi's golden rule.

The structure of a scattering probability is always like the expression (4.1), even in much more complicated cases: Some numbers, the transition matrix element, which may be more complicated beyond leading order, and a phase space contribution of the final states. If part of the final state is not observed, it has to be summed or integrated over.

---

[1] Otherwise a perturbative treatment is not justified.

E. g. if the particles in the final state have spin, which is not observed, the result has to be averaged over all possible spin orientations. Concerning the initial state, in most of this text it will be assumed that is known as precisely as quantum mechanics permits.

4.2 Non-relativistic

Specializing now from general transitions to scattering experiments, it is useful that many of the necessary concepts can already be understood non-relativistically. This offers a simpler formalism.

4.2.1 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, between which two momenta the scattering angle is measured, and in which direction.

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.

4.2.2 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 various quantities, which can be derived from the cross section, which are of particular importance. One is the total cross section

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

which may still depend on further parameters, like the total center-of-mass energy. The others 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. Of course, differential cross-sections can also be defined from the total cross-section.

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. Nowadays, usually 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 contributions 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.

4.2.3 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 expected number of events N for a process 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 luminosities at the LHC are 25 fb^{-1} until the end of 2012, and at least 300 fb^{-1} until 2020.

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. Then one can define for each possible pairing an individual partial luminosity.

4.3   Relativistic

4.3.1   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,

$$
\begin{aligned}
p &= \gamma\beta m \\
E^2 &= \vec{p}^2 + m^2
\end{aligned}
\tag{4.2}
$$

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 (4.2) 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 (4.2) 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 (4.2) defines a surface in momentum space, called the mass-shell. Particles fulfilling (4.2) are consequently called on-shell.

One further consequence relevant to particle physics is time dilatation, i. e. that 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 by

$$\tau' = \gamma \tau. \tag{4.3}$$

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, for an observer, the faster they are. This has especially consequences for the experimental accessibility of unstable particles.

Elastic scattering proceeds then in the relativistic case just as in the non-relativistic case. The only change 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[2].

---

[2]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.

4.3.2 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,

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

This implies especially

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

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 with any reasonable rate. 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 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.

4.4 Formfactors

In practice, a very important way to describe cross-sections is by means of so-called form factors. Define q as the amount of momentum exchanged between two spin-less particles colliding into two other particles. Then a form-factor F is defined as

$$\frac{d\sigma}{d\Omega} = \frac{d\sigma}{d\Omega}_{\text{simple}} |F(q^2)|^2$$

The differential cross-section called 'simple' is the cross-section of two particles with point-like structure and exactly solvable interactions. Using Lorentz symmetry, it is possible to show that any influence of the internal structure can only depend on the momentum transfer q^2 between both particles, and therefore the deviation from such a structure can be parameterized by this single additional function, the form-factor. The interesting physics is hence entirely encoded in this form-factor.

Of course, if there are different particles involved, spin, or more than two particles in the final state, the structure is not as simple. However, Lorentz symmetry always permits to write the differential cross-section as a sum of products of trivial kinematic cross-sections and form factors. Hence, even in the general case, it is possible to parametrize a cross-section by a set of form factors. Since in this way all 'trivial' effects are removed, this is very useful to identify new effects.

4.5 Decays

In a sense a very particular type of inelastic processes is the decay of a single particle, i. e. a process 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 particles, plus possibly a certain amount of energy to satisfy spatial momentum conservation. Furthermore, there may be additional constraints like that electric charge is conserved. The following is thus a generalization of the nuclear decays treated in section 2.6.1.

Again, 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 smaller 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, already for technical reasons.

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. This is also true for nuclear decays.

For the experimental measurement of the properties of resonances time dilatation (4.3) 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.

A simple, and heuristically, mathematical description of decays in quantum physics can be achieved in the following way. Though sketchy, the results agree with one obtained in a more rigorous treatment.

For a time-independent Hamilton operator, the time-dependence of a quantum-mechanical state is given by the oscillatory factor $\exp(iE_0t)$ in the wave function $\Psi(x)\psi(t)$. This implies that the probability to find a particle is constant in time, as $|\psi|^2$ does not depend on the time anymore. If a particle should decay, This must change. This can be introduced by

giving the energy a small imaginary component, $E_0 \to E_0 - i\Gamma/2$. Then the wave-function is damped in time, characterized again by the decay width Γ.

A decaying probability is however hard to interpret. It is better to perform a Fourier transformation in time. The Fourier-transformed wave-function becomes, dropping the spatial part $\Psi(x)$ which only acts as a spectator,

$$\psi(E) = \frac{1}{(2\pi)^{\frac{1}{2}}} \frac{i}{E - E_0 + \frac{i\Gamma}{2}}.$$

From this the (normalized) probability to find the particle with energy E can be obtained

$$P(E) = \frac{\Gamma}{2\pi} \frac{1}{(E - E_0)^2 + \frac{\Gamma^2}{4}}.$$

This shows that the additional imaginary part of the energy broadens the distribution in energy space of the particle. A resonance has thus no associated fixed energy, and has a so-called natural linewidth.

Such decays do not only occur for free particles. There are also known, e. g., in the context of atoms. If an electron in an atom is in an excited state, the atom will also decay into its ground state by emission of a photon, and the electron dropping back to the lowest possible level. In this case, E_0 will play the role of the energy of the excited atom, with $1/\Gamma = t_{1/2} \ln 2$ its life-time, or half-life $t_{1/2}$.

4.6 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

$$\sigma \sim \frac{1}{((p_1 + p_2)^2 - m^2)^2} \tag{4.4}$$

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 diverges, 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 the center-of-mass energy s, this is also called an s-channel process.

In contrast, there are also processes where one particle can be transmuted into another one by emitting a third particle. In these cases the relevant variables to be tuned are different, e. g. 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 its finite, though possibly small, width Γ. Including this width leads to a Breit-Wigner cross section in, e. g. the s-channel

$$\sigma \sim \frac{\Gamma^2}{(\sqrt{s} - m)^2 + \frac{\Gamma^2}{4}}. \tag{4.5}$$

Thus, the divergences becomes a peak with a width determined by the decay width of the particle, which is therefore in principle 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 state. 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 lead beyond the present scope.

One should note that the formula (4.5) 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.

4.7   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. The lowest order is

usually called, under the assumption of quick convergence of the series, leading order, abbreviated by LO. The next order is called next-to-leading order, NLO, and so on as N...NLO or N^nLO. 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.

However such an expansion is not able to capture all features of essentially all quantum field theories. But this by no means implies that perturbation theory is useless. If performed carefully, it is easily, and in fact very often, possible that quantitatively this already gives the bulk of the result. In fact, for most aspects of particle physics, perturbation theory suffices to be competitive to the experimental precision.

But there are cases, in which this is not the case. Most prominent among them is bound state physics - since perturbation theory starts usually around a non-interacting theory, it knows only about elementary particles, but not about (stable) bound states. It is therefore very advisable to analyze a problem before choosing a particular method to deal with it.

Since for much of the remainder of this lecture perturbation theory is indeed providing already the dominant contribution, is is worthwhile to introduce the language of perturbation theory. Furthermore, non-perturbative physics is almost always a generalization of perturbation theory. Hence, first analyzing a problem in a perturbative language and then generalizing is in most cases the wisest course of action.

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. An interesting observation is that the LO contribution is always a graph without loops, and therefore LO is also called tree-level. On the other hand, all higher orders at most as many loops as the order, i. e. NLO at most one loop, and thus the expansion is also called loop expansion. Finally, it can be shown that every loop introduces one power of $\hbar$. Hence, the expansion is also an expansion in quantum effects, especially as at tree-level there is no non-trivial power of $\hbar$. Hence, tree-level is also called the classical contribution.

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 computer algebra system can perform. Especially vertices are restricted by conservation laws, like four-momentum conservation.

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.

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 5

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 implements special relativity.

5.1   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 5.8.

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, the vector potential does not appear in the Maxwell equations

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

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

$$
H = \frac{1}{2m}(i\partial_i - eA_i)^2 + eA_0, \tag{5.1}
$$

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

$$
A_\mu \rightarrow A_\mu + \partial_\mu\phi, \tag{5.2}
$$

---

[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. However, this shows that the physical degrees of freedom are the electromagnetic fields, rather than the vector potential. Hence, using the vector potential is just a convenience.

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

$$\psi \to \exp(-ie\phi)\psi. \tag{5.3}$$

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 experimentally supported theories 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 (5.1) 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]. This procedure is also known as minimal coupling. In principle, more complicated possibilities exist to obtain a gauge theory. But since so far no experimental support for such theories has been found, they will be skipped here.

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 (5.2-5.3) 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 3.4, 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.

---

[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.

5.2   The Dirac equation

As has been emphasized, the electron is a fermion. However, (5.1) only describes a scalar. The description of fermions in quantum field theory is somewhat more complicated. Hence, here only some of the most pertinent features will be given, and the remainder is left to a quantum field theory course.

The basic features distinguishing fermions from bosons is their half-integer spin, especially 1/2, and the Pauli principle. Half-integer spin is impossible for ordinary rotations. However, the Lorentz group permits the construction of such particles. The most important property is that it is not possible to write down a one-component wave-function for this purpose. At least, two (complex) components are needed. If a particle with an anti-particle should be described, as will be required exclusively for the standard model, two more components for this anti-particle are necessary. Thus, the simplest wave-function for QED is a four-dimensional complex object, a so-called spinor ψ. The name indicates that this quantity is different from an ordinary four-vector, despite having the same number of components. Especially, it does not transform in the same way as a four-vector under Lorentz transformations.

The dynamical equation describing a (free) fermion of mass m is given by the celebrated Dirac equation

$$(i\gamma^\mu\partial_\mu - m)\psi = 0, \tag{5.4}$$

where the γ^μ are some fixed 4×4 matrices of pure numbers. Note that the product $\gamma^\mu\partial_\mu$ is not a scalar, since the γ_μ are not a four-vector. This operator transforms covariantly under Lorentz transformation. Though the Dirac equation is a postulate, it has been confirmed experimentally, and is a viable description of a free fermion. It yields that the relativistic relation between energy and momentum is obeyed also by fermions, making them reasonable particles. It also implies that the fermion and the antifermion have the same mass. Finally, the solutions of this equation automatically obey the Pauli principle, completing the necessary property of fermions.

