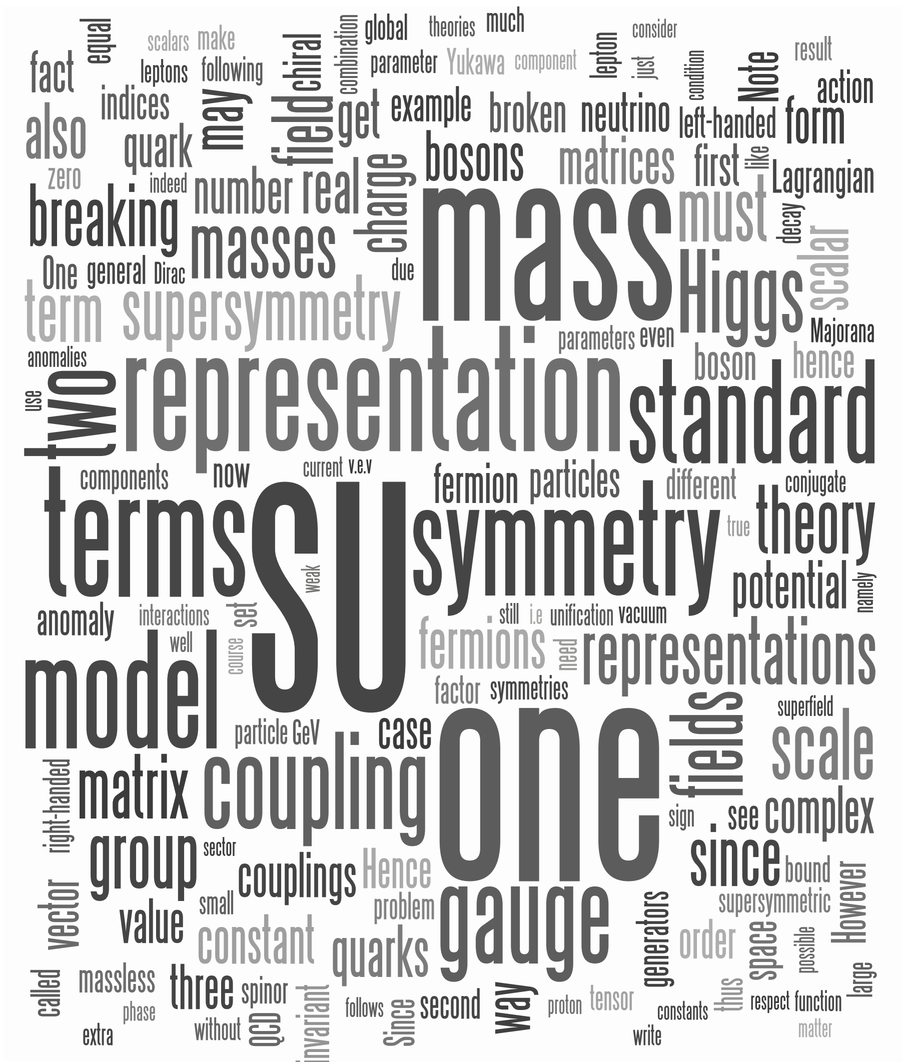
Beyond the Standard Model

A.N. Schellekens



[Word cloud by www.worldle.net]
Last modified 12 June 2017

Contents

Preface

The first version of these notes was written up for lectures at the 1995 AIO-school (a school for PhD students) on theoretical particle physics. Later they were adapted for lectures at the Radboud University in Nijmegen, aimed at undergraduate students in their fourth year. This means that no detailed knowledge of quantum field theory is assumed, only some basic ideas like the intuitive notion of Feynman diagrams and their relation to Lagrangians. Most of the current version was updated during the spring of 2015.

The purpose of the lectures is to explain the essence of current ideas about possible physics beyond the Standard Model. Although such ideas often have a finite life-time, there are many that have been around for a decade or more, and are likely to play an important rôle in particle physics at least for another decade. The emphasis is on those ideas that are likely to survive for a while, not only due to lack of data, but also because of intrinsic importance.

Another purpose is to describe the Standard Model as a special point in the huge space of quantum field theories, and explain which alternatives are possible.

Not too much time will be devoted to the huge number of models existing in the present literature, but only a limited set of 'standard' ones is explained. In comparison with other lecture notes, more attention is paid to Standard Model physics, and furthermore most explanations are a bit more basic. A lot of background material is included in the appendices.

The list of references is still extremely limited. Only the sources on which these notes were based are listed. These may be consulted for a more complete set of references.

Conventions

The metric signature we use is $(1, -1, -1, -1)$. This means that for on-shell momenta $p^2 \equiv p^\mu p_\mu = m^2$. The standard Dirac action is $i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi$ and the standard action for a massive real scalar is $\frac{1}{2}\partial^\mu\phi\partial_\mu\phi - \frac{1}{2}m^2\phi^2$. Repeated indices are always to be summed over, but in a few equations the sums are written out explicitly anyway. In most cases raised or lowered indices have no special significance. The exceptions are space-time indices, which are always raised and lowered with the metric $g_{\mu\nu}$, and $SU(2)$ spinor indices, which are raised or lowered with the ϵ-tensor $\epsilon_{\alpha\beta}$. Except for a few pages discussing supergravity, the metric is equal to the flat metric $\eta_{\mu\nu}$. Conventions regarding superspace generally follow [2]. Covariant derivatives are of the form $\partial_\mu - ieA_\mu$ for positively charged particles in electromagnetism (note that some texts use the opposite sign for the gauge field term). The meaning of "+ c.c" is "add the complex conjugate". In an expression involving operators this is to be interpreted as the hermitean conjugate. The terms to be conjugated are either indicated by brackets, or if there are no brackets "c.c" applies to all terms. Several other conventions are stated in the appendices.

1 Introduction

Our field – considering its name "High Energy Physics" – is perhaps best characterized by the quest for the fundamental laws of physics. Now that we have, in principle, a very satisfactory description of all natural phenomena occurring on this planet in terms of the "Standard Model", it is natural for us to ask what lies beyond that model.

1.1 A Complete Theory?

But before doing that we should appreciate the remarkable situation that we are in. The current time can without exaggeration be called a historical moment in the history of physics. Never before did we have any right to entertain the thought that we are close to a fundamental theory of all phenomena in our universe. Compare the Standard Model to its predecessors, Atomic Physics, Nuclear Physics and Hadronic Physics. Atomic physics lacked an explanation for radioactivity and the energy source of the sun. Nuclear physics was never even a theory, and neither was hadronic physics. Furthermore, unlike the Standard Model, all of these theories break down if one tries to extrapolate them to higher energies.

On the 4th of July 2012 CERN announced the discovery of the last Standard Model particle that was missing, the famous Higgs boson. It was found after a decades-long quest, fifty years after its first theoretical description. This particle completes the Standard Model. After its discovery, there are no other concretely defined particles on the search list: there is no particle with definite properties (spin, color, charge, mass) that we are still looking for. The Standard Model remains consistent even if we extrapolate it all the way to the Planck scale, about 10^{19} GeV; no new particles are needed for that. This also implies that to the best of our knowledge all features of the world around us can be derived, in principle, from the quarks, leptons and interactions of the Standard Model. The words "in principle" are important: of course there are plenty of phenomena that we do not really understand well, such as high T_c conductivity, some astrophysical phenomena, strong interaction physics, the origin of life, and the nature of consciousness, but few people doubt that we do know the basic laws of physics that underlie these phenomena. They are all incorporated in a very simple Lagrangian involving fields of spin $\frac{1}{2}$, 1, and 0. There is no convincing reason to doubt that all atomic, molecular and solid state physics, all chemistry and biology, and all nuclear and hadronic physics ultimately follows from this Lagrangian, even though actually deriving it may be far beyond our capacities. Given these achievements, the Standard Model has a rather modest name. Perhaps "The theory of almost everything" [25] would be more appropriate.

This "theory of almost everything" does not contain gravity, but for all practical purposes this is easy to remedy by coupling it, classically, to Einstein's general relativity. We then have a complete theory for everything in our solar system.

This special moment may pass, and at any moment new experimental or observational information may change everything. In fact it is rather surprising that this has not happened already. Many ideas regarding physics beyond the Standard Model predicted the first appearance of "new physics" at several orders of magnitude below the energy scale the LHC can currently reach. We may still find evidence for such new physics, and indeed at this moment (early 2016) there exist some tantalizing results that put the Standard Model under stress. None of these has reached the limit of five standard deviations that we require for observations in particle physics. But if it happens, the current moment is merely a window in time, whose existence is rather puzzling. There is no obvious reason why there would be an energy gap between new physics and old physics.

1.2 Gravity and Cosmology

In any case, there must be more than just the Standard Model. The Standard Model with gravity may describe the solar system correctly, it fails it larger scales. Only about one sixth of the mass that affects the rotation of galaxies consists of Standard Model matter. The rest is called "dark matter", and we do not know what it is, or even if it really exists at all. And then there is the fact that the expansion of the Universe appears to be accelerating, a phenomenon first observed in 1998 by studying distant type-Ia supernovae. This can be explained by postulating something called "dark energy", providing 70% of the energy density of the universe (see the next section for more details). Perhaps it is less mysterious than the name suggests, but it is hard to be sure.

Furthermore, adding classical gravity to the Standard Model is not satisfactory, even though it works in practices. But adding gravity renders the theory internally inconsistent, since we do not know how to quantize it. The most immediate problem, how to do perturbation theory without encountering non-renormalizable infinities, has perhaps been solved already in string theory, but may be the least profound one. Much more difficult are questions like "what is the meaning of geometry and topology in a quantum theory", or "what happens quantum mechanically near a black hole horizon".

Cosmology has other unsolved problems. One of them is to find the correct theoretical description of inflation: the hypothetical exponential expansion of the early universe, that led to the remarkable spatial flatness observed today, and which would be difficult to understand otherwise. Most cosmologists – but not all – believe inflation requires something beyond the Standard Model. Another unsolved problem is why we see a huge surplus of baryons over anti-baryons. Mechanisms to explain that go by the name of "baryogenesis". Most of them require additional particles or interactions, beyond the Standard Model.

Most of these problems – dark matter, dark energy, inflation, consistency of quantum gravity – are obviously related to gravity: without gravity they do not exist. Perhaps this suggests that gravity is the culprit, and not the Standard Model. Indeed, we should keep in mind that "inflation", "dark matter" and "dark energy" are only the names of solutions to the problem (as is "baryogenesis") and not the name of the problems themselves. For each of them there exist alternative ideas. Some of these involve modifications of the theory of gravity. Although it may seem almost like blasphemy to tinker with Einstein's theory, if a well-motivated modification is found that addresses the remaining problems, we may need even less Beyond the Standard Model physics than most people think.

The problems associated with (quantum) gravity are completely irrelevant for our accelerator experiments until we reach energies as large as $M_{\text{Planck}} = 1.2 \times 10^{19}$ GeV, the Planck mass (the precise definition is $M_{\text{Planck}} = \sqrt{\hbar c/G_N}$, where G_N is Newton's constant). At this scale we should expect the Standard Model to break down in any case.

1.3 The Energy Balance of the Universe

Information about mass or energy that is not described by the known matter in the Standard Model can be obtained from anomalous gravitational attraction in galaxies, clusters of galaxies, colliding galaxies, the formation of the aforementioned structures, gravitational lensing and the structure of the Cosmic Microwave Background. But in addition to all this a very interesting piece of information comes from the expansion of the entire universe. Obviously, this is sensitive to anything that interacts gravitationally.

The expansion of the universe is described by first making the assumption, based on observation, that spatially it is isotropic and spherically symmetric. It is assumed that this holds at any time, not just now. This means that at any moment in time the universe can be spatially flat, or a 3-sphere (positive constant curvature), or a hyperbolic surface (negative constant curvature), with a scale factor $a(t)$ that may depend on time. Here a 3-sphere is a sphere embedded in four* dimensions, on whose surface we live. In this case the scale factor $a(t)$ can be chosen equal (or proportional) to the radius. After eliminating the fourth, auxilliary spatial coordinate and transforming to polar coordinates one gets a space-time metric given by

$$ds^2 = c^2dt^2 - a(t)^2 d\Sigma^2 \tag{1.1}$$

where

$$d\Sigma^2 = \frac{dr^2}{1-kr^2} + r^2 d\theta^2 + \sin^2\theta \; d\phi^2 \tag{1.2}$$

Here positive k corresponds to a 3-sphere, negative k to a hyperbolic surface and $k = 0$ corresponds to flat space.

In the literature one finds several conventions for the dimensionality of a, k and r. Note that we may rescale a, r and k by factors λ, λ^{-1} and λ^2 respectively without changing $d\Sigma^2$. This allows us to set $a(t_0) = 1$ at a preferred time t_0 (for example: now), or distribute length dimensions over the three parameters, or to set k to a fixed value. Common conventions are to set $k = 0, \pm 1$ with r dimensionless, while a has the dimension of length, or to make a dimensionless and and give r a dimension of length. Then k has the dimension of (length)$^{-2}$. In the latter case, for positive k and $a = 1$ we find $k = 1/R^2$, where R is the radius of the 3-sphere.

This metric ansatz is now plugged into the Einstein equations

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{1.3}$$

* The fourth dimension should not be confused with time, the fourth coordinate in Minkowski space. It is simply used for a mathematical description of the surface.

One assumes the energy momentum tensor to be of the form $T = \mathrm{diag}(\rho, p, p, p)$ where ρ is the energy density and p the pressure. This is called the perfect fluid approximation, and holds for example for a gas of particles. Depending on the kind of matter considered, one gets $p = w\rho c^2$, where w is a parameter. For massive particles ("dust" or "matter") one has $w = 0$ and for massless particles ("radiation") one gets $w = \frac{1}{3}$. The Einstein equations reduce to two separate equations, one determining the time evolution of matter densities (see (1.9) below), and one equation that takes the form

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} \tag{1.4}$$

The ratio on the left hand side is the rate of change of the scale of the universe, the quantity that Hubble measured by plotting velocity (determined from Doppler shifts) versus distance. It is called the Hubble constant, although it is not really constant. The physical dimension of H is $[\mathrm{time}]^{-1}$. The quantity H^{-1} is a time, called the Hubble time, which is roughly the age of the universe.* Note that – in agreement with the discussion above – this equation is dimensionally correct if we either make k dimensionless, and give a the dimension of [length], or make a dimensionless, and give k the dimension of $[\mathrm{length}]^{-2}$. The density ρ is actually the sum of the densities ρ_i of all contributing kinds of matter. It is customary to rewrite this equation by dividing both sides by H^2, and defining a "critical density" ρ_c as

$$\rho_c = \frac{3H^2}{8\pi G} \tag{1.5}$$

Just as H this is of course not quite constant. Now we get

$$1 = \frac{\rho}{\rho_c} - \frac{kc^2}{a^2H^2} \tag{1.6}$$

We define

$$\Omega_{\mathrm{curv}} = -\frac{kc^2}{a^2H^2}; \qquad \Omega_i = \frac{\rho_i}{\rho_c}; \qquad \Omega = \sum_i \Omega_i \tag{1.7}$$

and then we get the deceptively simple equation

$$1 = \Omega_{\mathrm{curv}} + \Omega \tag{1.8}$$

Clearly, if we could measure the curvature of the universe, and hence Ω_{curv}, we can measure using this equation the sum of all matter and radiation densities. This is like weighing the entire universe. One can get information about curvature by considering the apparent size of distant objects. For example, by comparing the apparent size of nearby and far away galaxies one can get information about the curvature, but nowadays the most accurate information came from the fluctuations in the cosmic microwave background. The size of these fluctuations can be computed, and serves as a standard measuring unit. Since this

* To be precise, if we extrapolate backward in time assuming the current velocity, we find that a was zero a Hubble time ago.

comes from the most distant visible feature in the universe, it gives the best measurement for curvature. According to the latest Planck satellite data the universe is spatially flat with a precision of about .5% ($\Omega_{\rm curv} = 0.000 \pm .0005$). Since the FRW metric and perfect fluid approximation for matter is clearly just an approximation, it is implausible that the universe is *exactly* spatially flat. Whether the deviation is positive or negative is obviously of utmost interest for cosmology, but we may never know. Unlike LHC, we have only one event to look at, our universe. This implies intrinsic statistical errors, which means that there is a fundamental limit on the accuracy we can reach.

However, the importance of this measurement for particle physics lies in the second term in eqn. (1.8). It tell us that the sum of all the contribution to Ω must be very close to 1. A small part of this (about 4.9%) can be accounted for by baryonic (*i.e.* Standard Model) matter. In the past, an important piece of information comes from the deuterium abundance in the universe. Deuterium is produced during big bang nucleosynthesis, the production process being $p + n \rightarrow d + \gamma$. This process can also run in the opposite direction: photons destroy deuterium. Therefore it is not surprising that the abundance depends strongly on the baryon-to-photon ratio. Since we know the number density of photons (most of them are from the CMB), and can fairly accurate estimates of the ratio of deuterium to hydrogen in the universe, this information can be used to determine the total amount of baryonic matter. Nowadays the details of the CMB fluctuations also offer important information about the amount of baryonic matter.

From various sources (such as galaxy rotation curves, clusters of galaxies, structure formation, gravitational lensing and the CMB) we get information about the total fraction of matter. This is about 30%, including baryonic matter. Therefore there is about 70% of the total Ω missing.

Above we have discussed two kinds of contributions (apart from $\Omega_{\rm curv}$) to Ω: matter and radiation. These contributions have a different "equation of state", which in this context just means a different value for the parameter w introduced above. From general relativity one does not just get eqn. (1.4) but also an equation describing the time evolution of densities

$$\dot{\rho} = -3\left(\frac{\dot{a}}{a}\right)\rho(1+w) \tag{1.9}$$

which implies

$$\rho \propto a^{-3(1+w)} \tag{1.10}$$

The two components we have discussed so far scale as follows with a: matter as a^{-3} and radiation as a^{-4} This is intuitively clear. Matter densities scale according to volume, but radiation has an additional dependence on scale because with increasing scale their wavelength increases with a and hence the energy of each photo decreases with a. For massive particles the energy is bounded from below by their mass. There can be other contributions to the energy density of the universe. A gas of strings has $w = -\frac{1}{3}$ and scales with a^{-2}, and a gas of membranes has $w = -\frac{2}{3}$. But there is no evidence for contributions of these latter two kinds.

One contribution that we have not yet discussed in this section is a cosmological constant. The cosmological constant Λ is a parameter of classical general relativity that is allowed by general coordinate invariance. It has dimension $[\text{length}]^{-2}$ and appears in the Einstein equations as

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R - \Lambda g_{\mu\nu} = 8\pi G_N T_{\mu\nu} \ . \tag{1.11}$$

Without a good argument for its absence one should therefore consider it as a free parameter that must be fitted to the data. It contributes to the equations of motion with an equation of state $p = w\rho$, with $w = -1$. Hence it does not scale with a at all! The cosmological constant is an obvious candidate for providing the missing contribution to Ω, and indeed the data seem in agreement with an extra component with $w = -1$.

Unlike dark matter, where the Standard Model offers nothing, dark energy is provided in abundance by the Standard Model. The parameter Λ contributes to the equations of motion in the same way as vacuum energy density $\rho_{\rm vac}$, which has an energy momentum tensor $T_{\mu\nu} = \rho_{\rm vac} g_{\mu\nu}$. Vacuum energy is a constant contribution to any (quantum) field theory Lagrangian. It receives contributions from classical effects, for example different minima of a scalar potential and quantum corrections (*e.g.* zero-point energies of oscillators). However, it plays no rôle in field theory as long as gravity is ignored. It can simply be set to zero. Since vacuum energy and the parameter Λ are indistinguishable it is customary to identify $\rho_{\rm vac}$ and Λ. The precise relation is

$$\frac{\Lambda}{8\pi} = \frac{G_N \rho_{\rm vac}}{c^2} := \rho_\Lambda \ . \tag{1.12}$$

This immediately relates the value of Λ with all other length scales of physics, entering in ρ_Λ.