5.3   Formulation as a field theory

So far, the formulation at the level of (5.1) is quantum-mechanical. As noted, quantum mechanics cannot deal with relativistic effects, and one has to pass to quantum field theory (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 like electromagnetism, this changes again to a Lagrangian density $\mathcal{L}$, the Lagrangian. Classical Maxwell theory takes the form

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{5.5}$$
$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{5.6}$$

which is explicitly invariant under the gauge transformation (5.2). The equations of motion, as obtained in classical mechanics, yield the Maxwell equations. 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 vector potential since it carries a single Lorentz index, and therefore transforms like a vector, and is described by an ordinary field.

The total Lagrangian of QED is obtained by coupling electrons and positrons to the electromagnetic field. As in the quantum-mechanical case, this is achieved, and experimentally confirmed, by minimal coupling. This yields

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}\gamma_\mu(i\partial^\mu - ieA^\mu)\psi - m\bar{\psi}\psi, \tag{5.7}$$

where m is the mass of the electron. The spinors representing the electron transform under gauge-transformations like the quantum-mechanical wave-functions by the phase (5.3), and therefore this Lagrangian becomes gauge-invariant in the same way as in the quantum-mechanical case.

Besides the presence of photons and (massive) electrons and positrons another property of the theory can be read off the Lagrangian (5.7). 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 factor $i\gamma_\mu$ ensures that the vertex 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.

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.

5.4   Indistinguishable particles

The formulation in terms of the Lagrangian (5.7) 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.

5.5   Gauge symmetry

The gauge transformations (5.2-5.3) 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). Since the two-dimensional plane is equivalent to the complex plane, two-dimensional rotations

---

[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.

can also be seen as phase rotations, and the group of all phase rotations is called U(1). This equivalence is also expressed as SO(2)≈U(1). Since phase factors are more useful in quantum physics, the latter formulation is usually employed. 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.

Such structures will again reappear for the other parts of the standard model, and are also present in gravity. The precise formulation is mainly an exercise in group theory. A somewhat more formal discussion of this will be given below in section 6.1.

For comparison, gravity in the form of the general theory of relativity 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.

5.6 Bound states and the positronium

Already the simple system of QED in form of (5.7) 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, either parallel or anti-parallel. Since the life-time of these states is substantially different, they are also called differently, orthopositronium and parapositronium.

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

---

[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.

5.7 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 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 (5.7),

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \sum_f \bar{\psi}_f\gamma_\mu(i\partial^\mu - ieA^\mu - m_f)\psi_f, \qquad (5.8)$$

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. 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.

5.8 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, called radiation. 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, NLO, 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 not 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 (5.8) 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.

What is known is that whether a theory requires only a finite number of parameters can in most cases be deduced by simply looking at the Lagrangian. This is known as superficial power-counting. It can then be shown that, in most cases, a theory requires only a finite number of parameters if all numbers, masses and coupling constants in the Lagrangian have zero or positive dimension in terms of energy in natural units. Hence, QED is renormalizable. The notable exceptions to the rule will be encountered later. E. g., the Higgs theory was invented because without the Higgs particle the standard model would be non-renormalizable despite being superficially renormalizable. On the other hand there exist some theories, possibly including quantum gravity, which are superficially non-renormalizable, but in fact are renormalizable.

5.9   Fixing the inputs and running quantities

In the previous section 5.8, 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 (5.8) such as to reproduce this scattering cross-section, one of the parameters, α, 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 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, 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 in energy, 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 the four running quantities of QED 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. Usually, this choice

---

[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.

is also normalized such that in the absence of quantum corrections the classical coupling emerges, as done here when taking the Thompson limit. There is no necessity 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.

5.10 Landau poles

One of the mainstays of modern particle physics remains perturbation theory. However, as noted before, perturbative calculations can, strictly speaking, never be entirely correct, 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 aforementioned running of the coupling $\alpha(\mu)$. It shows that the running coupling satisfies the differential equation

$$\frac{de}{d\ln\mu} = \beta(e) = -\beta_0 \frac{e^3}{16\pi^2} + \mathcal{O}(e^5), \tag{5.9}$$

where β is the so-called β-function, and β_i are its Taylor coefficients. The latter can be calculated in perturbation theory. In fact, the value of the coefficients are known up to order 10 in α, i. e. 20 in e, in the expansion. Truncating the series at the lowest order, the differential equation can be integrated, and yields

$$\alpha(\mu^2) = \frac{e(\mu^2)^2}{4\pi} = \frac{\alpha(\mu_0^2)}{1 + \frac{\alpha(\mu_0^2)}{4\pi}\beta_0 \ln\frac{\mu^2}{\mu_0^2}} \equiv \frac{4\pi}{\beta_0 \ln\frac{\mu^2}{\Lambda^2}}, \tag{5.10}$$

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 detail later.

Chapter 6

Invariances and quantum numbers

6.1 Symmetries, groups, and algebras

As has already been seen in section 5.5, 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 not 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 Hamil-

ton 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 very few notions of group theory. 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 $T_2 T_1 |s\rangle$. 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, T^{-1}. 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 5.5. There, the group was SO(2)$\approx$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,

$$T_1 T_2 |s\rangle \neq T_2 T_1 |s\rangle.$$

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

The gauge field did not transform in the same way, see equation (5.2). 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 transforma-

tion. Any unitary transformation can be written as[1]

$$T(\alpha) = e^{i\alpha_i\tau^i}, \tag{6.1}$$

where the τ^i are Hermitian operators, and the possibility of a multi-parameter transformation has been included. 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.

A particular important special case arises, when the parameters α_i are (infinitesimally) small. The relation (6.1) can then be expanded to give the infinitesimal transformations

$$\begin{aligned} T &\approx 1 + i\alpha_i\tau^i \\ T^{-1} = T^\dagger &\approx 1 - i\alpha_i\tau^i. \end{aligned}$$

In many cases, it is sufficient to use such infinitesimal transformations to investigate the properties of a theory. However, this is not always possible, already in the standard model, though this will not be needed in this lecture. It just serves as a cautionary remark.

To illustrate the concept, the example of translational symmetry from quantum mechanics is useful. If a theory is invariant under translations, $x \to x + \Delta x$, a state changes under infinitesimal translations by

$$\psi(x + \Delta x) \approx \psi(x) + \Delta x\frac{d\psi(x)}{dx} = \psi(x) + i\Delta x\frac{d\psi(x)}{idx} = \psi(x) - i\Delta x\hat{p}\psi(x) \approx e^{-i\Delta x\hat{p}}\psi(x),$$

where $\hat{p}$ is the momentum operator. In this case, the momentum operator is the generator of translation, and the group elements are $e^{i\Delta x\hat{p}}$, forming the (Abelian) translation group. If the space is three-dimensional, the generators are the three spatial momenta $\hat{p}_i$, and the group elements become $\exp(i\Delta\vec{x}\cdot\hat{\vec{p}})$, a multi-parameter group.

6.2 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

---

[1]Note that this relation is in general only unique up to a discrete group.

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. E. g. a state of positive charge will never mix with a state of negative charge in the sense that under time evolution an electron will not change into a positron.

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. In this way, conserved currents are linked to symmetries of the theory. This will be encountered throughout the standard model of particle physics.

6.3   Electric charge

It is worthwhile to consider the consequences of this for QED. Applying Noether's theorem to the gauge symmetry yields that the electric charge is conserved. Thus, the gauge symmetry entails in the end the conservation of electric charge.

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 (5.2) 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. As 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.

6.4 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[2]. 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 should also 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.

6.5 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. $\vec{x} \to -\vec{x}$. 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

---

[2]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.

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. 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 (5.8), 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.

6.6   Time reversal

Of course, the corresponding partner to parity is the discrete time-reversal symmetry T, i. e. the exchange of $t \rightarrow -t$, without changes to the spatial coordinates. This is physically equivalent to reversing all speeds and momenta, and as a consequence also angular momenta: Either objects move backward or the time moves backward. Similarly to parity, fields can be classified as being even or odd under time-reversal.

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.

6.7 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. Note that properties like rest mass, momentum, spin and angular momentum 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 charge parity at all, which is particularly true for charged states. 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. Its 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.

6.8 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. 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 $J^{PC} = 1^{--}$. Of course, if any of the quantum numbers is not well-defined, e. g. the charge parity quantum number of particles with anti-particles of half-integer spin is not entirely trivial, it is omitted.

6.9 Symmetry breaking

Before continuing to further symmetries a brief intermission about breaking symmetries is necessary. Take a system 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 addition to the system does not obey the symmetry. Then, of course, already at the classical level the theory no longer 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 addition is such that it only takes effect in a certain energy regime, but becomes negligible in another one. The most common case will be seen to be masses, 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 the addition is quantitatively 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 sym-

metry 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 known in particle physics, though most of them are rather obscure.

If a symmetry exists 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 once more 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 the Curie temperature is an often cited example for this phenomenon. Again, there is some fine-print in a full quantum-field theory, but the simpler picture will suffice for this lecture.

6.10   Fermion number

An inspection of the QED Lagrangian (5.8) 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 unchanged. This is a continuous U(1) symmetry, and therefore a conserved quantity has to exist according to Noether's theorem. Calculating it, its is found to be proportional to the electric current. 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.

6.11 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 numbers of 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.

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.

6.12 Chiral symmetry

The QED Lagrangian (5.8) 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, essentially due to the Pauli principle, and there is no analogue for bosons of arbitrary spin

This adds an additional symmetry to the theory, which is called the axial symmetry. In addition, like the generalization of the fermion number symmetry to the flavor symmetry for N_f flavors, it is possible to enlarge the axial symmetry to an axial flavor symmetry, called 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[3], the name chiral, Greek for handedness, is assigned.

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-

---

[3]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 badly broken that it has almost no relevance for QED. This will change later when discussing the strong nuclear interaction.

Chapter 7

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 larger mass. Since also gravity was not an option, there had to be another force at work, 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 substructure. To understand the details was, however, a rather complicated challenge. While nuclei can be described using the effective nuclear interaction discussed in chapter 2, this is no longer possible when attempting to describe the substructure of protons and neutrons. This required a more fundamental theory.

The formulation of this fundamental theory in its currently final form is quantum chromo dynamics (QCD). It is, in a sense, the latest addition to the standard model. Because of the phenomenon of confinement, discussed below in section 7.8, it was much longer formulated than experimentally established. The reason behind is that the strong nuclear force is, as it name suggests, 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 reliably. Hence, 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.

7.1   Nuclei and the nuclear force

To motivate the development of QCD, it is useful to return once more to 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.

Experimentally, as discussed in section 2.4, the force follows a Yukawa form (2.3). Such a potential emerges in a QFT when it is mediated by exchanging a massive particle of mass of order m. This lead to the prediction of an exchange boson for the nuclear force with a mass of the order of roughly hundred MeV, given the range of the nuclear force of about a few fm. 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.

7.2   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 (2.3), 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 experiments.

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 anni-

hilation under emission of photons, this is somewhat awkward. A more elegant resolution of this will be given in the quark model below in section 7.4.

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, 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.

7.3 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. This can be considered as a kind of flavor symmetry, but only weakly 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 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 I, and then there are different charge states. 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 subsumed 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.

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, defining a so-called baryon number B, which is one for the nucleons and zero for the mesons, it is found that

$$Q = I_3 + \frac{B}{2}. \tag{7.1}$$

This is so far a phenomenological identification, but will become quite relevant in section 7.4. 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, they are commonly called 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.

7.4   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 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. In fact, similar to the case of the nuclei, the experiments showed that the number of constituents in baryons is three, and in mesons it is two. However, when trying to resolve the

scattering centers using more energy, it was found that more and more scattering centers where identified, called partons, if the nucleon is probed as large energies. This effect could not come from an interaction like (2.3), requiring a new type of interaction.

In addition, in contrast to the nuclei 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 constituents and put into a scheme like the periodic table of chemical elements. The zoo could originate from just a few 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 7.8.

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. These two types were denoted up and down type, in correspondence to their isospin values. 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, respectively. 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. 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.

However, not all particles fit into this simple model. The Δ turns out to pose a serious challenge.

7.5 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 containing a quark of each color, e. g. a proton contains a red and a blue up quark, and a green down quark. Similar, the Δ^{++} 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 be formulated in a mathematically well-defined way.

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 photon. 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 photon these are massless bosons with spin one. However, they differ from photons in a very important property. While photons are only mediating the 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 carried by gluons, and none of these eight are either the quark charges, nor is there 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 interaction of nucleons in a nuclei can be traced back to the gluons.

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, its quantum version quantum chromodynamics, or QCD for brief.

That this construction makes sense can also be read off from experiment. Consider the ratio

$$\frac{\sigma_{e^+e^-\to\text{Hadrons}}}{\sigma_{e^+e^-\to\mu^+\mu^-}}$$

as a function of the center-of-mass energy s. To leading order, the lepton pair is annihilated into a photon, which then splits either into two quarks or two leptons. Thus, the difference comes from the number of quarks. If there are three quark colors, this ratio will behave for not too large ($\sqrt{s} \lesssim 3.2$ GeV) energies as

$$3 \left(\frac{1}{9} + \frac{4}{9} + \frac{1}{9} \right) = 2.$$

Without color, this ratio, also known as R ratio, would be smaller by a factor of three. This is experimentally well confirmed.

7.6 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 it 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[1]. 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 dynamical effect of the strong interactions. To understand it, recall that massless free fermions have a chiral symmetry, see section 6.12. 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. However, it is possible that the strong interactions spontaneously break chiral symmetry. This possibility is explicitly confirmed. Since the breaking turns out to be soft, this explains why the direct measurements yields the smaller masses, as they are performed at higher energies.

The effect of the larger mass at low energies proceeds in the following (hand-waving) way. The gluon interaction creates a strong binding of the quarks, which creates a so-called vacuum condensate, that is the vacuum is filled with quark-antiquark pairs. Since the condensate is made from the quarks, any other quark will interact with it. Since the

---

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

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, 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.

7.7   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. The details are somewhat intricate, and will therefore be skipped. The final result is that whenever a symmetry is spontaneously broken there appears a certain number of massless particles, where the number is fixed by the type of the broken symmetry.

This would yield massless pions. The fact that they have a (small) mass comes from a corollary to Goldstone's theorem: If there is an additional explicit breaking of the symmetry, the massless states will gain a mass determined by this explicit breaking. In the case of QCD, this turns out to be true due to the current masses of the quarks, which are small. Therefore, pions are light, but not massless in QCD.

7.8   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 works very differently than electromagnetism or gravity. These two forces have the property that the potential 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 (2.3) 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

$$V = -\frac{4}{3}\frac{\alpha_s}{r} + \sigma r, \tag{7.2}$$

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 distance, 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.

This 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 behave as described, and are quite heavy. This interaction becomes weaker at higher energies, i. e. quarks moving through the quark condensate at high enough energies they will no longer be slowed down by it. The quarks will thus behave as if they only have their current mass. This explains the difference of both masses. The 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 (7.2), 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 (5.10) 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 $10^{15} - 10^{16}$ GeV.

The potential (7.2) also explains why no individual quarks are seen, at least to some extent. Since the potential energy between two quarks rises indefinitely 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 (7.2), this parameter is therefore also called string tension. This

explains 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 favorable 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. 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.

A consequence of the potential (7.2) is that for each hadronic state, there can be further excited states. These are the hadronic resonances, and exist for both baryons and mesons. 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.

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% of the cases 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 deexcite 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[2].

The study of the decays of hadrons has been very useful in the understanding of the strong force. As will be seen later in section 7.11, today decays of hadronic resonances have also become an important window into the study of other processes.

7.9   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

---

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

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, and many more. There is indeed no a-priori reason why such bound states should not exist. However, the are not yet any unequivocally observed tetraquarks or pentaquarks, though at least for tetraquarks and one pentaquark there is by now substantial circumstantial evidence available. That an experimental observation is so complicated can be motivated theoretically by two arguments.

The first effect is mixing. E. g. for a tetraquark, 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[3]. It is a light neutral meson, with quantum numbers compatible, e. g., 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 such states. The question which of these states contribute most is highly non-trivial. It can, in principle, be experimentally measured or theoretically 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 (very short-lived) 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.

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 it is thus challenging to establish their properties beyond doubt. Only very few candidates have been found so far, but some of them, to be discussed later in section 7.11, appear very promising.

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 add at most 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

---

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

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 combine only gluons to a colorless object. 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 their mass, 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 are only partially glueballs.

There are some glueball states which have quantum numbers which cannot be formed by only two or three 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.

7.10   Flavor symmetry and strangeness

So far two quark flavors, up and down, have been introduced. As discussed in section 6.11, this implies the presence of an approximate 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[4] 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 is about 400-450 MeV. It follows 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 8.10.

Besides the broken chiral symmetry, there is also the now enlarged flavor symmetry. Since the quarks have different masses, its is broken. Hence, the individual quark flavors are conserved, but bound states with differing quark content have differing masses. The 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.

---

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

7.11   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[5] quark. However, 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, chiral symmetry is 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 said to have hidden charm and are called charmonia. 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 therefore cannot 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. The reason is the so-called Zweig rule. It can be understood in the following way. When two quarks annihilate, this is a colorless state. Since gluons carry color, they cannot annihilate into a single gluon, though a single photon is possible. Two gluons is also not possible, since this state has not the correct quantum numbers in terms of the combination of spin, parity, and charge-parity. This is only possible starting with three gluons. But this requires three interactions. And though the interaction strength of

---

[5]The name originates from the fact that it solves several experimental mysteries observed in the weak interactions, to be discussed in chapter 8, and because at that time it appeared to complete the quark picture.

the strong interaction is larger than for QED, three interactions is still rare, and therefore not a substantial effect.

Hence, these hadrons are extremely stable compared to hadrons made from lighter quarks, where the pions offer a simpler decay channel. Thus, these meta-stable charmonia states 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 (7.2). Similarly to the hydrogen atom[6], 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 (7.2), 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 (7.2). 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 appears not to be able to explain all hadrons. These are candidates for tetraquarks. Recently, also a candidate for a pentaquark was found in this system.

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

---

[6]Spin-angular momentum couplings are much more relevant than for atoms, and have to be taken into account already at leading order.

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.

Other than the mass and the electric charge, the bottom quark behaves essentially as the charm quark. Especially, there is a rich spectroscopy with open and hidden beauty[7], the latter also called bottonium in analogy to charmonium. Similar to the case for the charm quark, the lightest mesons with open beauty are rather heavy, $B^{\pm}$ and B^0 at 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 up 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.

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 their production and decay take place so far apart from each other, a few μm, that both processes can be experimentally resolved. They and the mesons 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 8.5, 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.

7.12 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 is extended by the top (or in old texts truth) quark. This is the last quark, which so far has been

---

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

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 surprising 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 8. 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.

7.13　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 a certain symmetry group, the so-called SU(3) group. It is this symmetry structure which helped in the formulation of QCD, showing it to generalize QED.

Is it thus sufficient to just generalize the QED Lagrangian (5.7)? The answer to this is both yes and no. The quarks carry an additional color index like ψ^i. This requires to generalize the corresponding covariant derivative, yielding

$$D_\mu^{ij} = \delta^{ij}\partial_\mu - i\frac{g}{2}A_\mu^a(\tau^a)^{ij}, \tag{7.3}$$

where the A_μ^a are now the eight gluon fields, g is the coupling constant, and the τ^a are eight fixed 3×3 matrices.

Replacing the spinors by color vectors of spinors in the QED Lagrangian (5.8), and the flavors by the six quark flavors, is already sufficient to obtain the quark part of the QCD Lagrangian. However, it turns out that just taking the same field-strength tensor as in QED (5.6) with replacing the fields by their generalization with an index is not sufficient. In fact, in this case the Maxwell term (5.5) is not gauge-invariant.

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

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c, \qquad (7.4)$$

where the f^{abc} are 8^3 numbers uniquely related to the τ^a, and fully determined by the color symmetry. It is a straightforward, though tedious, calculation to show that with this replacement the Maxwell-term (5.5) becomes gauge-invariant. However, in contrast to QED, this field-strength tensor in itself is still not gauge-invariant, but $F_{\mu\nu}^a F_a^{\mu\nu}$ is.