Vacuum energy is a notoriously divergent quantity in quantum field theory. One may think of it as the sum of the ground states energies of all the harmonic oscillators in the mode expansion of all the fields. Alternatively, and equivalently, it may be decribed by the contribution of loop diagrams without external lines, that one usually throws away in QFT. The contribution of such a loop diagram is proportional to

$$\int d^4k \, \log(k^2 - m^2) \tag{1.13}$$

To understand the logarithm note that an n-point graph with external momenta is correctly obtained by differentiating n times with respect to m^2, and hence a zero-point amplitude corresponds to not differentiating at all. If we cut off the integration at some scale M, we get a contribution proportional to M^4. Such a cut off could be physically inspired by some new physics, such as a discrete structure of space-time. But surely the scale of such new physics must lie beyond the range of LHC, because otherwise we should have seen it already. This would suggest that $M > 1$ TeV. Not only quantum vacuum energy contributes to ρ_Λ, but also classical vacuum energy like the shift in the potential that occurs in the Higgs mechanism.

The value of ρ_Λ is irrelevant in QFT, but it has important effects on the time evolution of the universe and on its size. Another relation obtained from the Einstein equations (derivable from the foregoing two equations) is

$$\ddot{a} = -\frac{4\pi G}{3}(1 + 3w)\rho \tag{1.14}$$

From this equation we see that matter and radiation decelerate the expansion of the universe ($\rho > 0$ and $w = 0$ or $\frac{1}{3}$), while a cosmological constant with $\rho_\Lambda > 0$ accelerates the expansion. Unlike matter densities, ρ_Λ can have both signs, as we can already see from the previous paragraph: the loop diagrams have opposite signs for bosons and fermions. Hence for positive Λ the universe undergoes accelerated expansion, and for negative Λ it collapses. The value of Λ becomes relevant as soon as it dominates all other contributions. But since all other contributions scale with negative powers of a, in a universe that starts expanding this eventually happens. This implies that the simple observation that our universe exists for billions of years and has a size of billions of light years means that we know an experimental upper limit on $|\Lambda|$, and that we know about this limit for a long time already.

It is entertaining to use Planck units to specify Λ. Then the natural value of Λ is about one Planck mass per Planck volume. The limit obtained from the size and lifetime of the universe described above is about 10^{-120} in Planck units. Contributions from particle physics cut off at 1 TeV yield a value of about 10^{-60} in Planck units, far above the observational upper limit. For this reason many people believed that if Λ is so small, it would actually vanish for a reason still to be discovered. But in 1998 it was discovered that the universe is undergoing accelerated expansion. By now we know that the value of Λ needed to explain this is about the right quantity needed for Ω.

Interestingly, the current discrepancy in the value of Ω of about 70% was already known for decades, albeit less precisely. People did not know that the universe was as close to flatness as precisely as we know today. In Alan Guth's famous paper on inflation [15] he assumes that $0.01 < \Omega < 10$. That seems hardly "close to 1". However, if one extrapolates backwards in time, the the contribution of Ω_{curv} relative to matter and radiation approaches zero. Hence it would seem that Ω_{curv} must be extremely close to zero in the early universe. Indeed, $\Omega = 1$ means that the density is equal to the critical density. The term "critical density" indicates that being above or below this value makes a huge difference. Indeed this is correct. This value turns out to be a point of instability. If one starts with Ω just above one, the universe starts expanding, but recollapses. If one starts just below $\Omega = 1$ the universe expands very rapidly, an all matter gets diluted very fast. To get a universe that still exists after 13.8 billion years and that has a substantial matter density, one has to start with Ω very close to one. How close depends on how early one starts. According to [15], if one starts at a temperature corresponding to 1 MeV, one has to tune Ω to the value 1 with fifteen digit precision.

To explain this apparent fine-tuning, one may invent a mechanism that puts it very close to zero in the early universe. Inflation is such a mechanism. Then one would expect Ω to be very close to 1 today. This theoretical expectation did not agree with the known matter contributions to Ω, and it was also known that dark energy could fill in the gap. Hence one could claim that inflation predicted a positive cosmological constant of roughly the observed size. But still, it seems that nobody was courageous enough to predict that.

1.4 Environmental Issues

A remarkable fact about the current situation in our understanding of the universe, is that almost all remaining problems are "environmental". We are puzzled about values of parameters that are sometimes rather peculiar, but there is not really a concrete problem associated with these values. Physics would be equally consistent if we change these values.

We may almost have forgotten what a real problem looks like. But if we go back to the middle of last century, when people were trying to understand nuclear physics, the situation was very different. Nuclear physicists were so desperate that one of them exclaimed: "Even a wrong theory would be tremendous progress."

We still have some *real* problems left, but the list is very short: what is the correct theory of quantum gravity, and what are the constituents of dark matter? In the latter case, and alternative possibility is that we have to modify gravity somehow, but no matter how one looks at it, there is a discrepancy between the left-hand side and the right-hand side of Einstein's equations. This is a real problem. On the other hand, "dark energy" can be viewed as an environmental problem. We can describe it by simply choosing an already existing parameter appropriately, but of course that does not imply that there is no new physics that describes it. But anyone who tries to explain dark energy with new physics will first have to argue away the old physics.

There is perhaps one other real problem: stability of the Higgs potential. With the current values of the Higgs mass and the top quark mass (to which this issue is most sensitive), we are two or three standard deviations beyond the boundary line of stability. Beyond that line the quantum-corrected Higgs potential develops a second minimum, to which our universe could tunnel. This does not mean that the entire universe tunnels instantaneously, but that somewhere a tiny bubble of "false vacuum" appears, that starts expanding to cover the entire universe. One can compute the life-time of the universe under these conditions, and with current data this is expected to be far more than 13.8 billion years. However there are several theoretical uncertainties, and furthermore one has to worry not just about the current situation, but also about the history of the universe. So this is potentially a real problem.

Finally, neutrino masses are a real problem for the "classic" Standard Model, which was defined to have only left-handed neutrinos and no neutrino masses. Then, by definition, neutrino oscillations imply non-zero neutrino masses and hence new physics. But in principle neutrino masses can easily be introduced in a manner analogous to quark masses, which requires assuming the (still unproven) existence of right-handed neutrinos. This is an alternative definition of the Standard Model we might have adopted. In that case the actual mass of the neutrino and its smallness becomes another environmental problem.

All the rest can be called environmental problems. This list includes:

- Horizon problem: Why is the the early universe homogeneous, although there are many causally disconnected regions?

- Flatness problem: Why was the energy density in the early universe so close to the critical density?
- Baryons: Why are there only baryons and leptons, but essentially no anti-particles in the known universe?
- Dark energy: Why is it so small in comparison to natural scales?
- Dark energy vs. dark matter versus baryonic matter: why are there contributions to Ω today comparable in size? (the "why now" problem)
- Strong CP violation: Why is $\theta_{\rm QCD}$ extremely small, possibly zero?
- The Hierarchy problem: why is the Higgs mass so much smaller than the Planck mass?
- The Weak/Strong coincidence: why is the QCD scale close to the weak scale? Or more precisely: why are light quark mass differences of the same order of magnitude as nuclear binding energies?
- Neutrino masses: Why are they so much smaller than charged lepton masses?
- Quark and lepton masses: Strange hierarchies, for example $m_e \ll m_t$.
- Quark and lepton mixing angles: Why are quark mixing angles very small, while lepton mixing angles are not?
- Standard Model gauge group: Why $SU(3) \times SU(2) \times U(1)$?
- Standard Model family structure: Why this particular choice of representations?
- Charge quantization: Why is the proton charge exactly equal to minus the electron charge?
- Number of families: Why three?

These are all "why" questions. It is not guaranteed that we will ever get an answer to that kind of question, and there is no way to force nature to provide an answer. The Standard Model as we know it today, in 2016, is perfectly consistent. We get sensible answers for any physical process for energies far beyond those of the LHC, as long as we do not get too close to the Planck scale. Depending on the precise values of the parameters, we may have to conclude that our universe is not absolutely stable, but even that is not an inconsistency that requires a solution.

Perhaps the Standard Model is just the way it is, and we will have to accept that. But perhaps there is a multiverse, a plethora of universes with different "Standard Models", of which ours is just one. Then probably most of these alternatives cannot support observers to be puzzled about the why questions.

1.5 Baryogenesis

But not all remaining issues are likely to be merely "environmental". Consider for example the baryon excess. One could claim to solve it by simply assuming that the universe started with just baryons and leptons. But that seems a rather strange assumption, even if we allow for a universe with regions with some baryon excess and others with anti-baryon excess. If we randomly distribute baryons and anti-baryons over a large volume, the chance of finding a subregion like our universe is ridiculously small. Worse yet, at the end of inflation the pre-existing matter has been diluted by many orders of magnitude, so that the initial state is reset to zero. So we do not have an initial condition to choose, nature already set it up for us. All matter we observe is generated by "re-heating" at the end of inflation. So it seems clear that we need a mechanism to generate a net baryon number from nothing. Such a mechanism is generically called "baryogenesis". There are many ideas about this in the literature.

What all of these different ideas must have in common is that they must satisfy the Sakharov conditions. These are

1. Baryon number violation
2. C-violation
3. CP-violation
4. Processes out of thermal equilibrium

The first point is obvious. If charge conjugation is exact, then for very process that violates baryon number, for example $\phi \to p + e^-$ there is a process $\phi^* \to \bar{p} + e^+$. Even if ϕ^* is different from ϕ, we will have to assume that the initial ensemble contains equal numbers of ϕ and ϕ^*, for the same reason why we assume that we start with equal densities of p and $\bar{p}$. But then C-invariance makes both processes occur equally often, and we still cannot generate a new number of baryons. The argument for CP-non-invariance is identical, except that it would imply that for every left-handed proton there is a right-handed anti-proton in the final state. Finally the last condition is needed because in thermal equilibrium the reactions destroying baryon number would compensate the reactions creating it.

The standard model has baryon number violation (see section 5.5.8), C-violation and CP-violation in the weak interactions. The fourth condition depends on details of the evolution of the universe. It is realized if suitable phase transitions occur. The standard model constrains the options for phase transitions, but to know hat really occurs we need some cosmological input. It appears that the ingredients we have in the standard model are not sufficient to explain the observed asymmetry.

1.6 Beyond the Standard Model

There are many ideas that address the problems listed above, and that require some kind of "new physics" at higher energy scales. Some of these look very appealing and suggest beautiful underlying structures, with ambitious sounding names like "Grand Unified Theories" and "Supersymmetry". Verification of some of these ideas has seemed tantalizingly close at various times during the past three decades, and nevertheless it has not happened yet. Does nature not like symmetry? Whatever the answer, these ideas will be around for the foreseeable future, and will continue to be explored at the LHC and other experiments. Any particle theorist should know about them, and have a basic understanding of their good and not-so-good features. This is the main focus of these lectures.

2 Gauge Theories

In this section we present a brief introduction to non-abelian gauge theories, one of the main ingredients of the Standard Model. This assumes some basic knowledge of classical electrodynamics, which will be generalized from abelian symmetry groups ($U(1)$, or just phases) to non-abelian ones. Furthermore the notion of Euler-Lagrange equations for classical fields is assumed, and basic canonical quantization of free field theories.

2.1 Classical Electrodynamics

Classical electrodynamics can be derived from the following simple Lagrangian (or more properly, Lagrangian density):

$$\mathcal{L} = -\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} + J^{\mu} A_{\mu} \ , \tag{2.1}$$

with

$$F_{\mu\nu} = \partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu} \ . \tag{2.2}$$

To verify this statement we simply derive the Euler-Lagrange equations that follow from this Lagrangian

$$\partial_{\rho} \frac{\partial \mathcal{L}}{\partial(\partial_{\rho} A_{\sigma})} = \frac{\partial \mathcal{L}}{\partial A_{\sigma}} \ . \tag{2.3}$$

This yields

$$\partial^{\nu} F_{\mu\nu} = J_{\mu} \ . \tag{2.4}$$

Now define electric and magnetic fields

$$E_i = F_{0i} \ , \quad B_i = \tfrac{1}{2} \epsilon_{ijk} F_{jk} \ , \tag{2.5}$$

and the equation takes the form

$$\begin{aligned} \vec{\nabla} \times \vec{B} - \partial_t \vec{E} &= \vec{J} \\ \vec{\nabla} \cdot \vec{E} &= J_0 \end{aligned} \tag{2.6}$$

These are two of the four Maxwell equations (the other two,

$$\vec{\nabla} \times \vec{E} + \partial_t \vec{B} = 0$$

$$\vec{\nabla} \cdot \vec{B} \;\; = \;\; 0 \; . \tag{2.7}$$

are trivially satisfied if we express the electric and magnetic fields in terms of a vector potential A_μ). Consistency of Eq. (2.4) clearly requires

$$\partial^\mu J_\mu = \partial^\mu \partial^\nu F_{\mu\nu} = 0 \; . \tag{2.8}$$

because of the antisymmetry of $F_{\mu\nu}$. This implies that J must be a conserved current. For such a current one can define a charge

$$Q = \int d^3 x J_0 \; . \tag{2.9}$$

where the integral is over some volume V. This charge is conserved if the flux of the current $\vec{J}$ into the volume vanishes.

2.2 Gauge Invariance

Consider the bi-linear terms in the Lagrangian (2.1). If we quantize it naively, it seems that we will end up with particles having 4 degrees of freedom, since A_μ has four components. However, this is incorrect for two reasons. First of all, one degree of freedom is not dynamical, *i.e.* does not appear with a time derivative, namely A_0. This means that the corresponding canonical momentum does not exist, and one will not obtain creation/annihilation operators for this degree of freedom. In addition to this there is one degree of freedom that does not really appear in the action at all. Suppose we replace A_μ by $A_\mu + \partial_\mu \Lambda(x)$, where $\Lambda(x)$ is some function. It is easy to see that $F_{\mu\nu}$ does not change at all under this transformation, and therefore the action is also invariant. This is called *gauge invariance*. Hence the action does not depend on Λ, which removes another degree of freedom. We conclude that there are just two degrees of freedom instead of 4. These two degrees of freedom correspond to the two polarizations of light. The quanta of A_μ are called photons.

If we add a mass term $m^2 A_\mu A^\mu$ to the Lagrangian it is still true that A_0 is not dynamical, but gauge invariance is broken. Therefore now we have three degrees of freedom. Just as fermions, massless and massive vector fields have very different properties.

Now consider the coupling $A_\mu J^\mu$. This is *not* invariant under gauge transformations, but observe what happens if instead of the Lagrangian density we consider the action,

$$S_J = \int d^4 x A_\mu J^\mu \; . \tag{2.10}$$

This transforms into itself plus a correction

$$\delta S_J = \int d^4 x \partial_\mu \Lambda J^\mu \; . \tag{2.11}$$

Integrating by parts, and making the assumption that all physical quantities fall off sufficiently rapidly at spatial and temporal infinity, we get

$$\delta S_J = -\int d^4x \Lambda \partial_\mu J^\mu \ , \tag{2.12}$$

which vanishes if the current is conserved, as we have seen it should be.

Gauge invariance (or current conservation) is our main guiding principles in constructing an action coupling the electromagnetic field to other fields. Consider for example the free fermion. It is not difficult to write down a Lorentz-invariant coupling:

$$\mathcal{L}_{\text{int}} = eqA_\mu \bar{\psi}\gamma^\mu\psi \ , \tag{2.13}$$

which is to be added to the kinetic terms

$$\mathcal{L}_{\text{kin}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} \ . \tag{2.14}$$

Note that we have introduced two new variables here: the coupling constant e and the charge q. The latter quantity depends on the particle one considers; for example for the electron $q = -1$ and for quarks $q = \frac{2}{3}$ or $q = -\frac{1}{3}$. The coupling constant determines the strength of the interaction. This quantity is the same for all particles. It turns out that the combination $\alpha = \frac{e^2}{4\pi}$ is small, $\approx 1/137.04$. This is the expansion parameter of QED, and its smallness explains why perturbation theory is successful for this theory. Although only the product eq is observable, it is convenient to make this separation.

With this choice for the interaction, the current is

$$J^\mu = eq\bar{\psi}\gamma^\mu\psi \ . \tag{2.15}$$

Using the equations of motion (*i.e.* the Dirac equation) one may verify that this current is indeed conserved, so that the theory is gauge invariant. But there is a nicer way of seeing that. Notice that the fermion kinetic terms as well as the interaction are invariant under the transformation

$$\psi \to e^{ieq\Lambda}\psi \ ; \quad \bar{\psi} \to e^{-ieq\Lambda}\bar{\psi} \ , \tag{2.16}$$

if Λ is independent of x. Because of the derivative this is not true if Λ does depend on x. However, the complete Lagrangian $\mathcal{L}_{\text{kin}} + \mathcal{L}_{\text{int}}$ is invariant under the following transformation

$$\begin{gathered} \psi \to e^{ieq\Lambda(x)}\psi \ ; \quad \bar{\psi} \to e^{-ieq\Lambda(x)}\bar{\psi} \\ A_\mu + \partial_\mu\Lambda(x) \ . \end{gathered} \tag{2.17}$$

This is the gauge transformation, extended to act also on the fermions. This is sufficient for our purposes: it shows that also in the presence of a coupling to fermions one degree of freedom decouples from the Lagrangian, so that the photon has only two degrees of freedom.

2.3 Noether's Theorem

Actually all these facts are related, and the missing link is *Noether's theorem*. Simply stated, this works as follows. Suppose an action is invariant under a global (x-independent) transformation of the fields. Suppose it is not invariant under the corresponding local (x-dependent) transformation. Then the variation must be proportional to the derivative of the parameter $\Lambda(x)$ of the transformation (for simplicity we assume here that only first derivatives appear, but this can be generalized). Hence the variation of the action must have the form

$$\delta S = \int d^4x \partial^\mu \Lambda(x) J_\mu[\text{Fields}] \tag{2.18}$$

where $J_\mu[\text{Fields}]$ is some expression in terms of the fields of the theory. The precise form of J_μ depends on the action under consideration, and follows in a straightforward way from the symmetry.

The equations of motion are derived by requiring that the action is a stationary point of the action, which means that terms linear in the variation, such as Eq. (2.18) must vanish. Integrating by parts we get then

$$\int d^4x \Lambda(x) \partial^\mu J_\mu[\text{Fields}] = 0 \; . \tag{2.19}$$

Since $\Lambda(x)$ is an arbitrary function, it follows that the *Noether current* $J_\mu[\text{Fields}]$ is conserved. It is an easy exercise to show that the symmetry (2.16) of the free fermion action does indeed yield the current (2.15).

2.4 Covariant Derivatives

Checking gauge invariance can be made easier by introducing the *covariant derivative*

$$D_\mu = \partial_\mu - ieqA_\mu \tag{2.20}$$

This has the property that under a gauge transformation

$$D_\mu \rightarrow e^{ieq\Lambda} D_\mu e^{-ieq\Lambda} \tag{2.21}$$

If we now write the Lagrangian as

$$\mathcal{L} = i\bar{\psi}\gamma^\mu D_\mu \psi \tag{2.22}$$

checking gauge invariance is essentially trivial. One can simply pull the phases through D_μ, even if they are x-dependent!

Replacing normal derivatives by covariant ones is called *minimal substitution*, and the resulting interaction terms *minimal coupling*. It is a general principle: an action can be made gauge invariant by replacing all derivatives by covariant derivatives. For example the coupling of a photon to a complex scalar is given by the Lagrangian

$$\mathcal{L} = (D_\mu(q)\varphi)^*(D^\mu(q)\varphi) \ , \tag{2.23}$$

where q is the charge of φ. Note that φ must be a complex field since the gauge transformation multiplies it by a phase. Note also that the field φ^* has opposite charge.