QCD is therefore described by the Lagrangian

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu} + \sum_f \bar{\psi}_f^i \gamma_\mu (iD_{ij}^\mu - m_f)\psi_f^j. \qquad (7.5)$$

This theory has a number of further remarkable features, which can 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, just adorned by the τ^a matrices. Since the interaction potential between quarks (7.2) 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 (7.4) the coupling constant is fixed, as otherwise gauge invariance turns out not to be maintained. This feature, which arises from the underlying color symmetry, is called coupling universality, and has been experimentally verified.

Third, since the field strength tensor (7.4) 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 7.8, this was not expected. But this makes confinement a necessity. Still, this leaves open the question why confinement operates in the way it does, with the potential (7.2), and not in any other way.

Fourth, the appearance of a quadratic term in the field-strength tensor (7.4) 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 QCD, as this behavior is almost exclusive to theories similar to QCD. Because anti-screening dominates this leads to the difference between QED and QCD, especially the feature of asymptotic freedom discussed in section 7.8. In the terms of section 5.9, the coupling does not become stronger but weaker with energies.

Especially, it is possible to calculate the β function (5.9) also of QCD. To leading perturbative order only the first Taylor coefficient 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 (5.10), 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.

Because of this QCD at low energies is hard to treat. There are various methods available. The most successful one is arguably numerical simulations, which confirms that the QCD Lagrangian (7.5) 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.

7.14   The QCD phase diagram

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[8]. 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 together with the X-ray spectrum 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 gaseous and liquid 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 hadrons separated far by vacuum.

There are now several interesting directions to move on. Usually, they are signified by temperature and the baryon-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

---

[8]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 technical problems, but not a physics problem.

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, first not much happens up to the nuclear liquid-gas transition.

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 still 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 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 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. Heating up from zero, a rapid cross-over at a temperature of about 150-160 MeV occurs.

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 high-energy scatterings of the constituents. 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. At the same time, 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.

As a consequence of 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 transition to this behavior is very slow, and even at a few times the transition temperature not completed. The system then behaves then not like a gas, but rather like an almost ideal fluid.

This is also 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. Still, the point where it became possible to form stable nucleons is an important point, as this fixed the relative abundances of elements in the early universe. This process is called nucleosynthesis. The relative amount of nuclei created at this point, essentially only hydrogen, helium, lithium, and their isotopes, have been both observed and calculated. Both theory and experiment agree for most isotopes rather well.

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 8

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 sector 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 sector 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.

8.1 β decay and parity violation

The weak interaction surfaced for the first time in certain decays, particularly the β-decay discussed in chapter 2 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 8.2.

The first attempts of a theoretical formulation were based on a direct interaction of these four particles. The characteristic scale for the weak process was set by a new coupling, the Fermi constant, which is of order 1.14×10^{-5} GeV^{-2}.

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

---

[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.

violation.

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 one of the helicity states would be affected by the weak interactions. Experiment finds that only left-handed states are involved. Furthermore, the weak interactions are found to violate also the charge conjugation symmetry maximally.

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.

8.2   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 8.9, 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 that their mass is essentially below 0.2 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 many orders of magnitude 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 8.9. 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 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.

8.3 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 8.8.

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 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 lepton and the neutrino of each generation. This doublet structure has been already observed with the isospin relation (7.1), and it will be seen later how the electromagnetic charge enters this. Note, 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 they have been observed nonetheless.

8.4 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. 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 (surprisingly) unrelated.

Since weak charges are not confined, another mechanism is required to give the weak interaction a finite range. According to section 7.1, a possibility is that the charge carriers are massive. Another problem with this theory was that such a theory is unstable at large energies, see section 5.8. 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

$$G_F \approx \frac{g'^2}{M_W^2}, \tag{8.1}$$

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. That the $W^\pm$ carry electric charge indicate a relation to QED, which will be discussed in section 8.7.

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 7.7.

To understand that this is the right way, consider a photon A_μ coupled to some external current j. Going through the technical details, eventually a wave equation for the photon field is obtained,

$$\partial^2 \vec{A} = -\vec{j}. \tag{8.2}$$

It can now be shown that the solutions of this equation describe a massive photon of mass M if and only if[2]

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

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 - a time-independent medium filling all of space-time.

This is a screening process as can be seen from the following 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

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

---

[2]The details of this result are gauge-dependent, but the qualitative features remain in (most) other gauges.

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.

8.5   The Higgs effect and the Higgs boson

Motivated by the previous discussion, and by the considerations in section 7.7 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 Brout-Englert-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

$$
\begin{aligned}
\mathcal{L} &= \frac{1}{2}((\partial_\mu + iqA_\mu)\phi)^\dagger(\partial^\mu + iqA^\mu)\phi - \frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) \\
&\quad + \frac{1}{2}\mu^2\phi^\dagger\phi - \frac{1}{2}\frac{\mu^2}{f^2}(\phi^\dagger\phi)^2.
\end{aligned}
\tag{8.3}
$$

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 $\phi^\dagger\phi A_\mu A^\mu$ 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

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

This is just a reparametrization of the scalar field. 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 (8.3) yields

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

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

By technical manipulations[3] this takes the form

$$
\begin{aligned}
\mathcal{L} &= \frac{1}{2}\partial_\mu\rho\partial^\mu\rho + \frac{q^2}{4}A_\mu A^\mu\left(\frac{f}{\sqrt{2}}+\rho\right)^2 \\
&\quad -\frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) + \frac{1}{2}\mu^2\left(\frac{f}{\sqrt{2}}+\rho\right)^2 - \frac{1}{2}\frac{\mu^2}{f^2}\left(\frac{f}{\sqrt{2}}+\rho\right)^4 . \quad (8.4)
\end{aligned}
$$

The second term can be shown to provide exactly a current which screens, yielding an effective mass $qf/4$ for the photon field. Furthermore, if ρ is neglected, the Lagrangian would be just

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

This actually describes a massive photon.

Colloquial, this process is referred to as the breaking of the (gauge) symmetry. The precise details show that this is a misnomer, and hidden symmetry would be more correct. However, the previous terminology has become prevalent, and will therefore be used.

The Goldstone theorem of section 7.7 implies that for a given number of Higgs fields a certain number of gauge bosons will become massive, though this is not entirely trivial to show. Hence, if one desires a certain number of massive gauge bosons, this requires a pre-defined number of Higgs fields. Note that in contrast to section 7.7 no massless particles are created, but rather massless particles are turned into massive ones. The reason is that all the massless particles are combined into massive ones, which is also called pictorially that the Goldstone bosons are eaten by the gauge bosons.

This indicates how the weak gauge bosons will receive their mass. The necessary Higgs boson has by indeed now been discovered, and its mass is found to be about 125 GeV.

8.6 Parity violation and fermion masses

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

---

[3]Again a suitable gauge choice is made.

However, there is one serious flaw. This flaw originates from the combination of four things: The existence of fermions, the non-zero masses of these fermions, parity violation, and the requirement to have a gauge theory for the weak interaction for a consistent theory of massive vector bosons. For the known fermions, this combination is not possible in a consistent theory. Therefore, one of these has to be dropped.

Since massless fermions can be accommodated in the theory, it will be the masses. 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 rest.

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 a complicated interaction

$$\mathcal{L}_h = g_f \phi_k \bar{\psi}_i \left(\alpha_{ij}^k \frac{1 - \gamma_5}{2} + \beta_{ij}^k \frac{1 + \gamma_5}{2} \right) \psi_j + \left(g_f \phi_k \bar{\psi}_i \left(\alpha_{ij}^k \frac{1 - \gamma_5}{2} + \beta_{ij}^k \frac{1 + \gamma_5}{2} \right) \psi_j \right)^\dagger,$$

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 is fine with all other properties. 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.

Note that the quarks also have an additional contribution to their mass from the strong interactions, as discussed in section 7.6. In fact, the dominant contribution of the masses of hadrons is due to this effect, and not the Higgs, and in fact 99% of the mass in the universe is due to the strong interactions alone.

8.7 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 and fermions. 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 QED. 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.

8.7.1 Constructing the gauge group

Phenomenologically, as discussed in section 8.3, the weak interaction 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. Note that due to parity violation only left-handed fermions will be charged. 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 W, which changes one into the other, must carry a charge of one unit. The structure of gauge theories forces than the third, the Z, to be electrically neutral.

The quantum number distinguishing the two eigenstates of a doublet is called the third component of the weak isospin t_3. The weak interactions are therefore also called the weak isospin.

However, the weak gauge bosons are charged. Therefore, ordinary electromagnetic interactions have to be included also. It turns out that this cannot be directly ordinary electromagnetism, and actually electromagnetism will be an emergent phenomena. Rather, it requires first to introduce a new interaction, which is similar to QED, and which has a new quantum number, the so-called hypercharge. The ordinary electromagnetic charge is then given by

$$eQ = e\left(t_3 + \frac{y}{2}\right). \tag{8.5}$$

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 necessary to reproduce the experimental findings:

- Left-handed neutrinos: $t_3 = 1/2$, $y = -1$ ($Q = 0$)

- Left-handed leptons: $t_3 = -1/2$, $y = -1$ ($Q = -1$)

- Right-handed neutrinos: $t_3 = 0$, $y = 0$ ($Q = 0$)

- Right-handed leptons: $t_3 = 0$, $y = -2$ ($Q = -1$)

- Left-handed up-type (u, c, t) quarks: $t_3 = 1/2$, $y = 1/3$ ($Q = 2/3$)

- Left-handed down-type (d, s, b) quarks: $t_3 = -1/2$, $y = 1/3$ ($Q = -1/3$)

- Right-handed up-type quarks: $t_3 = 0$, $y = 4/3$ ($Q = 2/3$)

- Right-handed down-type quarks: $t_3 = 0$, $y = -2/3$ ($Q = -1/3$)

- W$^+$: $t_3 = 1$, $y = 0$ ($Q = 1$)

- W$^-$: $t_3 = -1$, $y = 0$ ($Q = -1$)

- Z: $t_3 = 0$, $y = 0$ ($Q = 0$)

- γ: $t_3 = 0$, $y = 0$ ($Q = 0$)

- Gluon: $t_3 = 0$, $y = 0$ ($Q = 0$)

- Higgs: Weak hypercharge $y = 1$ and two copies with $t_3 = \pm 1/2$. 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 antiparticle

This concludes the list of charge assignments for the standard model particles. 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.

Since at the present time the photon field and the Z boson are not yet readily identified, it is necessary to start out with a different set of particles, which will be denoted by W and B for the weak isospin and hypercharge, respectively. Both come with their own interaction strength, called g and g'.