The Lagrangian of the vector bosons can also be written down in terms of covariant derivatives. We have (for any $q \neq 0$)

$$-ieqF_{\mu\nu} = [D_\mu(q), D_\nu(q)] \ , \tag{2.24}$$

from which gauge invariance of the action follows trivially. Here q has no special significance, and any non-zero value can be used. This relation should be interpreted as a relation for differential operators acting on some function $\phi(x)$. The space-time derivatives in both covariant derivatives act on ϕ, but in the final result the action of the derivatives on ϕ cancels out.

2.5 Non-Abelian Gauge Theories

The field transformations we used to construct QED

$$\psi \to e^{ieq\Lambda(x)}\psi \tag{2.25}$$

are local (x-dependent) elements of the group $U(1)$, the group of phases $e^{i\theta}$. Since all elements of this group commute it is called *abelian*.

The whole formalism can be extended in a rather straightforward way to non-abelian groups. For simplicity we restrict ourselves here to the group $SU(2)$. This group is well-known as the rotation group for spinors, but here it will play a totally different rôle.

Suppose we have a field ψ^i with an extra index i. For definiteness we will assume that the field is a fermion (it could also be a complex scalar) and that i just takes two values, 1 and 2. The kinetic terms are

$$\mathcal{L}_{\text{kin}} = i\sum_{i=1}^{2} \bar{\psi}^i\gamma^\mu\partial_\mu\psi^i \ . \tag{2.26}$$

This Lagrangian is invariant under transformations

$$\psi^i \to \sum_j U^{ij}\psi^j \ , \tag{2.27}$$

where U is a unitary two-by-two matrix (note that $\bar{\psi}^i \to (\bar{\psi}U^\dagger)^i$). Under multiplication these matrices form a group, $U(2)$, and for simplicity we restrict ourselves to the subgroup $SU(2)$ of matrices with determinant 1 (the overall phase is just a $U(1)$ transformation, which we already considered earlier in this chapter).

Now suppose we consider a space-time dependent transformation $U(x)$. Because of the derivative this is no longer an invariance of the Lagrangian. Imitating QED, we can try to cure that by introducing a covariant derivative D_μ that must transform as

$$D_\mu \to U D_\mu U^{-1} \ . \tag{2.28}$$

Without loss of generality we may assume that D_μ has the form $\partial_\mu + A_\mu$, so that it reduces to the ordinary derivative for $A_\mu = 0$. Since A_μ must act on the indices i it must be a two-by-two matrix.

Therefore we can expand it into a complete basis of two-by-two matrices. Any such matrix can be written as $a + \vec{b} \cdot \vec{\sigma}$, where a and $\vec{b}$ are four complex constants and σ are the Pauli matrices. In this case we want A_μ to be anti-hermitean (just as ∂_μ) so the constants must be purely imaginary. Furthermore we will set $a = 0$. This is not necessary, but the constant component of A_μ corresponds to an abelian gauge field that belongs to the overall phase in $U(2)$ in comparison with $SU(2)$. Since we are only considering $SU(2)$ here, only the components proportional to $\vec{\sigma}$ are interesting for us. Instead of the Pauli matrices we will use the matrices

$$T^a = \tfrac{1}{2}\sigma^a \ . \tag{2.29}$$

This avoids several factors $\frac{1}{2}$ in formulas, and also prevents confusion with the Pauli-matrices used for spin. Then we write the gauge fields in the following way

$$A_\mu = -ig \sum_a A^a_\mu T^a \ , \tag{2.30}$$

where we have introduced a factor $-ig$ for future purposes. The component fields A^a_μ are real. The factor g will play the rôle of the coupling constant, just as e in QED. Note that there are three gauge fields, for $a = 1, 2, 3$.

To see how A_μ should transform, it is instructive to consider infinitesimal transformations

$$U(\vec{\xi}) = 1 + i\vec{\xi} \cdot \vec{T} \ . \tag{2.31}$$

Expanding Eq. (2.28) to first order in ξ we get

$$A_\mu \to A_\mu - i\partial_\mu \vec{\xi} \cdot \vec{T} + ig\vec{\xi} \cdot \left[\vec{T}, A_\mu\right] \ . \tag{2.32}$$

In terms of the components we get

$$A^a_\mu \to A^a_\mu + \frac{1}{g}\partial_\mu \xi^a - \xi^b \epsilon^{abc} A^c_\mu \ . \tag{2.33}$$

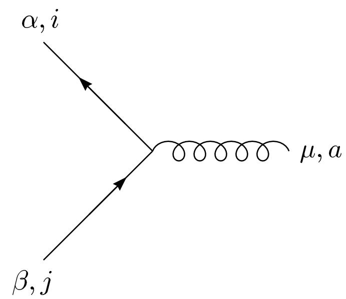
2.6 Coupling to Fermions

Replacing in the fermion action ∂_μ by D_μ we have coupled a three-component gauge field to the fermion. The action is explicitly

$$\mathcal{L} \ = \ i \sum_{i,j=1}^{2} \bar{\psi}^i \gamma^\mu D^{ij}_\mu \psi^j$$

$$
\begin{aligned}
&= i\sum_{i,j=1}^{2}\bar{\psi}^i\gamma^\mu\left[\partial_\mu\delta_{ij} - igA_\mu^a T_{ij}^a\right]\psi^j \\
&\equiv i\bar{\psi}^i\gamma^\mu\partial_\mu\psi^i + gA_\mu^a\bar{\psi}^i\gamma^\mu T_{ij}^a\psi^j \\
&= \mathcal{L}_{\text{kin}} + \mathcal{L}_{\text{int}} \ .
\end{aligned}
\tag{2.34}
$$

In perturbation theory the first term gives rise to the fermion propagator, which in comparison to the one of QED has an extra factor δ_{ij}. The second term is a perturbation, which yields the Feynman rule (the curly line represents a non-abelian gauge boson, see below)



$$ig\gamma^\mu_{\alpha\beta}T^a_{ij}$$

The fermion spinors $u, v, \bar{u}, \bar{v}$ now get extra indices $i, j, \ldots$ in addition to their spinor indices. The matrices T^a are multiplied together along a fermion line, starting at an outgoing arrow and following the line against the arrow direction. If there is a closed fermion loop, one obtains a trace of a product of matrices T. Combinatorically this works exactly as for the γ matrices.

2.7 Gauge Kinetic Terms

We can also write down a kinetic term for the gauge fields. First define

$$F_{\mu\nu} = [D_\mu, D_\nu] = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] \ . \tag{2.35}$$

Just like A_μ, the *field strength tensor* $F_{\mu\nu}$ is a two-by-two matrix, and it can be expanded in terms of Pauli matrices as

$$F_{\mu\nu} = -ig\sum_a F_{\mu\nu}^a T^a \ , \tag{2.36}$$

Now we can express the components of $F_{\mu\nu}$ in terms of those of A_μ:

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g\epsilon^{abc}A_\mu^b A_\nu^c \ . \tag{2.37}$$

The reason for writing $F_{\mu\nu}$ as in Eq. (2.35) is that it has a nice transformation law under gauge transformations

$$F_{\mu\nu} \to UF_{\mu\nu}U^{-1} \ . \tag{2.38}$$

Note that in contrast to the field strength of QED, the field strength tensor of non-abelian gauge theories is not gauge invariant. However, we can make a gauge invariant combination,

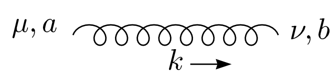
$$\mathcal{L}_{\text{gauge}} = \frac{1}{2g^2}\ \mathrm{Tr} F_{\mu\nu} F^{\mu\nu}\ . \tag{2.39}$$

where the trace is over the two-by-two matrices. Because of the cyclic property of the trace this quantity is gauge invariant. It is also manifestly Lorentz invariant, and hence it is a good candidate for the Lagrangian of the non-abelian gauge fields. If we write it out in components we get

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}\sum_a F^a_{\mu\nu} F^{\mu\nu,a}\ . \tag{2.40}$$

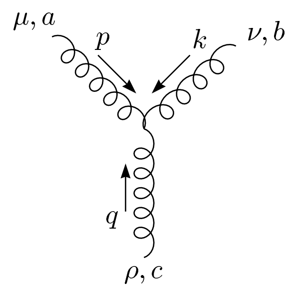
2.8 Feynman Rules

Note that the linear terms in $F^a_{\mu\nu}$ are just like those for QED. If that was all there was we just had three copies of QED, for $a = 1, 2, 3$. The quadratic terms in $F^a_{\mu\nu}$ give rise to cubic terms in the Lagrangian that are proportional to g, and quartic terms proportional to g^2. These are interactions. Just as in QED, we use the bi-linear terms in the action to define a propagator, which in fact is identical to the one of QED except for a factor δ^{ab}. To distinguish non-abelian gauge bosons from photons we use another kind of line:

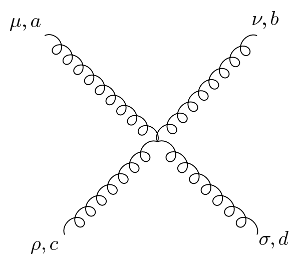


$$-\frac{i}{k^2} g_{\mu\nu} \delta^{ab}$$

The cubic and quartic term give rise to interactions, whose Feynman rules are



$$g\epsilon^{abc}\left[(q-p)_\nu g_{\mu\rho} + (k-q)_\mu g_{\nu\rho} + (p-k)_\rho g_{\mu\nu}\right]$$



$$\begin{aligned} -ig^2[&\epsilon_{eab}\epsilon_{ecd}(g_{\mu\rho}g_{\nu\sigma} - g_{\nu\rho}g_{\mu\sigma}) \\ &+\epsilon_{eac}\epsilon_{edb}(g_{\nu\rho}g_{\mu\sigma} - g_{\mu\nu}g_{\rho\sigma}) \\ &+\epsilon_{ead}\epsilon_{ebc}(g_{\sigma\rho}g_{\mu\nu} - g_{\mu\rho}g_{\nu\sigma})] \end{aligned}$$

Just like photons, non-abelian gauge bosons A^a_μ have two components (two for each value of the index a of course), and when they appear as external lines they are represented by polarization tensors ϵ^a_μ.

2.9 Other Gauge Groups

All the foregoing can easily be generalized to other symmetries. Instead of $SU(2)$ we may use other groups like $SU(N)$ or $SO(N)$. In general, one has instead of the Pauli matrices some other set of hermitean matrices T^a. These matrices satisfy a generalized set of commutation relations,

$$\left[T^a, T^b\right] = if^{abc}T^c \ . \tag{2.41}$$

where f^{abc} is a set of real numbers that are called the *structure constants* of the group. They are fully anti-symmetric in all three indices. In addition to the commutation relations, the only other property one needs to know about these matrices is their normalization. Often one uses

$$\mathrm{Tr}\ T^aT^b = \tfrac{1}{2}\delta^{ab} \ , \tag{2.42}$$

which is indeed satisfied by the $SU(2)$ matrices we used. In Eq. (2.39) this normalization is implicitly assumed.

To write down Lagrangians, transformations and Feynman rules for another group, simply make everywhere the replacement

$$\epsilon^{abc} \to f^{abc} \ . \tag{2.43}$$

An interesting special case is the group $SU(3)$, with fermions in triplet representations. There are eight traceless hermitean three-by-three matrices T^a. This yields QCD (quantum chromodynamics). Corresponding to the eight matrices there are eight gauge bosons, called gluons, while the fermions are called quarks. It is now completely straightforward to write down the QCD Lagrangian.

Note that the entire discussion of non-abelian gauge theories is completely analogous to that of QED. This is in fact a special case, obtained by replacing

$$\begin{array}{rcl} T^a & \rightarrow & q \\ g & \rightarrow & e \\ \epsilon_{abc} & \rightarrow & 0 \end{array} \quad . \tag{2.44}$$

3 The Higgs Mechanism

The second important ingredient of the Standard Model we will need to discuss is the Higgs mechanism.

3.1 Vacuum Expectation Values

The classical value of a field is the value it has when all quantum fluctuations are in their ground state. Often in quantum field theory it is implicitly assumed that the classical value of any field, $\langle 0|\phi|0\rangle$, vanishes. Indeed the usual mode expansion for scalar fields is

$$\phi(\vec{x},t) = \frac{1}{(2\pi)^3}\int \frac{d^3k}{2\omega(\vec{k})}\left(a_{\vec{k}}e^{i\vec{k}\cdot\vec{x}-i\omega(\vec{k})t} + a^{\dagger}_{\vec{k}}e^{-i\vec{k}\cdot\vec{x}+i\omega(\vec{k})t}\right) \quad , \tag{3.1}$$

where $\omega(\vec{k}) = \sqrt{\vec{k}^2+m^2}$. If one computes the vacuum expectation value (v.e.v.) of such a quantum field one finds zero, since the v.e.v. of any creation/annihilation operator vanishes. But this is not necessarily true. In general one can have $\phi = \phi_{\rm cl} + \phi_{\rm qu}$, with all quantum fluctuations in the second term, and $\phi_{\rm cl} \neq 0$. In general, if one quantizes a theory one considers the fluctuations of fields around minima of the classical action. These fluctuations define a set of harmonic oscillators, to which the quantization procedure is applied. For this to make sense, the change in energy must be quadratic (or higher order) in terms of infinitesimal fluctuations. In particular, there should not be any linear dependences. This implies that the classical field must be a solution to the equations of motion, or in other words a stationary point of the action. Usually $\phi = 0$ is a solution to the equations of motion, but in some cases there may be other solutions.

The classical value, $\phi_{\rm cl}$, serves as a new, non-trivial ground state of the theory. One defines the vacuum in such a way that $\langle 0|\phi_{\rm qu}|0\rangle = 0$. The properties of the quantum vacuum state, and in particular the symmetries it respects, are determined by those of the classical "background field" $\phi_{\rm cl}$. The possible values of $\phi_{\rm cl}$ are restricted by the symmetries the theory should have. In general, with $\phi_{\rm cl} \neq 0$ there will be fewer symmetries than with $\phi_{\text{cl}} = 0$. If some symmetry operation changes the classical vacuum, than this is not going to be a symmetry of the theory expanded around that vacuum.

In any case we want our vacuum to be translation invariant and Lorentz-invariant. This restricts ϕ_{cl} to be a constant over all of space-time, and it restricts ϕ to be a scalar field. If a vector field has a classical value, then it must point to some specific direction. This breaks rotation invariance.

In addition to space-time symmetries, fields may also transform under "internal" symmetries (by definition, this is anything else than Poincaré transformations). If a classical background field transforms non-trivially under such symmetries, and if this is used to define the classical vacuum of the theory, then the symmetry is broken. Historically, this is called *spontaneous symmetry breaking* of a symmetry. This means that there is a symmetry of the action that is not realized in the vacuum.

Now there are two possibilities one has to distinguish. The symmetry that is broken may be a *global* or a *local* symmetry. The physics implication of these two cases is rather different. Consider first global symmetries

3.2 The Goldstone Theorem

A very important consequence of spontaneous breaking of global, continuous symmetries is the appearance of massless scalar fields in the spectrum. This is governed by the Goldstone Theorem. Classically, what we are considering are transformations of fields and/or coordinates that leave the classical action invariant, but that change some solution of the classical equation of motion. Quantum-mechanically we have a transformation S that commutes with the Hamiltonian, $[S, H] = 0$, but does not leave the vacuum invariant, $S\,|0\rangle \neq |0\rangle$.

One can easily develop some intuition for what is going on in the classical picture. A transformation S changes ϕ_{cl} to $\phi_{\text{cl}} + \delta_S\phi$. If S is a symmetry, this implies that the action, as well as the classical Hamiltonian, has the same value for ϕ_{cl} and $\phi_{\text{cl}} + \delta_S\phi$. Then $\delta_S\phi$ is a fluctuation that does not cost any energy. Hence it must be a massless fluctuation.

The Goldstone theorem states that for any independent generator of a spontaneously broken continuous symmetry there is a massless scalar in the spectrum. We will not demonstrate this here in full generality, but show it in a concrete example.

The example is a complex scalar field theory with Lagrangian

$$\mathcal{L}_{\text{scalar}} = (\partial_\mu\phi)^*\partial^\mu\phi - m^2\phi\phi^* \ . \tag{3.2}$$

We have given this field a mass, since otherwise it would be hard to detect the appearance of massless modes. Note that there are two real fields, the real and imaginary parts of ϕ, with mass m.

Before continuing, a remark on the normalization of complex and real fields. The normalization chosen above is the standard one for complex scalar fields. A real scalar field η with mass m would have the Lagrangian

$$\mathcal{L}_{\text{real scalar}} = \frac{1}{2}\partial_\mu\eta\partial^\mu\eta - \frac{1}{2}m^2\eta^2. \tag{3.3}$$

In both cases the normalization is such that the propagator of these fields is $\frac{i}{k^2-m^2}$.

The complex scalar theory defined in Eqn. (3.2) has a global $U(1)$ symmetry: if we multiply ϕ with a constant phase, then the action does not change. The equations of motion are

$$\partial^\mu \partial_\mu \phi = -m^2 \phi \tag{3.4}$$

We are looking for constant solutions, $\partial_\mu \phi = 0$, and obviously this implies that $\phi = 0$. So this is not very interesting.

We can make it more interesting by adding interactions

$$\mathcal{L}_{\text{scalar}} = (\partial_\mu \phi)^* \partial^\mu \phi - m^2 \phi \phi^* - \frac{1}{4}\lambda(\phi\phi^*)^2 \ . \tag{3.5}$$

This is often written as

$$\mathcal{L}_{\text{scalar}} = (\partial_\mu \phi)^* \partial^\mu \phi - V(\phi) \ , \tag{3.6}$$

and $V(\phi)$ is called the scalar potential. The Hamiltonian derived from this action is

$$H = \int d^3x \left[|(\partial_{\vec{x}} \phi)|^2 + |(\partial_t \phi)|^2 + m^2 \phi\phi^* + \frac{1}{4}\lambda(\phi\phi^*)^2 \right] \ . \tag{3.7}$$

For this to be bounded from below for large values of ϕ requires $\lambda > 0$ (where λ is real).

Now the equations of motion are

$$\partial^\mu \partial_\mu \phi = -m^2 \phi - \frac{1}{2}\lambda(\phi\phi^*)\phi \tag{3.8}$$

For both $m^2 > 0$ and $\lambda > 0$ this still has only one constant solution: $\phi = 0$. As explained above, there is good reason why λ should be positive, but not for m^2. This is just a parameter in a Lagrangian. The fact that we wrote it as a square is not a valid argument for positivity. This was done so that in the previous case we could interpret m as a mass. But we can forget about that here, and simply choose $m^2 < 0$. Then the condition

$$\partial^\mu \partial_\mu \phi = -m^2 \phi - \frac{1}{2}\lambda(\phi\phi^*)\phi \tag{3.9}$$

has a non-trivial solution for constant ϕ

$$\phi\phi^* = \frac{-2m^2}{\lambda} \equiv \frac{1}{2}v^2 \tag{3.10}$$

Note that there is an entire circle of vacua, because the solution does not depend on the phase. We just choose one of them, for example ϕ is real and positive. Then $\phi_{\text{cl}} = \frac{1}{\sqrt{2}}v$, with v as defined above. The factor $\frac{1}{\sqrt{2}}$ seems awkward, and it might appear more natural to define $\phi_{\text{cl}} = v$, but this normalization is more convenient for future purposes.