8.7.2 Breaking the electroweak symmetry

To have a viable theory of the electroweak sector it is necessary to break the symmetry such that three gauge bosons become massive, and one uncharged one remains massless to become the photon. This is rather technical in detail. It requires the introduction of four Higgs fields, as noted above. Three of them will be eaten by three gauge bosons to give them mass using the Brout-Englert-Higgs mechanism as described in section 8.5.

Going through the details, it is ultimately found that the actually observed particles, the $W^\pm$, the Z, and the photon A are given by superpositions

$$
\begin{aligned}
A_\mu &= W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W \\
Z_\mu &= W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W \\
W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \pm iW_\mu^2).
\end{aligned}
$$

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

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

The masses can be calculated to be the experimentally observed W mass M_W and zero mass for the photon. Interestingly the mass of the Z is determined to be

$$
M_Z \cos\theta_W = M_W,
$$

and therefore not independent. According to the best measurements this effect is, however, only of order 15%. thus, the interaction determine the mass difference between the $W^\pm$ and the Z. Electromagnetism arose in the process, but the electric charge is uniquely given as

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

i. e., the observed electric charge is smaller than the hypercharge, and given by the original quantities. Electromagnetism is hence an emergent phenomena from a different theory. This is also the reason why sometimes this is called the electroweak unification.

8.8   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 neutrino 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 parameterizing 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 zero, while in the quark sector $\sin \theta_C$ is about 0.22. Corresponding observations are also made for other flavors. This also explains the inter-generation violation of flavor, the so-called Glashow-Iliopoulos-Maiani (GIM) mechanism, which incidentally predicted the charm quark.

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-violation and C-violation, it is necessary to include something into the interaction which provides this mixing. This can be done by considering instead of pure states (u, d, s) mixings

$$
\left(u \quad d \cos \theta_C + s \sin \theta_C \quad -d \sin \theta_C + s \cos \theta_C \right),
$$

and likewise for leptons. It should be noted that the up-type and down-type flavors of each generation are already mixed by the weak interactions themselves, since up-type and down-type are just the two different charge states of the weak isospin. In contrast to this intrageneration mixing, the effect here mixes different generations, and is hence an intergeneration effect, apart from the weak interaction.

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 in terms of angles of the mixing matrix, which is called the Cabibbo-Kobayashi-Maskawa (CKM) matrix. The matrix itself is given by

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}.$$

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

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

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

$$\begin{aligned} V_{CKM} &= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & e^{-i\delta_{13}}s_{13} \\ -s_{12}c_{23} - e^{i\delta_{13}}c_{12}s_{23}s_{13} & c_{12}c_{23} - e^{i\delta_{13}}s_{12}s_{23}s_{13} & s_{23}c_{13} \\ s_{12}s_{23} - e^{i\delta_{13}}c_{12}c_{23}s_{13} & -c_{12}s_{23} - e^{i\delta_{13}}s_{12}c_{23}s_{13} & c_{23}c_{13} \end{pmatrix}. \quad (8.6) \\ c_{ij} &= \cos\theta_{ij} \\ s_{ij} &= \sin\theta_{ij} \end{aligned}$$

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

$$
\begin{aligned}
\chi \;=\; & \sin\theta_{12}\sin\theta_{23}\sin\theta_{13}\cos\theta_{12}\cos\theta_{23}\cos^2\theta_{13}\sin\delta_{13} \qquad (8.7)\\
& \times(m_t^2-m_c^2)(m_t^2-m_u^2)(m_c^2-m_u^2)(m_b^2-m_s^2)(m_b^2-m_d^2)(m_s^2-m_d^2).
\end{aligned}
$$

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.

8.9   Particle 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 the standard model, the expression for the transition probability P 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 Cabibbo angle. If the probability, the energy and the distance is known for several cases, both the Cabibbo 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

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

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 occur if the involved particles have a mass. Thus, only with massive neutrinos there was a 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 not 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 heaviest. 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 from 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 (8.6), 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.

8.10   Anomalies

As noted in section 6.9, 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, part of the chiral symmetry described in section 6.12 is broken anomalously in the standard model. This breaking has 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 7.10, 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 of 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 gauge theories, and not for arbitrary ones, and only for particular assignments of charges to the fermions. However, this affected combination of charge assignments and theories 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. In 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). A second explanation is that only a theory like the standard model, with 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.

8.11   Theoretical problems

As has been noted, the three 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. The infinite rise could be countered, if there is indeed a GUT, like discussed in section 8.10, and then everything is well-behaved.

But 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[4]. 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[5].

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. 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 couplings would be taken, the Higgs mass 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

---

[4]It is assumed for simplicity here that all neutrinos are massive, though the experimental data are still compatible with one of them being massless.

[5]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 are no reasons for this assumption except technical ones. Hence, such statements express a hope rather than any reasonable physics.

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. It is found that in such a theory quantum effects always drive the self-interactions 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 5.8.

8.12 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 of 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$\sim 10^9$) amount.

8.13   The early universe

The weak interactions, and particularly 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. This implies that all particles have been massless in the early universe.

Chapter 9

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 desert.

9.1   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 for renormalization, as described in section 5.8: 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 thirtyfive 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 a structural difference of QED. 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 8.10. 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 8.8. 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 46 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?

9.2   Gravity

9.2.1   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 perturbatively not 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 roughly the Planck scale. 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

$$M_P = \frac{1}{\sqrt{G_N}} \approx 1.22 \times 10^{19} \text{ GeV}.$$

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.

9.2.2   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 arise. In fact, this includes the possibility that the Planck scale 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 improper limits above this scale.

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 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.

9.3 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.

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.

9.4   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.

9.4.1   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, especially nearby supernovas.

Also, there are some hints from stellar evolutions 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.

Finally, every once in a while, some unexpected structure is observed in the spectrum of cosmic rays. Any such structure could indicate processes involving unknown particles. However, these structures so far usually faded away once the observational techniques have been improved or astrophysical sources have been better understood, e. g. the influence of having an actual cosmical surrounding which on small scales is not isotropic.

9.4.2 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 and dynamics of galaxy clusters and the properties of the large-scale structures in the universe all support this finding. In fact, most (about 80%) 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, as well as gravitational lensing. 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 too low, and the flux of cosmic neutrinos too small, to make up even a considerable fraction of the dark matter.

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. This has lead to the idea that dark matter interacts with itself with a new force, which is stronger than just the gravitational interaction, but which does not couple to any standard model particles. In fact, it is also speculated that dark matter could consist out of many different particle species with many different forces.

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 de-
parture of gravity from its classical behavior also at long distances. Though parametriza-
tions 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.

9.4.3   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.

9.4.4   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,

---

[1]The homogeneity of the universe leads to a vanishing of the monopole moment and the dipole moment
originates from the observer's relative speed to the background.

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. This constant could also provide the remaining 70% of the mass to close the universe, and is in fact a vacuum energy. However, the known (quantum) effects contributing to such a constant provide a much too large value for Λ, about 10^{40} times too large. In addition, they have the wrong sign. What leads to the necessary enormous suppression is unclear.

9.4.5 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, the effect can only become relevant at very high temperatures.

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 ensures 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.

9.5   How can new physics be discovered?

A task as complicated as the theoretical description of new physics is its experimental observations. One of the objectives of theoretical studies is therefore to provide signals on which the experiments can focus on. Here, some of the more popular experimental signatures, which can be calculated theoretically, will be introduced. It should be noted that one can differentiate between generic signatures and predictions. A generic signature is a qualitative deviation of an experiment from the prediction of known physics. A prediction is a quantitative statement about an experiment's outcome. The latter are experimentally simpler, as the experiment can be focused on a particular target, and the required sensitivity is known a-priori. The disadvantage is that a null result will only invalidate this particular prediction, and cannot make a statement about any other, differing prediction. Generic signatures require an experiment to check all possible results for anomalies, what is already for combinatorial reasons a serious obstacle, and in practice limited by computational resources and availability of personnel. But it has the advantage that no deviation from known physics could escape, in principle, up to the sensitivity the experiment can provide.

An important point with respect to sensitivity is the statistical significance of an observation. Since the observed processes are inherently quantum, it is necessary to obtain a large sample (usually hundreds of millions of events) to identify something new, and still then an interesting effect may happening only once every hundredth million time. Experimentally identifying relevant events, using theoretical predictions, is highly complicated, and it is always necessary to quote the statistical accuracy (and likewise the systematic accuracy) for an event. Usually a three sigma (likeliness of 99.7%) effect is considered as evidence, and only five sigma (likeliness 99.9999%) are considered as a discovery, since in

the past several times evidence turned in the end out to be just statistical fluctuations.

Especially, a single event can only in the rarest of cases be a reliable signal. Usually, background processes will create signatures similar to the searched-for events. Thus, the true signal is the amount of events more than just from any (known) background source, and one has to take into account also the signal-to-background ratio S/B. Usually, this ratio is much smaller than one, and thus background suppression is necessary. The way how to suppress background is usually process-dependent, and this is therefore beyond the current scope.

Besides statistical errors, there are the much harder to quantify systematic uncertainties, e. g. the energy resolution of an experiment. There is no way to guarantee that a systematic error has been correctly estimated. Thus in the past at several occasions a too optimistic assumption has created a fake signal, which went away once systematic effects have been correctly accounted for. These problems have to be kept in mind when interpreting experimental results.

One of the best possibilities is, of course, to discover a new particle. This would be a unique signature for new physics, and is therefore the main focus of many searches. The second possibility is to look for the absence of something, especially of energy going into the collision. this missing energy would indicate the creation of a very weakly interacting new particle, e. g. a dark matter particle. However, neutrinos act very similarly, and are thus a problematic background for such searches.

A third possibility is the measurement of some quantities very precisely. Any deviation from the expected standard model value, if theoretically calculable with sufficient precision, is then indicating new physics. To identify the type and origin of such new physics, however, requires then careful theoretical calculations of all relevant models, and comparison with the measurement. Thus, a single deviation can usually only indicate the existence of new physics, but rarely unambiguously identify it. The advantage of such precision measurements is that they can usually be performed with much smaller experiments than collider experiments, but at the price of only very indirect information. Searches for dark matter or a neutron electric dipole moment larger than the standard model value are two examples of such low-energy precision experiments. But also collider experiments can perform precision measurements

There is the common expectation that something deviating from the standard model will show up in these measurements at an energy scale of one to a few TeV. One of the reason to expect this is that the Higgs plays a pivotal role to make the standard model consistent. If the Higgs should actually be only a particle (very) similar to the Higgs, this should show up at around 1 TeV. A second reason to suspect something new is the energy

scales so far encountered. In the standard model, so far the energy jumps between new particles is roughly an order of magnitude. There is no reason why this sequence should end now, so this lets it appear natural that something should happen at around 1 TeV. Finally, most theories which can accommodate several of the shortcomings of the standard model would require to have very finely tuned parameters to be not visible at the TeV scale. Though this is not impossible, there are very few examples known in nature where this occurs. This also motivates that new physics should be encountered at that scale. Especially, observations related to dark matter suggest that dark matter particles should have a mass in the TeV range, if they still interact with the standard model.

Hence, all in all, though there is no guarantee that something interesting beyond a rather light Higgs has to happen at the TeV scale, there is quite some motivation in favor of it. Time will tell. And what this something may be, the following will try to give a glimpse of. But it should always be kept in mind that all of the reasons, and all of the following proposals can be completely incorrect. After all, nature is what it is, and there is no logically compelling reasons in favor (and quite a number against) of nature being to be as expected. This was the lesson of the turn of the 19th to the 20th century. Be prepared for everything, and keep in mind that at most one of the following proposals can be correct in the presented form.

Chapter 10

Candidate theories beyond the standard model

10.1   Supersymmetry

One of the most favored extensions of the standard model is supersymmetry, mainly due to its elegance and that there is a natural way how to construct a supersymmetric extension of the standard model.

Supersymmetry is a symmetry which links the two basic types of particles one encounters in the standard model: bosons and fermions. In particular, it is possible to introduce operators which change a boson into a fermion and vice versa. Supersymmetric theories are then theories which are invariant under these transformations. This entails, of course, that such theories necessarily include both, bosons and fermions. In fact, it turns out that the same number of bosonic and fermionic degrees of freedom are necessary to obtain a supersymmetric theory. Hence to each particle there should exist a superpartner. Superpartners of bosons are usually denoted by the ending ino, so for a bosonic gluon the fermionic superpartner is called gluino. For the superpartners of fermions an s is put in front, so the super-partner of the fermionic top quark is the bosonic stop (and for quarks in general squarks).

Besides the conceptual interest in supersymmetric theories there is a number of phenomenological consequences which make these theories highly attractive. The most interesting of these is that quite a number of divergences, which have to be removed in ordinary quantum field theories by renormalization as discussed in section 5.8, drop out automatically. From this cancellation it is also possible to obtain a 'natural' explanation why scalar particles should be of about the same mass as the other particles in a theory. There are some other reasons why supersymmetric theories are interesting:

- There exists a theorem that without supersymmetry it is not possible to embed the standard model (Lie groups) and general relativity (Poincare group) into a non-trivial (i.e. any other than a direct product) common structure (Coleman-Mandula theorem)

- Supersymmetric theories arise necessarily in the context of string theories, candidate theories for quantum gravity to be discussed later

- Supersymmetry provides a simple possibility to extend the standard model

- Supersymmetry can be upgraded to include also gravity (called supergravity)

- Supersymmetry predicts relations between coupling constants which seem without relation without supersymmetry

- Supersymmetry provides such stringent constraints that many results can be shown exactly

In particular, it is widely expected that supersymmetry should be manifest at the 1-10 TeV scale, as this is the most likely range for the simplest extension of the standard model such that one ends up with the known particles with the right properties. However, there is no necessity for it.

Furthermore, there are also several reasons which make supersymmetry rather suspect. The most important one is that supersymmetry is not realized in nature. Otherwise the unambiguous prediction of supersymmetry would be that for every bosonic particle (e. g. the photon) a particle with the same mass, but different statistics (for the photon the spin-1/2 photino) should exist, which is not observed. The common explanation for this is that supersymmetry in nature has to be broken either explicitly or spontaneously. However, how such a breaking could proceed such that the known standard model emerges is not known. It is only possible to parametrize this breaking, yielding a enormous amount of free masses and coupling constants, for the standard model more than a hundred, while the standard model has only about thirty.

As a side remark of this introduction it should be mentioned that supersymmetry offers also technical methods which can be used even if the underlying theory itself is not supersymmetric (which is done, e. g., in nuclear physics). So the use of supersymmetric concepts extends far beyond just possible extensions of the standard model.

To show the enormous complexity of supersymmetric theories, consider a supersymmetric version of the standard model. This requires to construct for each particle in the standard model a new superpartner. In fact, also completely new particles are actually

required. Here, the supersymmetric version of the standard model with the least number of additional fields will be introduced, the so-called minimal supersymmetric standard model (MSSM). This requires the following new particles

- The photon is uncharged. Therefore, its superpartner has also to be uncharged, and to be a fermion. It is called the photino

- The eight gluons require eight fermions carrying the same color charges, which are called gluinos

- Though massive, the same is true for the weak bosons, leading to the charged superpartners of the W-bosons, the winos, and of the Z-boson, the zino, together the binos. Except for the photino all gauginos, the gluinos and the binos, interact with the original gauge-fields

- The superpartner of the leptons, the sleptons, are, called the sneutrinos, the selectron, the smuon, and the stau

- The same applies to quarks, requiring the squarks

- The Higgs requires a fermionic superpartner, the higgsino. However, requiring supersymmetry forbids that the Higgs has the same type of coupling to all the standard model fields. Therefore, a second set of Higgs particles is necessary, with their corresponding superpartners. These are the only new particles required

Since no mechanism is yet known how to generate supersymmetry breaking in a way which would be in a accordance with all observations, and without internal contradictions, this breaking is only parametrized in the MSSM, and requires a large number of additional free parameters. Together with the original about 30 parameters of the standard model, these are then more than a 130 free parameters in the MSSM. Exactly, there are 105 additional parameters in the MSSM, not counting contributions from massive neutrinos.

These parameters include masses for all the non-standard model superpartners, that is squarks and sleptons, as well as photinos, gluinos, winos, and binos, except for the Higgsinos. it is not possible to construct a gauge-invariant additional mass-term, due to the chirality of the weak interactions, in much the same way as for quarks and leptons in the standard model. One advantage is, however, offered by the introduction of these free mass parameters: It is possible to introduce a negative mass squared for the Higgs particles, reinstantiating the same way to describe the breaking of electroweak symmetry as in the standard model. That again highlights how electroweak symmetry breaking

and supersymmetry breaking may be connected. These masses also permit to shift all superpartners to such scales as they are in accordance with the observed limits so far. However, if the masses would be much larger than the scale of electroweak symmetry breaking, i. e., 250 GeV, it would again be very hard to obtain results in agreement with the observations without fine-tuning. However, such mass-matrices should not have large off-diagonal elements, i. e., the corresponding CKM-matrices should be almost diagonal. Otherwise, mixing would produce flavor mixing also for the standard model particles exceeding significantly the observed amount.

The arbitrariness of these enormous amount of coupling constants can be reduced, if some model of underlying supersymmetry breaking is assumed. One, rather often, invoked one is that the minimal supersymmetric standard model is the low-energy limit of a super-gravity theory. In this case, it is, e. g., predicted, that the masses of the superpartners of the gauge-boson superpartners should be degenerate, and also the masses of the squarks and sleptons should be without mixing and degenerate and these masses also with the ones of the Higgs particles. Furthermore, all trilinear bosonic couplings would be degenerate. If the phase of all three parameters would be the same, also CP violation would not differ from the one of the standard model, an important constraint.

Of course, as the theory interacts, all of these parameters are only degenerate in such a way at the scale where supersymmetry breaks. Since the various flavors couple differently in the standard model, and thus in the minimal supersymmetric standard model, the parameters will again differ in general at the electroweak scale, or any lower scale than the supersymmetry breaking scale.

This simplest form of the MSSM is by now ruled out by experiment. To further permit a supersymmetric extension of the standard model requires to relax the assumptions. This can be either done by relaxing the supergravity constraints on the parameters, or by introducing additional new particles to rescue the supergravity paradigm. Either way introduces at the current time more unknowns, and further experimental data will decide whether this is he right way to go or whether supersymmetry is not part of nature, at least at energies not too large compared to the electroweak scale.

Note that one big advantage of the MSSM, and other supersymmetric extensions, is that it is almost universally possible to have automatically a suitable dark matter particle, predicted. This is a mixture of various supersymmetric particles, called the neutralino, or also LSP for lightest supersymmetric particles, as to be stable it is required to be light enough.

10.2 Technicolor

Technicolor is the first prototype theory for compositness approaches. The idea is that the hierarchy problem associated with the mass of the Higgs boson can be circumvented if the Higgs boson is not elementary but a composite object. If its constituents in turn are made of particles with masses which do not suffer from a hierarchy problem, in particular fermions, then the hierarchy problem simply would not exist.

However, such models require that interactions are non-perturbative, such that the Higgs can be a bound state. It would, as atoms, appear as an elementary particle only on energy scales significantly below the binding energy.

Such a construction is actually rather intuitive, and even realized in the standard model already. In QCD, bound states of quarks occur which actually have the same quantum numbers as the Higgs. In fact, already within QCD condensates with the quantum numbers of the Higgs condensate can be constructed, which induce the breaking of the electroweak symmetry. Only because the size of such condensates is then also given by the hadronic scale, and thus of order GeV, this is not sufficient to provide quantitatively for the observed electroweak symmetry breaking. Qualitatively it is the case.

Thus, the simplest extension is to postulate a second, QCD-like theory with typical energy scales in the TeV range with bound states which provide a composite Higgs, in addition to the standard model. Such theories are called technicolor theories.

Technicolor theories are also prototype theories for generic new strong interactions at higher energy scales, since at low energies they often differ from technicolor theories only by minor changes in the spectrum and concerning interaction strengths. Also, most of these suffer from similar problems as technicolor. Studying technicolor is therefore providing rather generic insight into standard model extensions with strongly interacting gauge interactions above the electroweak scale.

Actually implementations of the technicolor idea tend, however, to be rather complex, similar to supersymmetry, if they should agree with all experimental results. Since they are theoretically harder to control, due to the strong interactions and non-perturbative features, there is much less known for these theories than for the supersymmetric theories.