To quantize this theory we expand the field around the vacuum: $\phi = \phi_{\text{cl}} + \phi_{\text{qu}}$. We may do this as follows

$$\phi = \frac{1}{\sqrt{2}}(v + \eta + i\rho) \tag{3.11}$$

Now η and ρ are two real fluctuations, treated as new field variables. But there is a more clever way of expanding around the classical vacuum:

$$\phi = \frac{1}{\sqrt{2}}(v+\eta)e^{i\xi} \tag{3.12}$$

Expanding to first order, we see that the two expansions are related: $\rho = v\xi$. But if we substitute Eqn. (3.12) into the Lagrangian, we see that ξ disappears from the action, except in the (canonically normalized) kinetic terms:

$$\frac{1}{2}\left|(\partial_\mu \eta + i(v+\eta)(\partial_\mu \xi)\right|^2 \ . \tag{3.13}$$

Note that η does not vanish from the action. If we expand the scalar potential we get

$$V(\phi) = \frac{1}{2}m^2 v^2 + \frac{1}{16}\lambda v^4 + \frac{1}{2}m^2\eta^2 + \frac{3}{8}\lambda v^2\eta^2 + \frac{1}{4}\lambda v\eta^3 + \frac{1}{16}\lambda\eta^4 \tag{3.14}$$

Substituting Eqn. (3.10) we get

$$V(\phi) = -\frac{m^4}{\lambda} - m^2\eta^2 + \frac{1}{4}\lambda v\eta^3 + \frac{1}{16}\lambda\eta^4 \tag{3.15}$$

There is a quadratic term in the scalar potential defining a positive mass $\sqrt{-2m^2}$ for η (remember that $m^2 < 0$ and that there is a factor $\frac{1}{2}$ in the canonical definition of mass terms for real scalars). But the field ξ is massless. This is the Goldstone boson.

Observe that the expansion Eqn. (3.12) is only valid if $v \neq 0$. Hence it cannot be used to show that there is a massless mode if we expand around $\phi = 0$. We would not get quadratic kinetic terms for ξ, but rather something like $(\eta\partial_\mu\xi)^2$. Indeed, the expansion around $\phi = 0$ and $m^2 > 0$ just yields two massive modes with mass m.

3.3 Higgs Mechanism for Abelian Gauge Symmetry

Now consider the same example with a local instead of a global symmetry. This implies that ϕ is coupled to an abelian gauge field. The Lagrangian is

$$\mathcal{L}_{\text{scalar}} = (D_\mu\phi)^* D^\mu\phi \ . \tag{3.16}$$

where $D_\mu = \partial_\mu - ieA_\mu$. The gauge symmetry of this Lagrangian is $\phi \to e^{ie\Lambda}\phi$. Let us assume that ϕ has a v.e.v. equal to v, which we will take to be real. If we expand around $\phi = v$, the fluctuations will not have the gauge symmetry anymore, since v is fixed and does not transform. This is puzzling at first sight, because we had argued before that the gauge symmetry was essential for having a massless photon with two physical polarizations.

To see what happens we rewrite the Lagrangian as before, choosing

$$\phi(x) = \frac{1}{\sqrt{2}}(v+\eta(x))e^{i\xi(x)} \ , \tag{3.17}$$

so that η are the real fluctuations and ξ the imaginary ones. In the quantum theory the quanta of η and ξ will yield the fluctuations, and they will have the usual expansion in terms of oscillators, as in Eq. (3.1). Substituting (3.17) into the Lagrangian we get

$$\tfrac{1}{2}|(\partial_\mu \eta + i(v+\eta)(\partial_\mu \xi - eA_\mu)|^2 \ .$$

Now we replace everywhere A_μ by

$$B_\mu = A_\mu - \frac{1}{e}\partial_\mu \xi \tag{3.18}$$

Now the Lagrangian becomes

$$\tfrac{1}{2}|(\partial_\mu \eta - i(v+\eta)eB_\mu)|^2 \ . \tag{3.19}$$

Expanding this yields

$$\tfrac{1}{2}\partial_\mu \eta \partial^\mu \eta + \tfrac{1}{2}e^2 v^2 B_\mu B^\mu + \tfrac{1}{2}e^2 B_\mu B^\mu \eta(2v+\eta) \ . \tag{3.20}$$

Now suppose that there are other terms in the Lagrangian in addition to (3.16). This includes in particular the kinetic terms for the gauge bosons. All the additional terms must be gauge invariant. The replacement $B_\mu = A_\mu - \frac{1}{e}\partial_\mu \xi$ can be realized on all other terms as a gauge transformation, which may include ξ-dependent transformations of other fields. This implies that the other terms in the Lagrangian remain unchanged, except that A_μ is replaced everywhere by B_μ.

To summarize, suppose we started with a Lagrangian

$$-\tfrac{1}{4}F_{\mu\nu}(A)F^{\mu\nu}(A) + \mathcal{L}_{\text{scalar}} + \mathcal{L}_{\text{rest}}(A) \ . \tag{3.21}$$

Then after shifting the vacuum and some changes of variables we end up with

$$-\tfrac{1}{4}F_{\mu\nu}(B)F^{\mu\nu}(B) + \tfrac{1}{2}\partial_\mu \eta \partial^\mu \eta + \tfrac{1}{2}e^2 v^2 B_\mu B^\mu + \tfrac{1}{2}e^2 B_\mu B^\mu \eta(2v+\eta) + \mathcal{L}_{\text{rest}}(B) \ . \tag{3.22}$$

Two observations can now be made:

- The field ϕ had two real degrees of freedom, η and ξ, but the latter has disappeared completely.
- The quadratic term in B_μ gives a mass ev to the vector field.

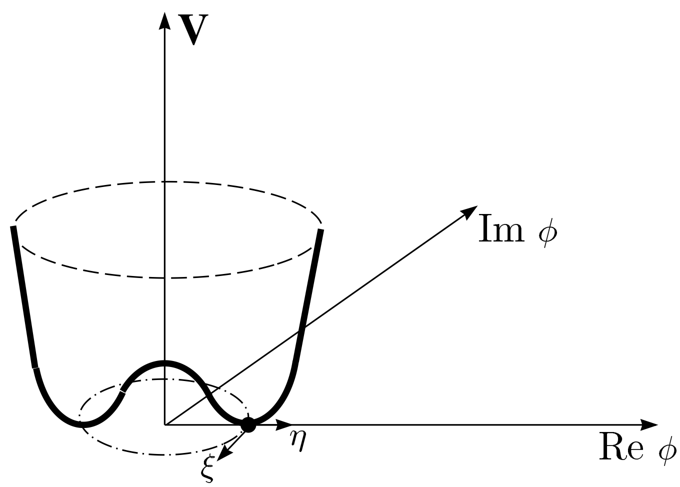
This magic is called the *Higgs mechanism*, after one of its inventors. It allows us to give a mass to the gauge boson, simultaneously breaking the gauge symmetry. The field ξ is not really gone. As we have seen, a massive gauge boson has three degrees of freedom, a massless one only two. When we made the transformation $B_\mu = A_\mu - \frac{1}{e}\partial_\mu \xi$ we have absorbed ξ into the gauge field to provide the extra degree of freedom needed to make it massive. One often says that ξ was "eaten" by the gauge field.

Massive vector bosons occur in the theory of weak interactions, the $W^\pm$ and Z bosons. You may wonder why we couldn't simply have added the mass term by hand. The reason is that such a procedure makes the theory inconsistent. It explicitly breaks gauge invariance, and gauge invariance is essential for consistency of theories with spin-1 particles. In the procedure explained above gauge invariance is not manifest anymore in the shifted ground state, but it is still present in a less obvious form.

3.4 The Mexican Hat Potential

In the previous section we simply put in a vacuum value v by hand, and examined its effect on the gauge boson masses. In section 3.2 we saw how to get such a v.e.v. from a suitable potential. Combining these results we got not only a massive vector boson, but also a massive scalar η. The latter corresponds to the Higgs scalar that has recently been discovered at CERN, although we still have to build the mechanism into the Standard Model.

As observed before, we have a continuous set of ground states to choose from, and above we have chosen just one of them, $v = \sqrt{\frac{-2m^2}{\lambda}}$ real. The continuum is illustrated in the following picture of the potential as a function of ϕ



The black dot indicates our choice of the ground state, but any choice on the bottom of the "Mexican hat" would have been fine as well. By making a choice we break the gauge symmetry, *i.e.* the phase rotations of ϕ. We have also indicated the directions of the small perturbations η and ξ.

If one shifts the value of ϕ one finds that η now gets a mass, and ξ disappears, as before. The mass of η, the Higgs boson, is a free parameter, and is in principle unrelated to the mass of the vector boson, ev.

Observe that in order to find a non-trivial ground state we had to take $m^2 < 0$. If one expands around the trivial ground state $\phi = 0$ this negative value of m^2 leads to trouble: the theory now contains particles with imaginary mass. Such particles are also known as "tachyons" because their velocity, given by the relativistic formula $(\frac{v}{c})^2 = \vec{p}^2/(\vec{p}^2 + m^2)$, can exceed the velocity of light. The presence of tachyons means that the theory with this vacuum choice is sick. Classically this is related to the fact that a field configuration on top of the hill inside the Mexican hat is unstable. The only correct vacuum is the non-trivial one, and all fluctuations around it have positive or zero mass. If on the other hand $m^2 > 0$ the only vacuum is $\phi = 0$. By continuously changing m^2 we can go from the symmetric vacuum $\phi = 0$ to the vacuum with broken gauge symmetry. Historically, this is what led to the name "spontaneous symmetry breaking".

All the above can be generalized to non-abelian gauge theories. The main features are the same. Some symmetries within a symmetry group are spontaneously broken, and the corresponding vector bosons acquire a mass by each "eating" a scalar. The resulting spectrum always contains (at least) one Higgs scalar, whose mass is a free parameter and hence cannot be predicted.

4 The Standard Model

Before trying to look beyond the Standard Model, let us examine it more closely, paying special attention to those features that might be relevant for attempts to go beyond it.

4.1 QED and QCD

At the lowest energies its exact gauge symmetries are $SU(3) \times U(1)$. The gauge fields are described by the Lagrangian (for conventions see the beginning of these notes).

$$\mathcal{L}_1 = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{4} G^I_{\mu\nu} G^{\mu\nu,I} \tag{4.1}$$

where $I = 1 \ldots 8$ labels the gluons. We denote the field strength tensor of QCD here as $G^I_{\mu\nu}$ They couple to six quark flavors, each in a triplet representation of $SU(3)$.

$$\mathcal{L}_2 = i\bar{\psi}^i \gamma^\mu (\partial_\mu - ig_3 A^I_\mu T^I) \psi^i \ , \tag{4.2}$$

where the implicit sum on i is over the six quark "flavors" u, d, c, s, b and t. These are all the quarks we know, and there are reasons to believe that there are no more. The hermitean matrix T^I is the $SU(3)$ generator in the triplet representation. Color indices of the quarks have been suppressed. The parameter g_3 is the QCD coupling constant.

Note that the Lagrangian $\mathcal{L}_1 + \mathcal{L}_2$ has an exact $U(6)$ symmetry $\psi_i \rightarrow U_{ij}\psi_j$. In fact it has an even larger symmetry since each quark has both left- and right-handed components which can be rotated completely independently. We can write the fermion Lagrangian as

$$\mathcal{L}_2 = i\bar{\psi}^i_L \gamma^\mu (\partial_\mu - ig_3 A^I_\mu T^I) \psi^i_L + i\bar{\psi}^i_R \gamma^\mu (\partial_\mu - ig_3 A^I_\mu T^I) \psi^i_R \ , \tag{4.3}$$

where $\psi_L = \frac{1}{2}(\mathbf{1} + \gamma_5)\psi$, $\psi_R = \frac{1}{2}(\mathbf{1} - \gamma_5)\psi$, which tell us that in fact there is an $U(6)_L \times U(6)_R$ symmetry. This is called *chiral symmetry*, and it is one of the most interesting and important aspects of the Standard Model.

In nature this symmetry is broken for at least six different reasons, as we will see. The first and most obvious one is that the six quarks have six completely different masses. Quark masses are hard to define theoretically, since quarks can only be observed inside hadrons and never as free particles. Therefore they are also difficult to measure experimentally. The particle data group [8] gives the following results*.

$$\begin{aligned} m_u &= 2.2 \pm .6 \text{ MeV} \\ m_d &= 4.7 \pm .5 \text{ MeV} \\ m_s &= 96 \pm 8 \text{ MeV} \\ m_c &= 1.27 \pm .03 \text{ GeV} \\ m_b &= 4.1 \ldots 4.7 \text{ GeV} \\ m_t &= 173 \pm 1 \text{ GeV} \end{aligned}$$

These masses are taken into account by adding the following terms to the Lagrangian

$$\mathcal{L}_3 = -\sum_i m_i \bar{\psi}^i \psi^i \ . \tag{4.4}$$

The fact that all masses are different implies a breaking from $U(6)$ to $U(1)^6$, but even if all masses were all equal these terms link the left and right-handed fermions and do not allow us to rotate them independently:

$$\mathcal{L}_3 = -\sum_i m_i (\bar{\psi}^i_L \psi^i_R + \bar{\psi}^i_R \psi^i_L) \ . \tag{4.5}$$

If all m_i were the same this would break the symmetry to $U(6)_V$, the vector symmetry, which acts by rotation ψ_L and ψ_R in the same way. The "orthogonal" combination, rotating ψ_L by U and ψ_R by U^{-1} is called the axial symmetry. The combination of these two effects (the existence of quark masses and their differences) leaves us with the global symmetry $(U(1)^6)_V$, which are the six separate flavor quantum numbers. They are conserved because QCD is flavor blind.

The third reason why $U(6)_L \times U(6)_R$ is broken (even if all quark masses were zero) is the coupling of the quarks to electromagnetism. This coupling adds the following terms to the Lagrangian

$$\mathcal{L}_4 = -i\bar{\psi}^i \gamma^\mu (ieqA_\mu)\psi^i \ , \tag{4.6}$$

where $q = \frac{2}{3}$ for the quarks u, c and t and $q = -\frac{1}{3}$ for d, s and b. QED is not flavor blind, but does not mix flavors, so that they are still conserved.

Also the leptons e, μ and τ, coupling with $q = -1$, need to be added now. The lepton part of the QED Lagrangian has a global $U(3)_L \times U(3)_R$ symmetry, which is broken down to $(U(1)^3)_V$ if we also add mass terms for the leptons. Their masses are known much more precisely than those of the quarks:

$$\begin{aligned} m_e &= 0.510998928 \pm .000000011 \text{ MeV} \\ m_\mu &= 105.658357 \pm .000002 \text{ MeV} \\ m_\tau &= 1776.82 \pm .16 \text{ MeV} \end{aligned}$$

* Using 2016 data, errors rounded off; note that there are two incompatible definitions of the b quark mass, hence the large range

The generators of the three $U(1)$'s that survive after masses are added are the three separate lepton numbers. The remaining fermions that we know are three species of neutrinos, but since they couple neither to QCD nor QED they will make their appearance later.

4.1.1 Chiral Symmetry Breaking

We have seen already three effects that break the chiral $U(6)_L \times U(6)_R$ symmetries, namely the presence of quark masses, the differences in the quark masses, and the electromagnetic interactions. But even if we switch off all these effects the chiral symmetries are broken. Switching off electromagnetism is quite a good approximation to the real world, since electromagnetic mass differences are only a few MeV, much less than the hadron masses themselves. Switching off the quark masses appearing in the QCD Lagrangian is a good approximation for u and d quarks, whose masses are a few MeV as well. If one could vary the parameters of the QCD Lagrangian (as one can in lattice simulations), and one could set the quark masses equal to zero, then one would find that the pion masses go to zero, while all other states in the hadron spectrum remain at some non-vanishing mass.

If we focus on these u and d quarks only, we might expect the bound state spectrum to show the same symmetry the quarks have in the limit of vanishing mass: an $U(2) \times U(2)$ chiral symmetry. However, the hadron spectrum only has a quite good $SU(2)_V$ symmetry called isospin, plus a $U(1)_V$ symmetry, baryon number. In the limit of vanishing quark mass, and if we switch off all electromagnetic (and weak) interactions, isospin is a perfect symmetry. In this limit, the proton and the neutron have equal masses, and the proton-neutron system can be effectively described by the Lagrangian

$$i \sum_{x=n,p} \bar{\psi}^x \gamma^\mu \partial_\mu \psi^x - M \sum_{x=n,p} \bar{\psi}^x \psi^x \tag{4.7}$$

This clearly has an exact $SU(2)_V$ symmetry that acts on the label x, that distinguishes protons and neutrons. The $SU(2)_A$ axial symmetry, however, is not realized in the spectrum. This symmetry rotates the left- and right-handed components of the baryon spinor ψ_x in opposite ways (if the left component is transformed with a 2×2 matrix U, the right one transforms with $U^\dagger$), and this is not a symmetry of the Lagrangian because the mass term couples the left and the right component. It is essential to know that $M \neq 0$, even if m_u and m_d vanish, and this fact is know for example from the aforementioned theoretical extrapolations.

Here we see an example of a symmetry of the Lagrangian that is not realized in the spectrum. This phenomenon is known as "spontaneous symmetry breaking". In general, there are two requirements for some infinitesimal operator T to be a symmetry of a physical system: T must commute with the Hamiltonian, $[H, T] = 0$, and T must annihilate the vacuum, $T\,|0\rangle = 0$ (if the infinitesimal symmetry operator annihilates the vacuum, its global form will leave the vacuum invariant).

The fact that $SU(2)_A$ is not realized in the baryon spectrum is understood as a result of a spontaneous symmetry breaking, which is dynamically generated by QCD. In other words, the QCD vacuum is not invariant under the $SU(2)_A$ transformations. There is a famous theorem, known as Goldstone's theorem, that applies to such a situation. The theorem states that such a symmetry breaking results in massless scalars in the spectrum, transforming like the derivative of the current of the broken symmetry. These particles are the pions, which indeed are quite light in comparison to other hadrons, but which are not completely massless. The reason is that because of the mass terms in the Lagrangian (sometimes called the "current quark masses") the $SU(2)_A$ symmetry was not exact to begin with, and hence the Goldstone bosons are only approximately massless. Often such particles are called pseudo-Goldstone bosons.

This is the fourth reason why the chiral symmetries are broken. In the process, QCD gives the quarks an effective mass of the order of one-third of the proton or Δ mass, which is called the "constituent mass" since it can be viewed as the mass of quarks as constituents of hadrons. The current mass is the relevant one in hard scattering, where soft QCD effect can be ignored.

It would be natural to expect a fourth Goldstone particle because the axial symmetry that is spontaneously broken is $U(2)$ and not just $SU(2)$. We will discuss later what happens to the extra $U(1)_A$ symmetry. If N flavors are present the mechanism extends straightforwardly from $SU(2)$ to $SU(N)$. In reality the masses of the other quarks can not be neglected, however, and hence this description becomes less useful.

Although this intuitive picture is appealing and leads to qualitatively and quantitatively satisfactory results, it has not been derived rigorously from QCD. However it is supported by lattice calculations.

4.1.2 The θ-parameter

There are still more terms one can add to this Lagrangian without destroying gauge invariance or renormalizability, namely

$$\theta \frac{g_3^2}{32\pi^2} \sum_{I=1}^{8} G_{\mu\nu}^I \tilde{G}^{\mu\nu,I} = \theta \frac{g_3^2}{16\pi^2} \text{Tr } G_{\mu\nu}\tilde{G}^{\mu\nu} \ , \tag{4.8}$$

where $\tilde{G}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}G^{\rho\sigma}$ and $G_{\mu\nu} \equiv G_{\mu\nu}^I T^I$. Here we used the relation

$$\text{Tr } T^I T^J = \tfrac{1}{2}\delta^{IJ}$$

which als defines the normalization of the $SU(3)$ generators T^I. This is the standard normalization, *c.f.* Eq. (2.42). This term is of the same order in fields and derivatives as the gauge kinetic terms. Hence it has mass dimension 4. We will see later that terms of higher order than 4 can be consistently dropped from the Lagrangian, because they have a coupling constant with dimension $[\text{mass}]^{-1}$. By assuming that the corresponding mass scale is as large as we want, we can always make such terms arbitrarily small in a natural way.