Generic features of technicolor theories is that besides the Higgs usually many more, heavier, bound states appear, similar to the hadrons of QCD. Thus, these theories provide a large number of predicted particles, usually also around the TeV scale. In addition, also dark matter candidates are found in this type of theories.

Note that a particular cumbersome challenge for technicolor is the large mass difference between neutrinos and the top, which in the standard model is created by the Higgs, and has here to be covered by non-perturbative effects. This is still a very active area of

research. For supersymmetry, where these masses are parameterized, this problem does not exist, but neither does it explain the masses, as technicolor attempts to do.

10.3   Other low-scale extensions of the standard model

While the previous two candidate theories are the most popular ones, there is an enormous amount of other possibilities. Here, only a few of the more relevant ones will be briefly mentioned.

Little Higgs theories construct situations which makes the Higgs similar to the pions of QCD, but without requiring to make them composite. This usually induces further Higgs particles in the spectrum, but also solves the hierarchy problem.

Hidden sectors introduce the idea that a whole non-trivial particle physics could exist besides our own, but any communication is extremely weak, since exchanged messenger particles are very heavy. Such a hidden sector is of course also a very good candidate for dark matter. If the exchange is only by the interaction of the hidden sector with the Higgs, these are also called Higgs portal models.

The idea of flavons decouples the mass generation for the weak gauge bosons and the fermions by introducing further Higgs fields, possibly in a hidden sector, which take care of the mass generation for the Higgs. This removes a completely coincidental appearing relation in the standard model, which allows to find different explanations for both phenomena.

See-saw mechanisms attempt to explain the extremely small mass of neutrinos by showing that neutrinos could actually be mixtures of two 'normal-weighted' particles, which then appear like a very light and a very heavy, yet undetected, other particle. This is sometimes also used for other particles, especially in context with the top mass.

Axions and milli-charged particles follow the idea that it is possible to include further particles into the standard-model, which interact weakly with it, without introducing new forces. Due to the structure of the standard-model, this is only possible with particles which have very small electric charge, much smaller than ordinary particles, or couple very differently to the standard model interactions than the other particles. Of course, such particles could also be dark matter candidates.

There is a generic trait for many BSM scenarios: The appearance of additional Higgs-like particles, being them elementary or composite. All of these models have a very similar low-energy behavior, essentially the standard model with an additional set of four more Higgs-like particles. This whole class of models is hence known as 2-Higgs doublet models. The name of doublet emerges from the fact that actually already in the standard model

there are four Higgs particles, but three of them join with the W and Z to create their mass. And of course, these include any models which have this structure from the outset. The different high-energy theories manifest then themselves in the various parameters of this extended Higgs sectors. Thus, such theories are heavily studied.

10.4 Grand unified theories

As outlined before, the fact that the electromagnetic couplings have small ratios of integers for quarks and leptons cannot be explained within the standard model. However, this is necessary to exclude anomalies, and make the standard model a valid quantum field theory. Thus, this oddity suggests that possibly quarks and leptons are not that different as the standard model suggests. The basic idea of grand unified theories is that this is indeed so, and that at sufficiently high energies an underlying symmetry relates quarks and leptons, making these ratios of electric charges natural. This is only possible, if the gauge interactions are all together embedded into a single gauge interaction, since otherwise this would distinguish quarks from leptons due to their different non-electromagnetic charges. Another motivation besides the electromagnetic couplings for this to be the case is that the running couplings, the effective energy dependence of the effective gauge couplings, of all three interactions almost meet at a single energy scale, of about 10^{15} GeV. They do not quite, but if the symmetry between quarks and leptons is broken at this scale, it would look in such a way from the low-energy perspective. If all gauge interactions would become one, this would indeed require that all the couplings would exactly match at some energy scale.

These arguments are the basic idea behind grand unified theories (GUT). Since there are very many viable options for such grand-unified theories, all of which can be made compatible with what is known so far, there is no point as to give preference of one over the other, but instead just to discuss the common traits. Also, GUT ideas are recurring in other beyond-the-standard model scenarios. E. g., in supersymmetric or technicolor extensions the required new parameters are often assumed to be not a real additional effect, but, at some sufficiently high scale, all of these will emerge together with the standard model parameters from such a GUT. In these cases the breaking of the GUT just produces further sectors, which decouple at higher energies from the standard model. Here, the possibility of further sectors to be included in the GUT will not be considered further.

The basic idea is that such a GUT is as simple as possible, though even the simplest examples, which are not yet ruled out by experiment, are also quite complex, and usually

require several additional, but very heavy, particles.

It should be noted that the idea of GUTs in this simple version, i. e., just by enlarging the gauge group, cannot include gravity, due to the aforementioned Coleman-Mandula theorem.

The most interesting consequences of such theories is that they also yield additional baryon number violation, as required from cosmology. In fact, they often yield so much as to be incompatible with the proton lifetime. Thus, measuring the proton lifetime is so far one of the most sensitive tests of GUTs.

10.5   At the Planck scale

In the following some possibilities will be briefly presented how gravity as a quantum theory could become part of the standard model.

10.5.1   Large extra dimensions

One attempt to remedy the large difference in scale between gravity and the standard model is by the presence of additional dimensions. Also, string theories, to be discussed later, typical require more than just four dimensions to be well-defined. Such extra dimensions should not be detectable. The simplest possibility to make them undetectable with current methods is by making them compact, i. e., of finite extent. Upper limits for the extensions of such extra dimensions depend strongly on the number of them, but for the simplest models with two extra dimensions sizes of the order of micrometer are still admissible. These large extra dimensions are contrasted to the usually small extensions encountered with gravity, which are of the order of the Planck length. For now, start with those with large extension. These can also be considered as low-energy effective theories of, e. g., string theories.

These models separate in various different types. One criterion to distinguish them is how the additional dimensions are made finite, i. e. how they are compactified. There are simple possibilities, like just making them periodic, corresponding to a toroidal compactification, or much more complicated ones like warped extra dimensions. The second criterion is whether only gravity can move freely in the additional dimensions, while the standard model fields are restricted to the uncompactified four-dimensional submanifold, then often referred to as the boundary or a brane in contrast to the full space-time called the bulk, or if all fields can propagate freely in all dimensions.

One thing about these large extra dimensions is that they can also be checked by tests of gravity instead of collider experiments. If there are $4 + n$ dimensions, the gravitational

force is given by

$$F(r) \sim \frac{G_N^{4+n} m_1 m_2}{r^{n+2}} = \frac{1}{M_s^{n+2}} \frac{m_1 m_2}{r^{n+2}},$$

where G_N^{4+n} is the $4 + n$-dimensional Newton constant and correspondingly M_s the $4+n$-dimensional Planck mass. If the additional n dimensions are finite with a typical size L, then at large distances the perceived force is

$$F(r) \sim \frac{1}{M_s^{n+2} L^n} \frac{m_1 m_2}{r^2} = \frac{G_N m_1 m_2}{r^2},$$

with the four-dimensional Newton constant G_N. Thus, at sufficiently long distances the effects of extra dimensions is to lower the effective gravitational interactions by a factor of order L^n. That can explain why gravitation is perceived as such a weak force at sufficiently long distances.

On the other hand, by measuring the gravitational law at small distances, deviations from the $1/r^2$-form could be detected, if the distance is smaller or comparable to L. This is experimentally quite difficult, and such tests of gravity have so far only been possible down to the scale of the order of a little less than a hundred μm. If the scale M_{4+n} should be of order TeV, this excludes a single and two extra dimensions, but three are easily possible. Indeed, string theories suggest n to be six or seven, thus there are plenty of possibilities. In fact, in this case the string scale becomes the $4 + n$-dimensional Planck scale, and is here therefore denoted by M_s. The following will discuss consequences for particle physics of these extra dimensions.

One possibility are so-called separable gravity-exclusive extra dimensions. These have n additional space-like dimensions. Furthermore, these additional dimensions are taken to be separable so that the additional dimensions do not mix with the ordinary ones in the space-time structure. Furthermore, for the additional dimensions to be gravity exclusive the other particles have to be restricted to the 4-dimensional brane of uncompactified dimensions.

Mathematically, this results in an infinite number of copies of the ordinary particles, but with larger masses. These are called the Kaluza-Klein masses[1].

A serious problem arises when the universally coupling Kaluza-Klein modes show up in processes forbidden, or strongly suppressed, in the standard model, like proton decay. The standard model limit for proton decay is about 10^{15} GeV, thus much larger than the comparable effect from the larger extra dimensions if M_s should be of order TeV. Thus this leads to a contradiction if not either M_S (or n) is again set very large, and thus large

---

[1]Originally, Kaluza and Klein in the 1930s aimed at unifying classically gravity and electromagnetism by introducing a fifth dimension, which failed.

extra dimensions become once more undetectable, or additional custodial physics is added to this simple setup. The latter usually leads, like in the case of technicolor, to rather complex setups.

The alternative to gravity-exclusive extra dimensions are such which are accessible to all fields equally, so-called universal extra dimensions. This implies that the theory is without compactification like the conventional theory, just in more (space) dimensions. As a consequence, such theories can in general not resolve the hierarchy problem. However, they provide possibilities how anomalies can be canceled, e. g. in six dimensions, without need to assign specific charges to particles. In addition, one of the Kaluza-Klein modes can often serve as a dark matter candidate. On the other hand, since particle physics has been tested to quite some high energy with no deviations observed, this imposes severe restrictions on the size of extra dimensions, being usually of order inverse TeV, and thus sub-fermi range, rather than μm.

In models with warped extra dimensions, also known as Randall-Sundrum models, the additional dimensions have an intrinsic curvature in such a way that the energy scales depend exponentially on the separation Δy of two points in the additional dimensions, $\exp(-2\Delta y k)$. By positioning different effects at different relative positions, large scale differences can appear naturally. In particular, the different masses for the standard model fermions can be explained by having different localizations of the different particles in the extra dimensions. Such compactifications, depending on details, are also often referred to as orbifolds.

An alternative flavor of (large) extra dimensions are obtained from so-called deconstructed extra dimensions. In this case the extra dimensions are not continuous, but are discrete, i. e., contain only a finite number of points. This can also be viewed by a finite, in case of the extra dimension being compactified, or infinite set of four-dimensional space-times, which are distinguished by a discrete quantum number. As an example, take only one additional dimension, with N points and radius R. Then each of the N points is a complete four-dimensional space-time, and is also called a brane.

A rather popular possible signature for large extra dimensions are the production and decay of black holes. The Schwarzschild radius of a $4 + n$-dimensional black hole for n compact dimensions characterized by the $4 + n$-dimensional Planck scale M_s is given by

$$R_B \sim \frac{1}{M_s} \left(\frac{M_B}{M_s} \right)^{\frac{1}{n+1}},$$

with the black hole mass M_B. If in a high-energy collisions two particles with center-of-mass energy s larger than M_s^2 come closer than R_B, a black hole of mass $M_B \approx s$ is

formed. The cross-section is thus essentially the geometric one,

$$\sigma \approx \pi R_B^2 \sim \frac{1}{M_s^2} \left(\frac{M_B}{M_s} \right)^{\frac{2}{n+1}} .$$

It therefore drops sharply with the scale M_s. However, its decay signature is quite unique. It decays by the Hawking process, i. e., by the absorption of virtual anti-particles by the black hole, making their virtual partner particles real. The expectation value for the number of particles for the decay of such a black hole is

$$\langle N \rangle \sim \left(\frac{M_B}{M_s} \right)^{\frac{n+2}{n+1}} ,$$

and therefore rises quickly when the energies of the colliding particles, and thus the mass of the produced black hole, significantly exceeds the scale of the compactified dimensions.

10.5.2   Quantum gravity

In the following a short look will be taken at how a quantum theory of gravity can be obtained pursuing the normal way of covariantly quantizing it. This is difficult, since a consistent theory of quantum gravity is very hard to obtain. Because the Newton constant appearing is of dimension of inverse energy any perturbative way of quantizing the theory will produce a non-renormalizable theory, having little predictivity. Therefore, ultimately, a genuine non-perturbative approach will be necessary.

Quantum gravity as a direct quantized version of general relativity is also not the only proposed possibility of how to obtain classical gravity in the long-distance limit of everyday life and of stellar or interstellar objects. Other approaches include supergravity theories, string theories, or non-commutative geometry. All of them have their merits and drawbacks, but all are also measured in comparison to quantum gravity.

It should be noted, however, that there is no unique way yet known of how to quantize gravity, i. e. on which quantities e. g. canonical commutation relations should be imposed. This is actually not surprising. Even for ordinary quantum mechanics, there is no unique choice. It was rather the comparison to experiment which eventually decided what was the correct way to proceed. Quantization rules are postulates, not derived.

Given that there are currently no experimentally reliable statements about quantum effects in gravity, this option is not yet available. Hence, the only possibility is to quantize the theory, and obtain its low-energy limit, which must be classical general relativity, for which plenty of experimental and observational results exist.

As a consequence, several different proposals are currently pursued. One is the canonical quantization approach, which essentially just continues quantum mechanics. The

consequence of this quantization procedure is that the coordinates loose their meaning in the usual way. Since their values are subject to quantum fluctuations of the metric, it is no longer possible to assign to them an independent reality, as has been done in the classical theory. Thus the quantum uncertainty not only afflicts the particles but space-time as such: Not only it becomes impossible to assign a position to a particle, but it becomes impossible to give the term position a meaning. Thus, even if particles could be assigned a position, the position concept itself has now a quantum uncertainty.

Though the quantization procedure is in itself, irrespective of the particular version chosen, straightforward, the resulting quantum theory is not.

If quantized canonically, the major obstacle is that the theory cannot be treated perturbatively, as noted already in section 9.2.1. A possible solution is that this is only a technical problem, and the theory is treatable once non-perturbative methods are employed. This is the concept of asymptotic safety, discussed in section 9.2.2.

In other versions of quantization this is not necessarily true. One particular popular choice is the so-called loop quantum gravity approach. In this case, the quantization is not applied to the metric, itself, but rather to more complicated quantities, which are certain closed line integrals of the metric, therefore the reference to loops. It appears that this quantization procedure leads to a perturbatively more tractable theory, but the inherent non-localities when defining the dynamical variables have so far limited the progress.

One further possibility to quantize gravity is to postulate the existence of a minimal length, similar to the postulate of a minimal phase space volume $\Delta x \Delta p \sim \hbar$ in ordinary quantum mechanics. This is also similar to the idea of a maximum speed in general relativity. As there, the existence of such a minimal length, which is typically of the order of the Planck length 10^{-20} fm, has profound consequences for the structure of space-time. Especially, coordinate operators do no longer commute, just like coordinate and momenta do not commute in quantum mechanics. Thus, this ansatz is called non-commutative geometry. Since there is a minimal length, there is also a maximal energy, and hence all quantities become inherently finite, and renormalization is no longer necessary. On the downside of this approach, besides an enormous increase in technical complexity, is that in general relativity neither coordinates nor energies themselves are any longer physical entities, like in special relativity or in quantum (field) theories. Thus, the precise physical interpretation of a non-commutative geometry is not entirely clear. Furthermore, so far it was not possible to establish a non-commutative theory which, in a satisfactory manner, provides a low-energy limit sufficiently similar to the standard model. Particularly cumbersome is that it is very hard to separate the ultraviolet regime where the non-commutativity becomes manifest and the infrared, where the coordinates should again

effectively commute.

10.5.3 String theory

A complete alternative to the quantization of conventional gravity is string theory. Here, the problems encountered when attempting to quantize gravity are transcended by using a conceptual new form of theory, which only in the low-energy limit takes the form of an ordinary quantum theory. In these theories the concept of an elementary particle, i. e. point particles, are replaced by strings.

However, most string theories have a very complicated structure, and it is particularly non-trivial to understand the low-energy limit. This problem is made more severe by the fact that a given string theory can have multiple low-energy limits, which are hard to distinguish, the so-called landscape problem. Furthermore, string theories are technically complicated, especially when treated consistently. Hence, here only a brief glimpse can be given of their properties. Especially, any string theory which seems to be able to produce nature requires at some scale a (broken) supersymmetry, leading to superstring theories. Finally, all string theories found so far are related to each other. Furthermore, it turns out that they all need common objects, which are of higher-dimensional types than strings, the already alluded to branes. The description of branes and strings is conjectured to be unifiable in a single theory, the so-called matrix theory, or M-theory for short. Until now, it was not possible to formulate this theory.

String theories have a number of rather generic properties, like the natural appearance of gravitons, the need for additional dimensions, and problems due to tachyons, which appear already in their simplest form

This already shows, why string theories are particularly attractive: The natural appearance of the graviton makes string theories rather interesting, given the intrinsic problems of quantum gravity. Further advantages of more sophisticated string theories are that they have generically few parameters, are not featuring space-time singularities, such as black holes, on a quantum level, and often have no need for renormalization, thus being consistent ultraviolet completions. The price to be paid is that only rather complicated string theories even have a chance to resemble the standard model, their quantization beyond perturbation theory is not yet fully solved, and it is unclear how to identify a string theory which has a vacuum state which is compatible with standard model physics, again the so-called landscape problem due to enormous number of (classical) vacuum states in string theories. Furthermore, in general genuine string signatures usually only appear at energy levels comparable to the Planck scale, making an experimental investigation, and thus support, of stringy properties of physics, almost impossible with the current technology.

How comes the idea of string theory about? It is motivated by how perturbation theory fails in gravity. Generically, the understanding of physics has been increased by looking at ever shorter distances and at ever high energies. The current confirmed state of affairs is then the standard model. Going back to quantum gravity, a similar insight can be gained. In the perturbative approach, the ratio of free propagation to the exchange of a single graviton behaves as

$$\frac{A_{\text{free}}}{A_{1g}} = \frac{\hbar c^5}{G_N E^2} = \frac{M_P^2}{E^2}$$

where $M_P^2 = \hbar c^5/G_N$ is again the Planck mass, this time in standard units. Since M_P is once more of the order of 10^{19} GeV, this effect is negligible for typical collider energies of TeV. However, if the energy becomes much larger than the scale, the ratio of free propagation to exchange of a graviton becomes much smaller than one, indicating the breakdown of perturbation theory. This is not cured by higher order effects, but actually becomes worse.

The basic idea behind string theory is to try something new. The problem leading to the divergence is that with ever increasing energy ever shorter distances are probed, and by this ever more gravitons are found. This occupation with gravitons is then what ultimately leads to the problem. The ansatz of string theory is then to prevent such an effect. This is achieved by smearing out the interaction over a space-time volume. For a conventional quantum field theory such an inherent non-locality usually comes with the loss of causality. String theories, however, are a possibility to preserve causality and smear out the interaction in such a way that the problem is not occurring.

However, the approach of string theory actually goes (and, as a matter of fact, has to go) a step further. Instead of smearing only the interaction, it smears out the particles themselves. Of course, this occurs already anyway in quantum physics by the uncertainty principle. But in quantum field theory it is still possible to speak in the classical limit of a world-line of a particle. In string theory, this line becomes a sheet. As noted, string theories can also harbor world-volumes in the form of branes. One of the problems with a theory of such world-volumes is that internal degrees of freedom are also troublesome, and can once more give rise to consistency problems. Thus, these will be neglected here. String theory, on the other hand, seems to be singled out to be the theory with just enough smearing to avoid the problems of quantum field theory and at the same time having enough internal rigidity as to avoid new problems. Though without branes it does not appear to be able to describe nature.

One feature of string theory is that there is usually no consistent solution in four space-time dimensions, but typically more dimensions are required. How many more is actually a dynamical property of the theory: It is necessary to solve it to give an answer. In

perturbation theory, it appears that ten dimensions are required, but beyond perturbation theory indications have been found that rather elven dimensions are necessary. Anyway, the number is usually larger than four. Thus, some of the dimensions have to be hidden, which can be performed by compactification, as with the setup for large extra dimensions. Indeed, as has been emphasized, large extra dimensions are rather often interpreted as a low-energy effective theory of string theory.

Since the space-time geometry of string theory is dynamic, as in case of quantum gravity, the compactification is a dynamical process. It turns out that already classically there are a huge number of (quasi-)stable solutions having a decent compactification of the surplus dimensions, but all of them harbor a different low-energy physics, i. e., a different standard model. To have the string theory choose the right vacuum, thus yielding the observed standard model, turns out to be complicated, though quantum effects actually improve the situation. Nonetheless, this problem remains a persistent challenge for string theories.