But since the $G\tilde{G}$ term has mass dimension 4, the parameter θ is dimensionless. It turns out that θ is like an angle: all physics is periodic in θ. The factor g_3^2 and the normalization are chosen in such a way that that the periodicity of θ is 2π.

The term (4.8) explicitly violates parity P, but respects charge conjugation C, and hence it also violates CP. To see why it violates parity note that the ϵ-tensor transforms under Lorentz-transformation Λ^{μ}_{ν} to $\Lambda^{\mu}_{\mu'}\Lambda^{\nu}_{\nu'}\Lambda^{\rho}_{\rho'}\Lambda^{\sigma}_{\sigma'}\epsilon^{\mu'\nu'\rho'\sigma'} = \det(\Lambda)\epsilon^{\mu\nu\rho\sigma}$. This implies in particular that (4.8) is indeed Lorentz invariant. But the determinant is negative for space inversion $\vec{x} \to -\vec{x}$ and also under time reversal. This is consistent with CPT-invariance: if CP is violated, then T must be violated as well.

Topogical considerations. A noteworthy feature of this term is that it is a total derivative. Writing $G_{\mu\nu} = G^I_{\mu\nu}T^I$ one has

$$\frac{1}{4}\,\mathrm{Tr}G_{\mu\nu}\tilde{G}^{\mu\nu} = \partial_\mu K^\mu\ , \tag{4.9}$$

where

$$K^\mu = \epsilon^{\mu\nu\rho\sigma}\ \mathrm{Tr}[A_\nu\partial_\rho A_\sigma + \tfrac{2}{3}g_3 A_\nu A_\rho A_\sigma] \tag{4.10}$$

Normally one would drop such total derivative terms from the Lagrangian. However one has to be careful with boundary terms. It turns out that in non-abelian gauge theories there exist field configurations with finite (Euclidean) action for which the boundary integral on S_3 at infinite radius does not vanish. These are called *instantons*. They are characterized by an integral over all of Euclidean space that is always an integer

$$\mathcal{N}_{\mathrm{E}} = \frac{g_3^2}{16\pi^2}\int_{\mathrm{E}} d^4x \mathrm{Tr}\ G_{\mu\nu}\tilde{G}^{\mu\nu} = n \in \mathbb{Z}$$

Note that this looks very much the action (4.8), but the latter is of course defined in Minkowski space. The fact that this Euclidean integral is quantized is the reason that θ is periodic. This can intuitively be understood as follows (the following discussion assumes a basic understanding of path integral in quantum field theory). When going to Euclidean space the integration measure $\int_{\mathrm{M}} d^4x$ is changed to $i\int_{\mathrm{E}} d^4x$, where 'M' and 'E' denote Minkowski and Euclidean respectively. This turns the integrand into a negative exponential, dominated by the classical paths, and with exponential suppression for paths that deviate from it. But the θ term behaves a bit differently. Within the action, all contractions involving a g_μ are changed to Euclidean contractions involving $\delta_{\mu\nu}$. But terms involving a Levi-Civita tensor $\epsilon^{\mu\nu\rho\sigma}$ change by a factor i, because there is just one time component in every non-vanishing tensor. Hence schematically we get the following path intregrand for the path integral

$$e^{i(S_{\mathrm{M}}+\theta\mathcal{N}_{\mathrm{M}})} \to e^{-S_{\mathrm{E}}+i\theta\mathcal{N}_{\mathrm{E}}}$$

Since $\mathcal{N}_E$ is always an integral, we see that the Euclidean path integral is indeed periodic in θ. The underlying physics requires a lot more discussion, but that is beyond the scope of these lecture notes.

Theories with values of θ that differ by multiples of 2π are related by gauge transformations. These are not the local gauge transformations shown in eqn. (2.33), where the local parameters ξ^a are implicitly assumed to fall off rapidly towards infinity, but they are gauge transformations that do contribute to the boundary terms (4.10). Because such configurations exist and contribute to the path integral one cannot simply drop the total derivatives in the action. In QED there are no such configurations, and the CP-violating terms may indeed be dropped; it has no observable consequences.

Observable Consequences. All terms in the Lagrangian that we have seen before give rise to a vertex that can be used in perturbation theory. So it would be natural to construct the vertex corresponding to (4.8). But one finds that there is no such vertex. This is due to the fact that (4.8) is a total derivative. So we will never see the effect of (4.8) in any Feynman diagram. But QCD is more than just Feynman diagrams. There are contribution to physical processes that cannot be obtained by means of Feynman diagrams. These are called non-perturbative effects.

In QCD the term (4.8) does indeed have observable consequences, as it contributes to the electric dipole moment of the neutron, a CP-violating quantity. To see how electric and magnetic moments transform under P and CP it is most convenient to assume CPT, and consider T-invariance instead of CP invariance. This is because C changes quarks into anti-quarks, and hence it is a bit cumbersome to derive CP transformations of the neutron using C and P directly. Under parity, electric dipole moments flip. Magnetic dipole moments transform as $r \times p$, *i.e.* angular momenta, and hence they are invariant under parity. Under time reversal, magnetic moments flip (because magnetic moments are due to, or behave like, rotating charges, and under time reversal all rotations reverse directions), whereas electric dipole moments remain unchanged. Hence under either P or T the electric and magnetic dipole moment flip with respect to each other, and hence if a particle has both moments, the final state of the transformation must be different from its initial state. Since the neutron has a magnetic dipole moment, if it also had an electric one, this would violate both P and CP.

The electric dipole moment of the neutron, d_n, has not been observed. The current experimental limit is $|d_n| < 2.9 \times 10^{-26}e\text{cm}$, which puts a bound on θ: $\theta < 4 \times 10^{-10}$. The electric dipole moment of the neutron is approximately given by

$$\frac{e}{m_n}\theta\frac{m_u m_d}{m_u + m_d}\frac{1}{\Lambda_{\text{QCD}}} \tag{4.11}$$

where Λ_{QCD} is the QCD scale, and m_n the neutron mass. We have seen above that θ is like an angle, and hence its full parameter space is the interval $[0, 2\pi)$. It could have take any value in this interval, but nature has chosen it to be remarkably close to 0. To appreciate the point, define a new parameter $x = \theta/2$, and suppose the value of x were experimentally determined to be $3.1415926536 \pm 10^{-10}$. Wouldn't you think that this is remarkably close to π, and that this cannot be a coincidence? But $x = \pi$ is physically equivalent to $x = 0$, and hence this is essentially what we observe.

The Strong CP Problem. The fact that the angle θ is so close to zero seems to demand an explanation. This problem is called *the strong CP problem.* A first idea could be to simply declare that CP is a symmetry of the strong and electromagnetic interactions. Indeed, since all other terms in the $SU(3) \times U(1)$ Lagrangian respect CP the term (4.8) cannot be generated if it is set to zero. This is an important lesson, which will come back several times in these lectures: terms can be consistently removed from a Lagrangian if their removal leads to an enhanced symmetry. In this situation one says that the absence of such a term is "natural".

But note that the absence of P and CP violation is a property of the strong and electromagnetic interactions, but not a general of property nature, since the weak interactions do not respect these symmetries. Hence after switching on the weak interactions we *do* have to worry about this term.

Indeed, in the presence of CP-violating Yukawa couplings the discussion is rather different. It turns out that phase rotations of the quark masses, in order to make them real, end up changing θ. The experimental limit is in fact not on θ but on a parameter $\bar{\theta}$, which is the difference between θ and an overall phase in the quark mass matrix. Only this difference is observable. This will be discussed in more detail in section 5.6.

Even if one somehow manages to make $\bar{\theta}$ exactly zero in the Lagrangian, this still does not mean that $d_n = 0$. Weak interactions still make contributions to d_n of order $10^{-31}e$cm. That is about five orders of magnitude smaller than our current limits, but it is essentally inevitable that such an effect exists in the Standard Model. New physics, such as low energy supersymmetry, can make contributions as large as 10^{-25} e.cm, and hence current experiments are already constraining these options.

4.2 The Weak Interactions

The complete Lagrangian for the weak interactions *after* symmetry breaking would occupy several pages, and we will not present it here. However at high energies, in unbroken form, it is much simpler. Interestingly above the symmetry breaking scale we do not only gain symmetries, but we also loose some, namely C, P and T. The $U(1)$ gauge group of QED now becomes part of a larger group $SU(2) \times U(1)$. Following tradition we denote the gauge bosons as A^a_μ, $a = 1, \ldots, 3$, and B_μ. The action for the gauge fields is the canonical one, Eq. (2.40). The $SU(3)$ gauge group of QCD is not involved in the weak interactions.

4.2.1 Fermion Representations

The Standard Model fermions are in the $SU(3) \times SU(2) \times U(1)$ representations

$$(3, 2, \frac{1}{6})_L,\ (3, 1, \frac{2}{3})_R,\ (3, 1, -\frac{1}{3})_R,\ (1, 2, -\frac{1}{2})_L,\ (1, 1, -1)_R \text{ and } (1, 1, 0)_R.$$

The corresponding fermion fields are denoted

$$\psi^{\mathcal{Q}}_L,\ \psi^{\mathcal{U}}_R,\ \psi^{\mathcal{D}}_R,\ \psi^{\mathcal{L}}_L,\ \psi^{\mathcal{E}}_R \text{ and } \psi^{\mathcal{N}}_R$$

respectively. Here $\mathcal{Q}$ stands for quark, $\mathcal{L}$ for lepton, $\mathcal{U}$ for charge $\frac{2}{3}$ quarks, $\mathcal{D}$ for charge $-\frac{1}{3}$ quarks, $\mathcal{E}$ for leptons of charge -1, and $\mathcal{N}$ for neutrinos. Here we are using a bit of foresight regarding the final interpretation of these representations. These fields couple to the gauge fields as indicated by their representations, and we will need three copies of each to get the three fermion families observed experimentally.

The precise form of this coupling is as follows

$$\mathcal{L}_{\mathcal{Q}} = i\sum_{\alpha=1}^{3} \bar{\psi}_L^{\mathcal{Q},\alpha}\gamma^\mu(\partial_\mu - ig_3 A_\mu^I T^I - ig_2 A_\mu^a T^a - ig_1 B_\mu Y)\psi_L^{\mathcal{Q},\alpha} , \tag{4.12}$$

where T^I is an $SU(3)$ color generator, $T^a \equiv \frac{1}{2}\sigma^a$ is an $SU(2)$ generator, and Y is the generator of the $U(1)$ factor, whose eigenvalue is $\frac{1}{6}$ for $\psi_L^{\mathcal{Q},\alpha}$. The label α distinguishes the three different quark and leptons families (note that "families" are sometimes called "generations" in the literature). The coupling of the other fermions to the gauge bosons works analogously, with $T^I = 0$ for color singlets and $T^a = 0$ for $SU(2)$ singlets. The $U(1)$ and $SU(2)$ coupling constants are denoted g_1 and g_2.

Note that the fermion $\psi_R^{\mathcal{N}}$ (the right-handed neutrino field) does not couple to any of the gauge bosons. This means that it can only be observed gravitationally, since gravity couples to everything. However, as we will see, it may couple to the rest of the Standard Model via the Higgs field, to be introduced next. The existence of right-handed neutrinos has not been demonstrated experimentally yet. If they do exist, there is no good reason why their multiplicity should be three, as is the case for the other five multiplets.

4.2.2 The Higgs Field

As the final ingredient in the Standard Model we introduce a complex scalar ϕ in the representation $(1, 2, \frac{1}{2})$. If we require its Lagrangian to be gauge invariant and of renormalizable type the most general form is

$$\mathcal{L}_\phi = (D_\mu\phi)^\dagger(D^\mu\phi) - \mu^2\phi^\dagger\phi - \tfrac{1}{4}\lambda(\phi^\dagger\phi)^2 , \tag{4.13}$$

where $D_\mu = \partial_\mu - \frac{1}{2}ig_1 B_\mu - ig_2(\frac{1}{2}\sigma^a)A_\mu^a$ is the covariant derivative. This scalar field ϕ is called the Higgs field. Suppose that for some unknown reason the scalar mass μ^2 is negative. This may seem strange, but at this point μ^2 is just a parameter in the Lagrangian. By writing it as a square we were incorrectly suggesting that it must be positive, but actually μ^2 may have any real value. One may ask the question if the sign ultimately requires further explanation, but that explanation is in any case "beyond the Standard Model", and we will not worry about that in this section. If $\mu^2 < 0$, the true minimum of the potential is not $\phi = 0$, but some non-trivial value, which by $SU(2)$ rotations we can bring to the form

$$<\phi>= \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix} , \tag{4.14}$$

and which we can make real by $U(1)$ transformations (the normalization is a convention). The minimum of the potential is at $v = 2\sqrt{-\mu^2/\lambda}$.

4.2.3 Vector Boson Masses

We now expand ϕ around its classical value $<\phi>$, *i.e.* $\phi =<\phi> + \ldots$, but we will ignore the extra terms for the moment. The constant term introduces, via the covariant derivative terms, a mass matrix for the vector bosons A^a_μ and B_μ. Introducing a vector $V^i_\mu = (A^1_\mu, A^2_\mu, A^3_\mu, B_\mu)$, we find the following form for these mass terms

$$\mathcal{L}_{\rm mass} = \tfrac{1}{2} V^i_\mu (M^2)^{ij} V^{\mu,j} \tag{4.15}$$

The matrix is

$$M^2 = \frac{1}{4} v^2 \begin{pmatrix} g_2^2 & 0 & 0 & 0 \\ 0 & g_2^2 & 0 & 0 \\ 0 & 0 & g_2^2 & -g_1 g_2 \\ 0 & 0 & -g_1 g_2 & g_1^2 \end{pmatrix} . \tag{4.16}$$

(The minus sign of the off-diagonal terms is due to the fact that σ^3 acts on $<\phi>$ via its lower component.)

The mass matrix has off-diagonal terms, which means that the original vector bosons A^3_μ and B_μ mix. To find the mass eigenstates we must diagonalize the matrix M. The correct form of the bi-linear terms in the Lagrangian for a real vector field X_μ is

$$\mathcal{L}_{\rm massive\ real\ vector} = -\tfrac{1}{2}\partial_\mu X_\nu \partial^\mu X^\nu + \tfrac{1}{2}\partial_\mu X_\nu \partial^\nu X^\mu + \tfrac{1}{2} M_X^2 X_\mu X^\nu \tag{4.17}$$

For a conjugate pair of complex vectors $X^\pm_\mu$ this is

$$\mathcal{L}_{\rm massive\ complex\ vector} = -\partial_\mu X^+_\nu \partial^\mu X^{\nu,-} + \partial_\mu X^+_\nu \partial^\nu X^{\mu,-} + M_X^2 X^+_\mu X^{\nu,-} \tag{4.18}$$

After diagonalization we find that three of the four gauge bosons have acquired a mass, namely

$$\begin{array}{rclll} W^\pm_\mu &=& \frac{1}{\sqrt{2}}(A^1_\mu \mp i A^2_\mu) & \text{mass} & \frac{1}{2} g_2 v \\ Z_\mu &=& \frac{1}{\sqrt{g_1^2+g_2^2}}\,(g_2 A^3_\mu - g_1 B_\mu) & \text{mass} & \frac{1}{2}\sqrt{g_1^2+g_2^2}\,v \\ A_\mu &=& \frac{1}{\sqrt{g_1^2+g_2^2}}\,(g_1 A^3_\mu + g_2 B_\mu) & \text{mass} & 0 \qquad\qquad \text{(photon)} \end{array}$$

We may now express the coupling of the fermions to the gauge fields A^a_μ and B_μ in terms of the new fields $W^\pm_\mu$, Z_μ and A_μ (the coupling to the gluon is of course not affected). For the field $\psi^{\mathcal{Q}}_L$ this results in

$$\begin{aligned} \mathcal{L}_{\mathcal{Q}} = i \sum_{\alpha=1}^{3} \bar\psi^{\mathcal{Q},\alpha}_L \gamma^\mu \quad (\quad & \partial_\mu - i g_3 A^I_\mu T^I - i g_2 \frac{1}{\sqrt{2}}(W^+_\mu T^- + W^-_\mu T^+) \\ & - i \frac{g_1 g_2}{\sqrt{g_1^2+g_2^2}} A_\mu (T_3 + Y) - i \frac{g_2^2 T^3 - g_1^2 Y}{\sqrt{g_1^2+g_2^2}} Z_\mu) \psi^{\mathcal{Q},\alpha}_L \ , \end{aligned}$$

where $T^\pm = T^1 \pm i T^2$. The expressions for the other fermion fields are analogous.

4.2.4 Electromagnetism

The photon is found to couple to the fermions through the operator

$$Q_{\text{em}} = T_3 + Y \ , \tag{4.19}$$

where Y denotes the $U(1)$ generator before symmetry breaking. The reason the photon remains massless is that this generator annihilates the new vacuum:

$$(T_3 + Y) < \phi >= 0 \ , \tag{4.20}$$

and hence Q_{em} is an exact local symmetry of the theory. The electromagnetic coupling constant is found to be

$$e \equiv \frac{g_1 g_2}{\sqrt{g_1^2 + g_2^2}} \tag{4.21}$$

4.2.5 The Low-energy Spectrum

Let us now compute the fermion quantum numbers in the new vacuum. We find that the original $SU(3) \times SU(2) \times U(1)_Y$ representations decompose in the following way to $SU(3) \times U(1)_{\text{em}}$ representations

$$\begin{aligned}
(3, 2, \tfrac{1}{6})_L &\rightarrow (3, \tfrac{2}{3})_L + (3, -\tfrac{1}{3})_L \\
(3, 1, \tfrac{2}{3})_R &\rightarrow (3, \tfrac{2}{3})_R \\
(3, 1, -\tfrac{1}{3})_R &\rightarrow (3, -\tfrac{1}{3})_R \\
(1, 2, -\tfrac{1}{2})_L &\rightarrow (1, -1)_L + (1, 0)_L \\
(1, 1, -1)_R &\rightarrow (1, -1)_R \\
(1, 1, 0)_R &\rightarrow (1, 0)_R \ ,
\end{aligned}$$

and of course we get three copies of each fields. We denote these fields as $\psi_L^{\mathcal{U}}, \psi_L^{\mathcal{D}}, \psi_L^{\mathcal{E}}$ and $\psi_L^{\mathcal{N}}$ and similarly for the right-handed components. Just as before for the $SU(3) \times SU(2) \times U(1)$ representations, $\mathcal{U}$ stands for the three quarks u, c, t with charge $\frac{2}{3}$, $\mathcal{D}$ for the quarks d, s, b with charge $-\frac{1}{3}$, $\mathcal{E}$ for the leptons e, μ, τ with charge -1 and $\mathcal{N}$ for the three neutrinos.

Until 1998 most data were consistent with massless, purely left-handed neutrinos.* In the zero mass limit the right-handed neutrino decouples completely from all other fields in the Standard Model, and couples only to gravity. For this reason the existence of the right-handed neutrino components has been a matter of speculation. Nowadays we know that there must be mass differences between different neutrino species, and hence they cannot all be massless. We still do not know for sure if right-handed neutrinos exists, but it is the simplest possibility to explain the observed neutrino oscillations.

---

* The precise definition of left-handed is that the spin is oriented opposite to the direction of motion. This definition is convention-independent; however in the literature the corresponding projections are either $\frac{1}{2}(1 + \gamma_5)$ (our convention) or $\frac{1}{2}(1 - \gamma_5)$, and the definitions of γ_5 and the $\epsilon^{\mu\nu\rho\sigma}$ tensor may also differ by signs. If the neutrino is exactly massless this relative orientation is Lorentz-invariant.

4.2.6 Parameters

An often used parameter is $\tan\theta_{\rm w} = g_1/g_2$. The electromagnetic coupling constant e is then related to g_1 and g_2 as $e = g_2 \sin\theta_{\rm w} = g_1 \cos\theta_{\rm w}$. Experimental data are usually quoted in terms of $\sin^2\theta_{\rm w}$. The measured value is $.23117 \pm .00016$. The measured Z and W masses are $80.385 \pm .015$ and $91.1876 \pm .0021$. Using this experimental information we can compute the value of v, the Higgs v.e.v:

$$v = M_{\rm w}\sqrt{\frac{\sin^2\theta_{\rm w}}{\pi\alpha(M_{\rm w})}} \approx 246 \text{ GeV} \tag{4.22}$$

Here α is the QED fine structure constant, but one should not use the low energy value $\frac{1}{137}$, but the value at the mass of the W (or Z) boson, which is about $\frac{1}{128}$ (more about running coupling constants follows later).

4.2.7 The Higgs Boson

The scalar ϕ had originally four real (two complex) components. After symmetry breaking three of those four become the longitudinal components needed for the massive $W^\pm$ and Z vector bosons. The fourth one, the real field η which represents the component of ϕ in the direction of the vacuum expectation value appears in the spectrum as a scalar with mass $\sqrt{-2\mu^2}$. Its complete Lagrangian can be found by expanding $\phi(x)$ as

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + \eta(x) \end{pmatrix} \tag{4.23}$$

The purpose of the factor $\frac{1}{\sqrt{2}}$ is to make sure that the field η has the correct kinetic terms for a real scalar namely $\frac{1}{2}\partial_\mu\eta\partial^\mu\eta$. The complex field ϕ has no factor $\frac{1}{2}$ in its kinetic terms (see (4.13)). This is the last particle of the Standard Model that has been discovered, the famous Higgs boson. It was searched for during several decades, and for a long time it was also the only one that was missing. With its discovery the Standard Model is complete. This does not mean that it is correct and that it is certain to survive future experiments, but only that the list of definite particles still searched for is now empty. The Higgs boson was discovered using the ATLAS and CMS detectors and the LHC accelerator at CERN, and officially announced on 4 July 2012. It has a mass of about 125 GeV. In 2013, P. Higgs and F. Englert received the Nobel prize for their work from 1964 that first described the mechanism we now call the "Higgs mechanism" (the paper of F. Englert was co-authored with R. Brout, who passed away before the particle was discovered; other people who played an important rôle in the theoretical development of the Higgs mechanism are Anderson, Kibble, Guralnik and Hagen.)

Note that the name "Higgs" is overused. We already introduced a Higgs field ϕ, a two component complex scalar. Now we found a real scalar η, which is called the Higgs boson. It is closely related to ϕ, but not the same. In other contexts, we will find other scalar fields that acquire a vacuum expectation value, and which are also called "Higgs fields".

The third and fourth order terms in the kinetic action of the non-abelian gauge bosons give rise to interactions. For example one gets a coupling of the vector fields $W_\mu^\pm$ to the photon, confirming that the charge of these fields is indeed what is suggested by the upper index. There are many other terms giving rise to couplings among the W, Z and η fields which we will not all present here.

4.3 Masses and Mixing angles

4.3.1 Yukawa Couplings

All the Fermi fields are still massless and hence the left and right-handed modes are in principle completely unrelated. Before $SU(2)\times U(1)$ breaking it was impossible to write down a mass term of the form $\bar\psi_L\psi_R$ without violating one of the gauge symmetries. The quarks and leptons can only get their masses after the symmetry is broken, and in order to generate a mass term from the vacuum expectation value of ϕ they must couple to it. Such a coupling can indeed be written down without violating $SU(3)\times SU(2)\times U(1)$, namely (the sign follows the convention for mass terms)

$$\mathcal{L}_Y = -g_{\mathcal{U}}^{\alpha\beta}\bar\psi_L^{\mathcal{Q},\alpha}\ [\mathbf{C}\phi^*]\psi_R^{\mathcal{U},\beta} - g_{\mathcal{D}}^{\alpha\beta}\bar\psi_L^{\mathcal{Q},\alpha}\phi\psi_R^{\mathcal{D},\beta} - g_{\mathcal{E}}^{\alpha\beta}\bar\psi_L^{\mathcal{L},\alpha}\phi\psi_R^{\mathcal{E},\beta} + \text{c.c.}\ , \tag{4.24}$$

where α and β are family labels, and $g_{\mathcal{U}}$, $g_{\mathcal{D}}$ and $g_{\mathcal{E}}$ are complex coupling matrices. Here "c.c" stands for "complex conjugate". If right-handed neutrinos exist, there may be an additional term involving the neutrino fields. It contains the combination of fields $\bar\psi_L^{\mathcal{L},\alpha}\ [\mathbf{C}\phi^*]\psi_R^{\mathcal{N},\beta}$, and puts lepton and quark couplings more or less on equal footing. This is appealing, but it is not clear whether it is also true, and in addition there are other terms one can write down if one introduces right-handed neutrinos. Therefore we postpone the discussion of neutrino masses to the next chapter.

Note that the total charge of each term must be zero. This obliges us to use ϕ^* in the first term and ϕ in the second one. We also have to make sure that all terms are $SU(2)$ singlets. This is easy for terms of the form $\bar\psi_L\phi$, which are singlets automatically if we contract their $SU(2)$ doublet indices in the obvious way: $\bar\psi_L^i\phi^i$. This is because $\bar\psi$ transforms as the complex conjugate of ψ, so it transforms according to the complex conjugate representation, which in addition must be transposed. More precisely, $\bar\psi=\psi^\dagger\gamma_0$. If $\psi\to U\psi$ under some unitary transformation U, then $\psi^*\to U^*\psi^*$ and $\psi^\dagger\to\psi^\dagger U^\dagger$ and hence $\bar\psi\to\bar\psi U^\dagger$. Since $\phi\to U\phi$, the combination $\bar\psi\phi$ is invariant. However, in $\mathcal{L}_Y$ $\bar\psi_L$ couples not only to the field ϕ but also to ϕ^*. By the same logic, the combination $\bar\psi\phi^*$ is *not* invariant under arbitrary unitary transformations, but in the special case of $SU(2)$ (and not for other $SU(N)$ groups) there is a way out. The reason we can do this is that the doublet representation of $SU(2)$ is *pseudo-real*. In general this means that the representation matrices are not real (and cannot be chosen real), but satisfy

$$-(T^a)^* = \mathbf{C}T^a\mathbf{C}^{-1}\ , \tag{4.25}$$

for some orthogonal matrix $\mathbf{C}$. It is easy to check that the $SU(2)$ doublet representation, with representation matrices proportional to the Pauli matrices, satisfy this relation with

$\mathbf{C} = i\sigma^2$, or $\mathbf{C}_{ij} = \epsilon_{ij}$. In other words, the two $SU(N)$ representations (N) and (N^*) are not distinct in the special case $N = 2$; they are equivalent. Writing out the $SU(2)$ doublet indices that were suppressed above, the two couplings read thus $\bar{\psi}_L^i(\phi^j)\delta_{ij}$ and $\bar{\psi}_L^i(\phi^j)^*\epsilon_{ij}$. [The matrix $\mathbf{C}$ should not be confused with the charge conjugation matrix C that acts on spinors.] The vacuum expectation value of $\mathbf{C}\phi^*$ is

$$\langle \mathbf{C}\phi^* \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} v \\ 0 \end{pmatrix} \tag{4.26}$$

which is precisely what is needed to give a mass to the upper component of the doublet.

4.3.2 Mass Matrix Diagonalization

After symmetry breaking we get mass matrices

$$m_{\mathcal{D}} = \frac{g_{\mathcal{D}} v}{\sqrt{2}}, \quad m_{\mathcal{U}} = \frac{g_{\mathcal{U}} v}{\sqrt{2}} \text{ and } m_{\mathcal{E}} = \frac{g_{\mathcal{E}} v}{\sqrt{2}}$$

for the down quarks, up quarks and charged leptons respectively. We can diagonalize these matrices using $U(3)$ rotations in flavor space. Each set of three charge eigenstates and each chirality can be rotated independently to a set of new fields Ψ

$$\Psi_{\text{ch}}^{\alpha,q} = U_{\text{ch},q}^{\alpha\beta} \psi_{\text{ch}}^{\beta,q} , \tag{4.27}$$

where ch $= L, R$ denotes the chirality, and q the particle type or charge, *i.e.* q is either $\mathcal{U}, \mathcal{D}$ or $\mathcal{E}$. If we write the fermion bi-linears in terms of the new fields Ψ the mass matrices transform to

$$\begin{aligned} \hat{m}_{\mathcal{D}} &= \tfrac{1}{\sqrt{2}} U_{L,\mathcal{D}}^\dagger g_{\mathcal{D}} U_{R,\mathcal{D}} v \\ \hat{m}_{\mathcal{U}} &= \tfrac{1}{\sqrt{2}} U_{L,\mathcal{U}}^\dagger g_{\mathcal{U}} U_{R,\mathcal{U}} v \\ \hat{m}_{\mathcal{E}} &= \tfrac{1}{\sqrt{2}} U_{L,\mathcal{E}}^\dagger g_{\mathcal{E}} U_{R,\mathcal{E}} v \end{aligned} \tag{4.28}$$

In general the matrices g are complex and neither symmetric nor hermitean, but since we can rotate both their indices independently, we can make sure that the mass matrices $\hat{m}_{\mathcal{U}}, \hat{m}_{\mathcal{D}}$ and $\hat{m}_{\mathcal{E}}$ are diagonal.

To see that this is possible note that any complex matrix X van always be brought to the form $X = UH$, where U is unitary and H is Hermitean. To bring X to diagonal form we can multiply it from the left and right with *distinct* unitary matrices. Hence we can multiply X from the left with U, so that a Hermitean matrix H remains. Now we can multiply H from the left with a suitable matrix $S^\dagger$ and from the right with S, such that S diagonalizes H in the standard way. The eigenvalues of H are real, but not necessarily positive, but we can multiply from the left (or right) with a diagonal matrix of signs to make all eigenvalues positive.

Note that the matrices U are not completely determined by the requirement that the matrices $\hat{m}$ be diagonal. We may multiply each relation in (4.28) from the right with a diagonal unitary matrix diag $(e^{i\phi_1}, e^{i\phi_2}, e^{i\phi_3})$ and from the left with diag $(e^{-i\phi_1}, e^{-i\phi_2}, e^{-i\phi_3})$ without changing the masses in any way. Thus for each pair U_L, U_R there are three undetermined phases.

4.3.3 The CKM matrix

These rotation matrices do leave a trace in some other terms in the Lagrangian, namely in the couplings of the W bosons. To the unrotated quarks and leptons these bosons couple proportional to

$$\bar{\psi}^{\alpha}_{L,1}\gamma^{\mu}W^{+}_{\mu}\psi^{\alpha}_{L,2} , \tag{4.29}$$

where 1 denotes the upper component of an $SU(2)$ doublet, and 2 the lower. Thus 1 stands for the labels $\mathcal{U}$ or $\mathcal{N}$, and 2 for $\mathcal{D}$ or $\mathcal{E}$. Expressing this in terms of the mass eigenstate Ψ one gets

$$\bar{\Psi}^{\alpha}_{L,1}U^{\alpha\beta}_{\rm CKM}\gamma^{\mu}W^{+}_{\mu}\Psi^{\beta}_{L,2} , \tag{4.30}$$

where $U_{\rm CKM} = (U_{L,\mathcal{U}})^{\dagger}U_{L,\mathcal{D}}$ for quarks, and $U_{\rm CKM} = (U_{L,\mathcal{N}})^{\dagger}U_{L,\mathcal{E}}$ for leptons. Henceforth we reserve the notation $U_{\rm CKM}$ for the CKM (Cabibbo-Kobayashi-Maskawa) matrix of the quarks. We return to the leptonic equivalent in the next chapter.

Note that the coupling matrices g contain a large amount of redundant information. Only the mass eigenstates and the relative rotations of the left-handed quarks are observable. The rotations of the right-handed particles are not. Lepton rotations require more discussion, since we first have to decide the origin of neutrino masses.

4.3.4 Counting Free Parameters in the CKM Matrix

To write down Lagrangians in quantum field theory one always starts with the most general expressions that respect all desired symmetries. These symmetries can be local continuous symmetries (gauge symmetries), global symmetries or discrete symmetries. In the case of the Standard Model the only symmetries are the gauge symmetries $SU(3) \times SU(2) \times U(1)$. In general, it is the model builder who decides which symmetries a model should have.

The most general Lagrangian respecting all symmetries may have a large number of parameters, but often only a subset of those parameters can be measured. The remaining parameters can be absorbed by redefining fields. The set of fields in a QFT can be redefined by taking arbitrary (non-degenerate) linear combinations. To remove the arbitrariness as much as possible, one first brings the kinetic terms to their canonical form; then one does the same with the mass terms (which can always be diagonalized) and if any field redundancy is left one can use it to bring some interaction terms to a standard form.

A general gauge invariant expression for the kinetic terms of a set of fermions χ^{α}_{L} is

$$\mathcal{L} = i\sum_{\alpha\beta} H_{\alpha\beta}\bar{\chi}^{\alpha}_{L}\gamma^{\mu}D_{\mu}\chi^{\beta}_{L} \tag{4.31}$$

and similarly for the right-handed fields. The sum is over the family labels. The gauge symmetry allows an arbitrary matrix H here (which can be different for each species $\mathcal{U}, \mathcal{D}, \ldots$). Hermiticity of the Hamiltonian (reality of the energy) requires H to be Hermitean; positivity of energy requires it to be positive definite. Then H can be written as

$H = A^\dagger A$, for some complex matrix A. If we now define new fields $\psi = A\chi$ using matrix multiplication in family space we get the kinetic terms in their canonical form

$$\mathcal{L} = i \sum_\alpha \bar{\psi}_L^\alpha \gamma^\mu D_\mu \psi_L^\alpha \tag{4.32}$$

This is usually the starting point, but note that in principle there is a large set of free parameters H, which can be transformed away. Of course we replace $A\chi$ by ψ in all other terms in the Lagrangian, but in the Yukawa couplings this just redefines the matrices g_α. Since the covariant derivatives behave as ordinary derivatives, after bringing the kinetic terms in canonical form the gauge interactions are still diagonal in family space, as in Eqn. (4.29).

After writing the kinetic terms in canonical form, there is still some redundancy left in field space: we can apply to all fermions ψ_L a unitary transformation in family space. This can be done for all species and chiralities separately. We limit ourselves now to the quark sector, since we are considering the CKM matrix. Before weak symmetry breaking, we have at our disposal the unitary 3×3 matrices $U_{L,\mathcal{Q}}$, $U_{R,\mathcal{U}}$ and $U_{R,\mathcal{D}}$. After weak symmetry breaking we can transform the two components of the $SU(2)$ multiplet $\mathcal{Q}$ separately, since we do not have to respect the broken $SU(2)$ anymore. So then we have four matrices $U_{L,\mathcal{U}}$, $U_{L,\mathcal{D}}$, $U_{R,\mathcal{U}}$ and $U_{R,\mathcal{D}}$.

We use this freedom first of all to bring the mass matrices produced by the Higgs v.e.v. in diagonal form. Since mass matrices have the structure $\bar{\psi}_L \psi_R$ they are sensitive to unitary rotations of the left-handed components with respect to the right-handed ones. Note that we have enough freedom to make all masses positive. Since bringing the mass matrices to diagonal form usually requires non-trivial transformations $U_{L,\mathcal{U}}$ and $U_{L,\mathcal{D}}$, this implies that the W vertex becomes off-diagonal in family space: we get a non-trivial CKM matrix.

Once the masses are in their canonical form, we can determine how much freedom we have still left. We observed above that $U_{L,\mathcal{U}}$ and $U_{L,\mathcal{D}}$ are not completely determined by the mass diagonalization. We can multiply them each with a diagonal unitary matrix, provided that we compensate for this in the corresponding U_R. But these are unobservable, so that we are allowed to change U_{CKM} by multiplying it from the left as well as the right by two independent diagonal unitary matrices. In other words, the mass terms as well as the kinetic terms are invariant if we change ψ_L and ψ_R by the same diagonal phase matrix, and we can do that for the species $\mathcal{U}$ and $\mathcal{D}$ separately. If the mass eigenvalues are all different, these are the only transformations that leave the mass matrices invariant.

Let us now count the number of parameters in U_{CKM} for N families. A unitary matrix can be written as e^{iT}, where T is a Hermitean matrix. There are N^2 independent Hermitean $N \times N$ matrices, and hence such matrices are described by N^2 real parameters. The $2N$ undetermined phases can be used to fix $2N$ of these parameters to any desired value, so that we are left with $N^2 - 2N$ parameters. However, not all $2N$ phases can be used. If we multiply U_{CKM} from the left with a diagonal matrix $e^{i\phi}\mathbf{1}$ and from the left with the inverse of that matrix, they cancel each other. Hence we had only $2N-1$, and not $2N$ independent phases at our disposal, and the number of parameters is thus

$N^2 - 2N + 1 = (N-1)^2$. This gives $0, 1$ and 4 for 1,2 and 3 families respectively.

As an example, consider the CKM matrix for two families. The most general CKM matrix, a 2×2 unitary matrix, can be parametrized as

$$\begin{pmatrix} e^{i(\alpha-\gamma)} \cos\theta_c & -e^{i(\alpha-\delta)} \sin\theta_c \\ e^{i(\beta-\gamma)} \sin\theta_c & e^{i(\beta-\delta)} \cos\theta_c \end{pmatrix} , \tag{4.33}$$

This depends on four and not five parameters, since it only depends on the differences of α, β, γ and δ. By using the freedom to multiply this on the left and the right by diagonal phase matrices, we can bring it to the form

$$\begin{pmatrix} \cos\theta_c & -\sin\theta_c \\ \sin\theta_c & \cos\theta_c \end{pmatrix} , \tag{4.34}$$

and hence there is only one physical parameter. The fact that α, β, γ and δ can be transformed away by field redefinitions means that they can never be determined in any physical process. Hence there is just one real parameter, not four. The parameter θ_c is called the Cabbibo angle. More precisely, it was called that when only two families were known to exist; in fact even the c quark had not been discovered yet when θ_c was introduced. For three families there are more complicated expressions for the CKM matrix, usually involving three angles and a phase. The standard parametrization is

$$U_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}s_{23}s_{13}e^{i\delta} & -c_{12}c_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \tag{4.35}$$

where $s_{12} = \sin\theta_{12}$, $c_{12} = \cos\theta_{12}$, etc.

Experimentally the matrix U_{CKM} is nearly equal to 1, but there are small off-diagonal matrix elements; the largest of these is the old Cabbibo-angle, $\theta_c \approx 13^{\text{o}}$. It determines how strongly an up-quark couples to an s-quark (compared to its coupling to its own family member, the down quark). This coupling is small: $U^{us}_{\text{CKM}} \approx \sin(\theta_c)$. Nowadays one defines $\theta_c \equiv \theta_{12}$, and introduces three similar angles for mixing between the second and the third family ($\theta_{23} \approx 2.38^{\text{o}}$) and the first and the third ($\theta_{13} \approx 0.2^{\text{o}}$). The phase is: $\delta = 1.2 \pm .08$. The fact that the matrix is so close to 1 is not understood, although there are models that produce this, together with the mass hierarchies, starting from some assumptions about the Yukawa couplings. The structure of the matrix may contain important hints regarding physics beyond the Standard Model.

4.3.5 Flavor Changing Neutral Currents and the GIM Mechanism

The rotation matrices disappear completely in the couplings of the Z boson and the photon. These particles couple upper components of $SU(2)$ doublets to upper components, and lower ones to lower ones. Hence after rotation to mass eigenstates, the coupling matrices in flavor space are all of the form $U^{-1}_{L,\mathcal{U}} U_{L,\mathcal{U}} = \mathbf{1}$, and similarly for $\mathcal{D}, \mathcal{E}, \mathcal{N}$. This is known as the GIM-mechanism. One may wonder why a trivial looking identity like

$U^{-1}U = \mathbf{1}$ has the name of three famous physicists (Glashow, Iliopoulos and Maiani) attached to it. The reason is that at the time the Standard Model was constructed only three quarks, u, d and s were known, and the latter therefore was not considered part of an $SU(2)$ doublet. In that case one finds that the Z-boson can couple a d quark to a d quark or an s quark, just like the W-boson can couple the u-quark to either a d, s (or a b) quark. Such couplings are called "Flavor changing neutral currents" (FCNC), and they are not observed experimentally. The GIM-mechanism explained that and predicted the existence of the c-quark. The absence of FCNC's is nowadays one of the most important constraints in attempts to go beyond the Standard Model.

5 A First Look Beyond

In this chapter we discuss a variety of issues that might be relevant for attempts to understand the Standard Model.

5.1 The Left-handed Representation

5.1.1 Replacing Particles by Anti-Particles

The notation employed thus far suggest that there is some sort of distinction between left- and right-handed fields. Actually all fields are on the same footing. This has to do with the existence of anti-particles. The anti-particle of a right-handed fermion is a left-handed anti-fermion. The fermion action, $i\bar{\psi}_R\gamma^\mu D_\mu\psi_R$ does not treat particles and anti-particles symmetrically. This is most obvious if one considers the covariant derivative: for a $U(1)$ gauge field this looks like $\partial_\mu - ieqA_\mu$, where q is the charge of ψ (for non-abelian symmetries our conventions are such that $D_\mu = \partial_\mu - igT^a A^a_\mu$).

Here a definite choice is made among the particle and anti-particle charge. The $SU(3) \times U(1)$ Lagrangian is always written in such a way that the charge corresponds to what we call "particles", as opposed to anti-particles. Note that there is no such asymmetry in the action of a complex scalar.

What we call "particles" is simply the species that we see most abundantly in our own environment. We see protons and electrons, and very few anti-protons and positrons. It is still an unsolved mystery how this asymmetry has arisen (the "baryogenesis problem"), but there is no fundamental reason why we should prefer particles over anti-particles.

Of course we know that the same fermion action also describes the anti-particle. Hence it should be no surprise that it can be rewritten in such a way that the rôle of particle and anti-particle are interchanged. To do so we introduce new variables

$$\begin{aligned}
\psi &= C^{-1}(\gamma^0)^T(\psi^c)^* = C^\dagger(\bar{\psi}^c)^T \\
\bar{\psi} &= -(\psi^c)^T C \ ,
\end{aligned} \tag{5.1}$$

where C is the charge conjugation matrix introduced in appendix C, which is a unitary matrix satisfying

$$C\gamma_\mu C^{-1} = -(\gamma_\mu)^T \ . \tag{5.2}$$

The action of C on γ_5 is: $C\gamma_5 C^{-1} = (\gamma_5)^T$. The precise form of C depends on the explicit representation of the Dirac γ-matrices, but the only thing that matters is that such a matrix C exists in any representation. The relation for $\bar{\psi}$ is not independent, but follows from the one for ψ. Note that this changes right-handed fields to left-handed ones:

$$\begin{aligned}\psi_R &= P_R\psi = P_R C^{-1}(\gamma^0)^T(\psi^c)^* = C^{-1}(P_R)^T(\gamma^0)^T(\psi^c)^* \\ &= C^{-1}(\gamma^0)^T(P_L)^T(\psi^c)^* = C^{-1}(\gamma^0)^T(P_L)^*(\psi^c)^* = C^{-1}(\gamma^0)^T(\psi^c_L)^* \ .\end{aligned}$$

Substituting this into the action $i\bar{\psi}_R\gamma^\mu D_\mu\psi_R$ yields a new action

$$\begin{aligned}i\bar{\psi}_R\gamma^\mu D_\mu\psi_R &= -i(\psi^c_L)^T C\gamma^\mu D_\mu C^\dagger(\bar{\psi}^c_L)^T \\ &= -i(\psi^c_L)^T C\gamma^\mu D_\mu C^{-1}(\bar{\psi}^c_L)^T \\ &= i(\psi^c_L)^T(\gamma^\mu)^T D_\mu(\bar{\psi}^c_L)^T \end{aligned} \tag{5.3}$$

This expression is a number, *i.e.* a 1×1 matrix. So it is equal to its own transpose, and we may replace it by its transpose. Since the fermions anti-commute this requires some care. The identity we are using is

$$\chi^T M\eta = \sum_{i,j}\chi^i M_{ij}\eta^j = -\sum_{i,j}\eta^j M_{ij}\chi^i = -\sum_{i,j}\eta^j M^T_{ji}\chi^i = -\eta^T M^T\chi \ , \tag{5.4}$$

where χ and η are mutually anti-commuting spinors. In our case they correspond to ψ^c and $\bar{\psi}^c$, and the indices i,j represent the complete set of indices ψ has, *e.g.* spin (Dirac indices), gauge and flavor indices. For the ordinary derivative term in the covariant derivative this yields

$$i(\psi^c_L)^T(\gamma^\mu)^T\partial_\mu(\bar{\psi}^c_L)^T = -i\partial_\mu\bar{\psi}^c_L\gamma^\mu\psi^c_L \tag{5.5}$$

We now move ∂_μ to ψ^c_L by "partial integration", *i.e.* we pretend that the Lagrangian density is integrated over space-time. This gives a final minus sign. For the gauge boson coupling part of the covariant derivative we get

$$\begin{aligned}&i(\psi^c_L)^T(\gamma^\mu)^T[-igA^a_\mu T^a](\bar{\psi}^c_L)^T \\ &\quad= -i\bar{\psi}^c_L[-igA^a_\mu(T^a)^T]\gamma^\mu\psi^c_L \\ &\quad= i\bar{\psi}^c_L[+igA^a_\mu(T^a)^*]\gamma^\mu\psi^c_L \ ,\end{aligned} \tag{5.6}$$

where we made use of the fact that T^a is hermitean. The final result is thus

$$i(\bar{\psi}^c)_L\gamma^\mu(\partial_\mu + ig(T^a)^* A^a_\mu)\psi^c_L \tag{5.7}$$

This is the desired result since $-(T^a)^*$ is the generator of the complex conjugate representation.

5.1.2 The Standard Model in Left-handed Representation

Having done this we can now describe all physics in terms of ψ^c_L instead of ψ_R. This removes an arbitrary distinction between left- and right-handed fields. This distinction made sense below the scale of weak symmetry breaking, since the left and right-handed components are paired by the mass-terms, but not in the unbroken theory. Furthermore we can now consider transformations that take any field to any other fermion field. This would be quite hard to describe if part of the fields had opposite handedness. A Standard Model family now looks like this

$$
\begin{array}{rc}
(3,2,\frac{1}{6}) & \begin{pmatrix} u_L \\ d_L \end{pmatrix} \\
(3^*,1,-\frac{2}{3}) & u^c_L \\
(3^*,1,\frac{1}{3}) & d^c_L \\
 & \\
(1,2,-\frac{1}{2}) & \begin{pmatrix} \nu_L \\ e^-_L \end{pmatrix} \\
(1,1,1) & e^+_L \\
(1,1,0) & \nu^c_L
\end{array}
\tag{5.8}
$$

5.1.3 Fermion Masses in the Left-handed Representation

The fermion mass terms now look somewhat different. Consider a Dirac mass term of the form $\bar{\psi}m\psi$. If we use left- and right-handed components this looks like $\bar{\psi}_L m\psi_R + \bar{\psi}_R m\psi_L$. In general, if ψ is a field with several components, m is replaced by a matrix M. Since the Hamiltonian must be Hermitean, one should in general expect to find an expression of the form $\bar{\psi}_L M\psi_R + \bar{\psi}_R M^\dagger\psi_L$. In the Standard Model this structure comes out because of the "+ c.c" in formula (4.24). In general this means adding the complex conjugate, or the Hermitean conjugate for operators. We have already seen that by a unitary rotation of ψ_L one may always rotate M to a Hermitean matrix, but before doing that it will come out as a general complex matrix.

In this mass term, ψ_L and ψ_R are really distinct fields. To emphasize that, we can give ψ_L a different name, χ_L. Then a typical (off-diagonal) mass term looks like this

$$-\bar{\psi}_R M\chi_L - \bar{\chi}_L M^\dagger\psi_R \tag{5.9}$$

and after replacing ψ_R by its left-handed anti-particle this takes the form

$$(\psi^c_L)^T MC\chi_L - \bar{\chi}_L M^\dagger C^\dagger(\bar{\psi}^c_L)^T\ . \tag{5.10}$$

Note that the second term is the Hermitean conjugate of the first. All indices have been suppressed here, but note that M and C are respectively matrices in family and in spinor space. Mass terms clearly looked nicer in L-R notation, but we will see that the left-handed representation has other advantages that make it worthwhile paying this price.

In the rest of these notes we will use both representations, depending on what is most convenient. We will refer to the representation used in the previous chapter as the *particle representation*, since all fermi fields are particles, and their conjugates anti-particles. This is the most useful basis when masses are present, for example for physics at low energies. The other one, the *left-handed representation* is more useful above the weak scale, where the fermions are massless.

5.1.4 Yukawa Couplings in the Left-handed Representation

Yukawa couplings are very similar to mass terms. They look like Eq. (5.10) but with the matrix M replaced by a coupling matrix times a Higgs field. All contractions of family and gauge indices work as before. The diagonalization of the mass matrices goes exactly as before, but our previous notation is now a bit unattractive. Therefore we define

$$\begin{array}{rcl} U_x & \equiv & U_{L,x} \\ V_x & \equiv & U^*_{R,x} \ , \end{array} \tag{5.11}$$

where x denotes $\mathcal{U}, \mathcal{D}, \mathcal{E}$ or $\mathcal{N}$. In this notation the matrices U act on particles and V on anti-particles, and the fermions ψ transform to $U\psi$ and ψ^c to $V\psi^c$. The diagonalization is now $M_{\text{diag}} = V^T M U$.

5.1.5 Real Representations

Equation (5.10) shows how to write down mass terms directly in terms of left-handed fields. An expression like $m\psi_L^T C\chi_L$, where ψ and χ are two, in principle distinct, spinors is obviously Lorentz invariant, and can be added to the Lagrangian if it is also a gauge singlet. Previously our mass terms looked like $m\bar{\psi}_L\chi_R$, and because of the complex conjugation the requirement was that ψ_L and χ_R belong to the *same* representation. Now the requirement becomes that ψ_L and χ_L must belong to *mutually complex conjugate* representations. Note that two mutually complex conjugate representations together form a real representation. Concretely, if ψ_L has N components and transforms in the following way under the action of some symmetry

$$\psi_L \to U\psi_L \tag{5.12}$$

then the fact that ψ_L and χ_L belong to mutually complex conjugate representations means that χ_L transforms as

$$\chi_L \to U^*\chi_L \ , \tag{5.13}$$

where U is a $N \times N$ unitary matrix. In the $2N$-dimensional space spanned by ψ_L and χ_L this transformation takes the form

$$\begin{pmatrix} \psi_L \\ \chi_L \end{pmatrix} \to \begin{pmatrix} U & 0 \\ 0 & U^* \end{pmatrix} \begin{pmatrix} \psi_L \\ \chi_L \end{pmatrix} \tag{5.14}$$

This matrix can be made real by means of a unitary transformation (proof: first diagonalize U. Then, in each 2×2 block of conjugate eigenvalues $\text{diag}(e^{i\phi}, e^{-i\phi})$ one can transform this matrix to a two-dimensional rotation.). After this basis transformation (ψ_L, χ_L) becomes a spinor Ψ_L which transforms according to a $2N$-dimensional real representation

$$\Psi_L \to O\Psi_L; \quad O^T O = \mathbb{1} \tag{5.15}$$

If a field Ψ belongs to a real representation one can write down a mass term of the form

$$-m\Psi_L^T C\Psi_L + \text{c.c} \tag{5.16}$$

This is called a Majorana mass term. It is obviously invariant under (5.15).

Note that the representations of the Standard Model *after* symmetry breaking can also be written entirely in terms of left-handed fields.* For one family one gets then the $SU(3)\times U(1)$ representations $(3,\frac{2}{3}),(3^*,-\frac{2}{3}),(3,-\frac{1}{3}),(3^*,\frac{1}{3})$ for the u and d quark, $(1,-1),(1,1)$ for the electron, and $(1,0)$ for the neutrino. One can determine all possible mass terms by finding all real subspaces. These are $(3,\frac{2}{3})+(3^*,-\frac{2}{3})$, $(3,-\frac{1}{3})+(3^*,\frac{1}{3})$, $(1,-1)+(1,1)$ and $(1,0)$ by itself. The latter one, the Majorana mass term for the neutrino does not appear in the *standard* model. This term can be present without any enlargement of the field content. However, it would break a global symmetry, namely lepton number, and hence there are important constraints on such a term.

In general if one has a gauge theory with fermions written in left-handed representation, one can write down a mass term for any subset of the fields that form a real representation. The distinction between Majorana masses and Dirac masses is not very big in this language. One can speak of Majorana masses if a field is in a real representation that is irreducible, whereas one speaks of Dirac masses when a field is in a representation that is irreducible as a real representation, but that consists of two mutually complex conjugate components. An example of such a representation is that of a u quark, $(3,\frac{2}{3})+(3^*,-\frac{2}{3})$, which is a real representation because one can find a basis so that the $SU(3)\times U(1)$ representation matrices are real, but which is reducible as a complex representation.

5.1.6 Mirror Fermions

Before symmetry breaking the Standard Model fermi fields (except the right-handed neutrino, which many people do not regard a Standard Model particle anyway) are in a fully complex representation, so that no mass terms can be written down. This is presumably no coincidence. It is quite possible that the fermions we see are only the low energy remnants of a larger fermion representation, which contains some real parts. Since mass terms for the real parts are not forbidden, they might indeed be generated, and then the complex part is all that survives as low energies. Without further information it is of course not possible to say anything about the masses of such particles. The most common occurrence of this sort of real $SU(3)\times SU(2)\times U(1)$ matter in models is in the form of *mirror families*. These are families of fermions whose representation is the complex conjugate of those of the families we observe. Instead of living in a world with 3 families, we might live in a world with $N+3$ families and N mirror families, where N families have paired with N mirrors to form massive particles.

5.2 Neutrino Masses

In the original formulation of the Standard Model the neutrinos are massless. There are three neutrinos, one per family, and they are part of a doublet $(1,2,-\frac{1}{2})_L$, together with the left-handed electron, muon and τ. Since, by definition, the Standard Model does not

* Often we will call this the "broken Standard Model". It has a gauge group $SU(3)_{\rm QCD}\times U(1)_{\rm QED}$.

contain any right-handed singlet $(1,1,0)_R$, one cannot write down Yukawa coupling to the Higgs boson, as one can do for the quarks. Hence the neutrinos do not acquire a mass from the Higgs mechanism.

Meanwhile it has become clear in several experiments that neutrinos *oscillate*, which means that they may be created as one of the three species (ν_e, ν_μ or ν_τ), but may be observed as another species (see section 5.2.6 below for details). The notation of the three neutrinos indicates how they are produced. In the perturbative Standard Model individual lepton numbers are conserved. Hence in any Feynman diagram one can follow the flow of quantum numbers. In a coupling to the W bosons, an electron goes into a neutrino with electron number 1 (denoted ν_e), and analogously a muon goes into ν_μ and a τ into ν_τ. These are called the "interaction eigenstates". However, if neutrino's get a mass, the mass eigenstates are in general a different linear combination, and hence a ν_e is a linear combination of the three mass eigenstates. If these masses are different, the three components have a different time evolution, and this results in oscillating probabilities for finding the three interaction states as a function of the distance of propagation. The possibility that neutrinos may oscillate was anticipated decades earlier by Pontecorvo [26]. The neutrino oscillation probabilities are sensitive to differences of (mass)^2. Since so far oscillations are the only information we have about neutrino masses, it follows that only mass *differences* have been seen. The differences are compatible with three distinct neutrino masses. One of these could in principle be zero, but that would seem a bit strange. This differences are extremely small, of order $\Delta m^2 \approx 10^{-3}$ to 10^{-5} eV^2.

Since we only know mass differences, one could still allow for neutrino masses as large as charged lepton masses, in the MeV or even GeV range. But this is not possible. Obviously we know already since Pauli postulated it in 1930 that the neutrino emitted in β-decay is extremely light (that is why it was called "neutrino"). Meanwhile we know this much more accurately from precise measurements of Tritium decays. This imposes limits of about 2 eV on the particular neutrino combination that is emitted in β-decay. Clearly, given the very small mass differences, this essentially rules out large masses for all three mass eigenstates, in any simple extension of the Standard Model with three neutrinos. For more about mass limits see section 5.2.7.

Furthermore, there are limits from cosmology. If the sum of the three neutrino masses exceeds 40 eV, neutrino matter would "over-close" the universe, which means that they contribute too much to Ω as defined in eqn. (1.7). Note that neutrino mass limits based on this argument must necessarily depend on estimates of neutrino abundances, assumptions about neutrino stability and basic assumptions about cosmology, in contrast to direct observations. There are other cosmological estimates based on the properties of the Cosmic Microwave Background (CMB) and other astrophysical features. These too depend on some additional assumptions, and give an upper limit for the sum of the masses of less than an eV. The latest limit from the Planck satellite's observation of the CMB is $\sum m_\nu < .23$ eV.

From all this information we know that neutrinos have masses, and that these masses are smaller than those of the charged leptons by a factor of a million or more.

5.2.1 Modifications of the Standard Model

Only minor modifications of the Standard Model are required to get massive neutrinos. In the Standard Model, one has one left-handed neutrino (or, equivalently, a right-handed anti-neutrino) for each family. There are two basic changes that can be made.

- Add a right-handed neutrino to allow a Dirac mass term.
- Add a Majorana mass term for the Standard Model neutrino.

The first possibility puts the neutrinos on the same footing as the charged leptons and the quarks: *all* Standard Model particles would have left- and right-handed components, and the only thing strange is the smallness of the neutrino masses. Indeed, although it is often said that neutrino mass is the first example of Beyond the Standard Model physics, this is a matter of definition. Based on what we knew before the discovery of neutrino oscillations, two versions of the Standard Model could have been chosen. The first is to omit right-handed neutrinos and the Dirac mass term, so that neutrinos are massless. This is how most people define the Standard Model. Neutrino masses are then "Beyond the Standard Model" by definition. But an equally reasonable definition would have been to allow right-handed components and Dirac masses, just as for all other particles, and assume that the masses were too small to observe. With this definition, the observation of neutrino masses through oscillations would just be the first observation of a finite difference of Standard Model parameters, which were too small to be be observed until a few years ago. The only "BSM" aspect of this scenario is the existence of additional degrees of freedom, the right-handed neutrinos.

However, the second possibility definitely deserves the label "Beyond the Standard Model", for several reasons: the Majorana mass term breaks lepton number, and adds an additional mass parameter, which *a priori* is not related to the Higgs field.

Note that one can in principle choose between these two modifications for each family separately; the first family neutrino may be given a pure Dirac mass and the second a pure Majorana mass, etc. But we will soon see that most likely *both* options are realized simultaneously: there would then be Dirac as well as Majorana mass terms in the Lagrangian. Furthermore, in a scenario where families behave distinctly, it become difficult, if not impossible, to obtain the observed neutrino mixing. Therefore we will from now on assume that the neutrino mass generation mechanism is the same for all three families.

5.2.2 Adding a Dimension 5 Operator

The second possibility listed in the previous section can only be realized in the broken Standard Model ($SU(3) \times U(1)$). In the unbroken Standard Model the left-handed neutrino is part of a doublet $\psi_L^{\mathcal{L}}$ with non-vanishing Y-charge (in the following we omit the superscript $\mathcal{L}$ since the lepton doublet is the only one that appears). The Majorana combination* $(\psi_L)^T \mathbf{C} C \psi_L$ is not Y invariant, and cannot be added to the Lagrangian without

* Note that there is one matrix C to make a Lorentz invariant Majorana mass term in spinor space, and a matrix $\mathbf{C}$ to couple the $SU(2)_{\rm weak}$ indices to a singlet.

breaking Y-charge. Since Y is a gauge symmetry, this is unacceptable.

We have just seen that we cannot generate a Majorana mass for the neutrino using the standard dimension-4 Yukawa coupling to the Higgs. But there are other ways. One method is to complicate the Higgs sector. The simplest possibility is to add a new Higgs boson in the triplet representation of $SU(2)$. This leads to several problems. Why would both Higgses get a vev that breaks $SU(3) \times SU(2) \times U(1)$to $SU(3) \times U(1)$? Why would the vevs line up so that they choose the same unbroken direction in $SU(2) \times U(1)$: electromagnetism?

There is a more appealing possibility. A Majorana mass term for the Standard Model neutrino may also be generated by a coupling of two neutrino fields to *two* Standard Model Higgs fields rather than just one. The required invariant combination is

$$[(\psi_L)\mathbf{C}\phi]^T C(\psi_L)\mathbf{C}\phi \tag{5.17}$$

This is called a "Weinberg operator" [33]. Note that $\mathbf{C}$, the charge conjugation matrix in $SU(2)_{\mathrm{W}}$ space, is used here to couple ψ_L and ϕ to an $SU(2)$ singlet. This combination has vanishing Y-charge, and it is a fermion. The spinor space matrix C is used to couple the two fermionic combinations to a Lorentz singlet. This combination has dimension 5, and therefore there will always be a coefficient $\frac{g}{M}$ multiplying this operator, where g is a dimensionless coupling constant and M a mass scale. A theory containing such an operator is not renormalizable, which means concretely that it does not make sense at scales larger than M. However, it is perfectly acceptable as an effective theory below M. This means that we can use it as long as the typical energies in a process are smaller than M. Indeed, it would be a good idea to make M very large, in order to obtain naturally small neutrino masses: $m_\nu \approx gv^2/M$. This idea is realized more naturally in the see-saw mechanism discussed below.

Note that lepton number* is necessarily broken, because Eq. (5.17) contains two fields ψ_L, and not a field and its conjugate. One could try to avoid that by assigning lepton number to the Higgs field ϕ, but then lepton number is broken as soon as the Higgs gets a v.e.v; lepton number disappears into the vacuum.

5.2.3 Neutrino-less Double-beta Decay

The violation of lepton number in the foregoing two mechanisms is a general consequence of neutrino Majorana masses. An observable manifestation of this breaking is neutrino-less double β-decay. In normal β-decay, a neutron in a nucleus decay to a proton, and the nucleus emits an electron and an anti-neutrino. This can also happen twice in a single process, a phenomenon observed in some nuclear decays where the single β decay is not energetically allowed because the final state nucleus is too heavy. These two processes are illustrated in fig. 1. The third process makes use of the Majorana mass term, which can be

* In fact, as we will see later, lepton number is not an exact symmetry of the Standard Model anyway. However, the combination $B - L$ (baryon number minus lepton number) might be an exact global symmetry. The Majorana mass term also violates $B - L$. The absence of such a Majorana mass term would be therefore be natural if we assumed that $B - L$ is an exact symmetry of nature.

represented as a two-point vertex with two outgoing (or ingoing) neutrino external lines. Note that this vertex violates the rule that fermion lines can be followed throughout any Feynman diagram in the direction of the arrows. This is precisely the effect of a Majorana mass term.

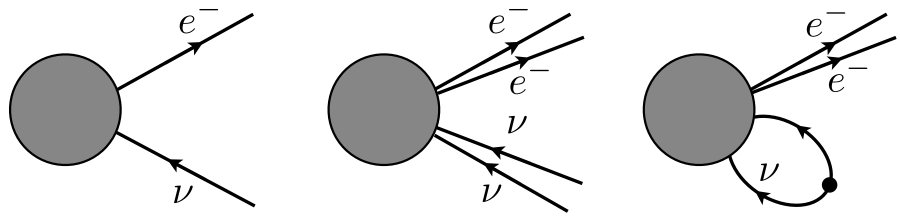


Figure 1: Ordinary β-decay, double β-decay and neutrino-less double β-decay

In the third case one sees two electrons but no neutrinos coming out, and hence a violation of lepton number by two units. Such decays have been looked for, but not found so far.

5.2.4 Adding Right-handed Neutrinos

The first possibility listed in section 5.2.1 is less exotic. One just adds a right-handed neutrino field ($i.e$ in the left-handed representation a left-handed anti-neutrino) and a Dirac mass-term. The extra field belongs to the $SU(3) \times SU(2) \times U(1)$ representation $(1, 1, 0)$, and the Dirac mass term can be generated by the Standard Model Higgs boson, in exactly the same way as it generates the up and down quark masses. The lepton sector looks then rather similar to the quark sector, and in particular it has its own CKM matrix. This is not very natural, however, since one would expect the neutrino masses to be roughly of the same order of magnitude as the other lepton masses and the quark masses. Although the hierarchies among the quark and lepton masses are large and not understood, a non-zero but small (< 2 eV for ν_e) electron-neutrino mass makes this hierarchy problem substantially worse.

5.2.5 The See-Saw Mechanism

A pure Dirac mass is also unnatural since as soon as we add the representation $\nu_L^c = (1, 1, 0)$ nothing forbids us to write down a Majorana mass term. This term has the following form

$$\tfrac{1}{2}(\nu_L^c)^T C M_m \nu_L^c + \text{h.c.} \tag{5.18}$$

When there are several flavors of fermions M_m is a matrix. Note that this matrix M_m must be symmetric, or, more precisely, only the symmetric part of M_m contributes. To see this more clearly, write the expression with explicit spinor indices α, β and flavor indices (i, j) as

$$\tfrac{1}{2}\sum_{\alpha,\beta}\sum_{i,j}(\nu_L^c)_{\alpha,i} C_{\alpha\beta} (M_m)_{i,j} (\nu_L^c)_{\beta,j} + \text{h.c.} \tag{5.19}$$

Because of fermi statistics, and because C is anti-symmetric, it follows then that M_m must be symmetric. There is no obvious reason why it should be real, however, so we cannot use standard results for diagonalizing real symmetric matrices using orthogonal transformations. However, a complex symmetric matrix M can be "diagonalized" as follows (this is known as Autonne-Takagi factorization)

$$M_{\text{diag}} = U^T M U \tag{5.20}$$

where U is unitary, and M_{diag} is diagonal and real. The reason for writing "diagonalized" with quotes is that this not the standard diagonalization of complex matrices. The standard way is to use $U^\dagger$ instead of U^T, because this is what correspond to a true basis transformation in a complex vector space. However, the proper procedure in QFT is to bring first the kinetic term in canonical form, and then use any remaining freedom to bring the mass terms in diagonal form. We are treating the field with representation $(1, 1, 0)$ here a a left-handed Weyl fermion, just as any other Standard Model fermions. Its kinetic term is

$$i\bar{\nu}_L^c \gamma^\mu \partial_\mu \nu_L^c \tag{5.21}$$

This is invariant under unitary transformations of the field ν_L^c. Applying this transformation to the Majorana mass term gives us precisely the correct transformation (5.20) to bring the mass matrix to real diagonal form.

Unlike all other direct quark and lepton mass terms, the mass term (5.18) is allowed by $SU(3) \times SU(2) \times U(1)$, and its mass scale M_m is not set by the Standard Model Higgs mechanism. The parameter M_m is unrelated to the Higgs mass parameter μ^2, and may *a priori* have any value. Note that the Majorana mass term $(\nu_L^c)^T C M_m \nu_L^c$ violates lepton number, just like the Weinberg operator (5.17). On the other hand, it is not unreasonable to assume that any term that is not explicitly forbidden by a gauge symmetry will indeed appear, even if such a term violates a discrete symmetry. The discrete symmetries of the Standard Model are merely a consequence of the fact that the Lagrangian terms of dimension four and less just happen to respect B and L. There is no profound reason why these symmetries should be sacred, unlike gauge symmetries, whose breaking renders the theory inconsistent. Furthermore, gravity has little respect for discrete symmetries: baryon and lepton number can disappear into a black hole, without leaving a trace. According to this philosophy, a term $(\nu_L^c)^T C M_m \nu_L^c$ should exist with M_m determined by some higher scale.

In addition to this Majorana mass term, we also have a Dirac mass term, which can be written in terms of left-handed fields as indicated in eqn. (5.10)

$$(\nu_L^c)^T M C \nu_L - \bar{\nu}_L M^\dagger C^\dagger (\bar{\nu}_L^c)^T \ . \tag{5.22}$$

Combining the Majorana mass term and the Dirac mass term, we get a mass matrix of the following form

$$\tfrac{1}{2}(\nu^T, (\nu^c)^T)_L C \begin{pmatrix} 0 & M_d \\ M_d^T & M_m \end{pmatrix} \begin{pmatrix} \nu \\ \nu^c \end{pmatrix}_L + \text{h.c.} \ , \tag{5.23}$$

It is assumed here that there are no direct Majorana contributions to the mass of ν_L, such as for example a Weinberg operator. That is why there is a 0 in the first row. But adding a non-zero entry here would not change anything quantitatively. The off-diagonal terms are simply the first term of (5.22), distributed symmetrically, with a factor $\frac{1}{2}$ to get the correct normalization. Note that the complete mass matrix for three families is a 6×6, symmetric, and complex matrix, which can be diagonalized by the method explained above. This diagonalization will mix ν and ν_c.

Consider first the simplest case, one family. We may assume that M_m and M_d are real, because if they are not we may multiply the fields ν and ν^c with appropriate phase factors to make them real. Then the matrix can be diagonalized by means of an orthogonal matrix and it leads to mass eigenvalues $\frac{1}{2}(M_m \pm \sqrt{M_m^2 + 4M_d^2})$. If we make the approximation $M_d \ll M_m$, which is reasonable according to the arguments given above, the eigenvalues are approximately M_m and $-\frac{M_d^2}{M_m}$ (the sign is irrelevant). Then we end up with one very massive neutrino (ν_L^c with a very small admixture of ν) and one very light one (essentially ν). If we take $M_d \approx 1$ GeV, the value $M_m = 10^{11}$ GeV leads to an naturally small neutrino mass of about 10^{-2} eV. This is called the "see-saw mechanism". In the limit $M_m \to \infty$, ν_L^c decouples from all interactions except gravity, and one recovers the Standard Model.

In the three family case one can solve the eigenvalue problem approximately in the limit where the determinant of M_m is much larger than that of M_d. One can then use the following *ansatz* for the light eigenvectors

$$\begin{pmatrix} \vec{v} \\ -M_m^{-1} M_d \vec{v} \end{pmatrix} , \tag{5.24}$$

Acting on this with the matrix (5.23) we get

$$\begin{pmatrix} -M_d M_m^{-1} (M_d)^T \vec{v} \\ 0 \end{pmatrix} , \tag{5.25}$$

Hence the vector (5.24) is transformed into $-M_d M_m^{-1}(M_d)^T$ times itself, up to corrections of order (M_d/M_m). Then the three light neutrino mass eigenvalues are approximately the eigenvalues of the 3×3 matrix $-M_d M_m^{-1}(M_d)^T$. In addition there are three heavy neutrino mass eigenvalues which are obtained by diagonalizing M_m.

5.2.6 Neutrino Oscillations

A very interesting consequence of neutrino masses and mixings is neutrino oscillations, observed fairly recently, but foreseen already in 1957 by Pontecorvo. This is a textbook application of quantum mechanics, which occurs whenever a state is created in a linear combination of two non-degenerate mass eigenstates. Each time we perform a measurement, we will find with a predictable probability one of the particles involved in the mixing. These probabilities evolve with time in a non-trivial way if the masses are different. The neutrino produced in, say, muon decay will in general be a linear combination of ν_e, ν_μ and ν_τ. If these three particles have the same masses, any linear combination is a mass eigenstate, and we might as well call the linear combination to which the muon decays ν_μ. This linear combination is often called the "interaction eigenstate". If all masses are equal, interaction eigenstates are mass eigenstates. When the muon neutrino subsequently scatters off a proton via W exchange it can only produce a muon, since the linear combination does not change, and hence it is still an interaction eigenstate. However, if the masses are distinct muon decay produces a linear combination of different mass eigenstates. The three components evolve differently with time, and hence after some time the state evolves to a linear combination of the three interaction eigenstates. This implies that in a scattering process the neutrino produced in muon decay can yield either an electron, a muon, or a tau, with probabilities that depend on the mixing angles, the neutrino masses and the length of the path traveled by the neutrino.

The PMNS matrix. Let us make this more precise. The coupling of the W boson is given by an expression completely analogous to the one for quarks, Eq. (4.30):

$$\bar{\Psi}_L^{\mathcal{L},i} U_{\text{PMNS}}^{i\alpha} \gamma^\mu W_\mu^- \Psi_L^{\nu,\alpha} \;, \tag{5.26}$$

where "P" refers to Pontecorvo, who first pointed out the possibility of neutrino oscillations in 1957 [26] and MNS stands for Maki, Nakagawa and Sakata, who proposed this matrix in 1962 [22] for two lepton flavors.* In (5.26) $\mathcal{N}$ and $\mathcal{L}$ denote, as before, the set of neutrinos resp. charged leptons. The labels $\alpha = e, \mu, \tau$ denote the charged lepton mass eigenstates, and $i = 1, 2, 3$ denote the neutrino mass eigenstates, in no particular order. One also writes (omitting the label "PMNS" for convenience)

$$\begin{aligned} |\nu_\alpha\rangle &= U_{\alpha i}\,|\nu_i\rangle \\ |\nu_i\rangle &= U^*_{i\alpha}\,|\nu_\alpha\rangle \end{aligned}$$

If the neutrino masses were purely of Dirac type, this matrix would have the same number of parameters as the CKM matrix, and can be parametrized in exactly the same way, although with very different values for the parameters $\theta_{12}, \theta_{23}, \theta_{13}$ and δ. If there are also Majorana components in the neutrino masses, there are two additional parameters (which can be transformed away if $M_m = 0$). They can be chosen as follows

$$U_{\text{PMNS}} = U(\theta_{12}, \theta_{23}, \theta_{13}, \delta) \times \text{diag}\,(e^{i\gamma_1}, e^{i\gamma_2}, 1) \tag{5.27}$$

where $U(\theta_{12}, \theta_{23}, \theta_{13}, \delta)$ is a standard form as used for the CKM matrix. The phases γ_1, γ_2 are CP violating (just as δ), but they do not contribute to neutrino oscillations. The standard parametrization of $U(\theta_{12}, \theta_{23}, \theta_{13}, \delta)$ is

$$\begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}s_{23}s_{13}e^{i\delta} & -c_{12}c_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \tag{5.28}$$

* Interestingly, the CKM matrix for quarks appeared later: Cabibbo introduced his angle for the two–family case in 1963, whereas Kobayashi and Maskawa published their paper in 1973.

where $s_{12} = \sin\theta_{12}$, $c_{12} = \cos\theta_{12}$, etc. Unlike the CKM matrix elements, some of the PMNS matrix elements are large: $\theta_{12} \approx 33^o$, $\theta_{23} \approx 45^o$, $\theta_{13} \approx 9^o$. The phase δ is essentially unknown.

The reason that the PMNS matrix has two extra phases is a direct consequence of the fact that the W-boson coupling is between a standard Dirac fermion on the one hand (the charged lepton) and a Majorana particle on the other hand. Let us compare this to the counting for the CKM matrix. The CKM matrix couples quarks to quarks. For N families, it is an $N \times N$ unitary matrix, which can be multiplied from the left and the right with diagonal phase matrices. These phase matrices are precisely the unitary transformations that leave respectively the up and down quark masses invariant. Since a Dirac mass term has the form $M\bar{\psi}_L\psi_R + \text{h.c.}$, one can multiply left-handed and right-handed Dirac fermions with compensating phases, without affecting M. Only the phase of the left-handed particle contributes to the CKM matrix, and hence this matrix can be changed. Fixing that phase is like a gauge choice: we have to agree on it in order to compare our results. There are N phases from the up-quark sector and another N from the down-quark sector. Since the overall up-quark and down-quark phase commute with the CKM matrix and can cancel each other, the net parameter reduction is by $2N - 1$, so that we get $N^2 - 2N + 1 = (N - 1)^2$ parameters. But if the W boson couples a Dirac fermion to a Majorana fermion, we do not get a diagonal phase factor from the Majorana side because a Majorana mass term contains the same fermionic field twice, and if we phase rotate this fermion this affects the mass (note that one could change the fermionic field with a sign, but not with a phase). Hence the number of parameters is $N^2 - N$, which for $N = 3$ gives 6. Note that if both Majorana and Dirac fermions are contributing, as in the seesaw mechanism, the neutrino masses are always of Majorana type. However, the two extra phases γ_1 and γ_2 cannot be observed unless one considers processes sensitive to the difference between Majorana and Dirac masses. In particular, they cannot be observed in neutrino oscillations [3, 7]. Indeed, even though we have observed oscillations, we still do not know if there exists a Majorana mass term (and hence a violation of lepton number).

Oscillations for two neutrino species. Although the subsequent discussion is easily generalized to three neutrinos, for simplicity we consider only two, namely the one that couples via a W boson to the electron and the one that couples to the muon. These are what one usually calls the electron neutrino and the muon neutrino. In collisions with other particles, a pure electron neutrino can only produce an electron, through the interaction $\nu_e \to e^- + W^+$ occurring as part of a more complicated process. Hence if we observe the electron in a detector, the neutrino is thereby identified as an electron neutrino. Similarly, if a muon scatters with matter and is converted into a neutrino by W exchange, this neutrino is by definition a muon neutrino. These are the interaction eigenstates. However, for generic mass matrices we cannot expect these to coincide with the mass eigenstates, and indeed it turns out that they do not. In fact we have

$$|\nu_e\rangle \;\; = \;\; \cos\theta\,|\nu_1\rangle + \sin\theta\,|\nu_2\rangle$$

$$|\nu_\mu\rangle \;=\; -\sin\theta\,|\nu_1\rangle + \cos\theta\,|\nu_2\rangle \ , \tag{5.29}$$

where ν_1 and ν_2 are the mass eigenstates. These mass eigenstates have the usual quantum mechanical time evolution:

$$|\nu_i, t\rangle = e^{iHt}\,|\nu_i, 0\rangle = e^{it\sqrt{(p_i)^2+(m_i)^2}}\,|\nu_i, 0\rangle \tag{5.30}$$

Since this time evolution is different for the two components, a pure electron neutrino will not stay a pure electron neutrino as it evolves in time. If it is detected, one may find that with some probability it has changed into a muon neutrino.

After some time interval T the interaction eigenstates has evolved to a state

$$|\nu_e, T\rangle = e^{iT\sqrt{p^2+m_1^2}}\cos\theta\,|\nu_1\rangle + e^{iT\sqrt{p^2+m_2^2}}\sin\theta\,|\nu_2\rangle \tag{5.31}$$

We can now compute the overlap of this state with an interaction eigenstate. The square of the amplitude is the probability for finding an electron neutrino in the final state

$$\begin{aligned} P(\nu_e \to \nu_e) &= \mid e^{iT\sqrt{p^2+m_1^2}}(\cos\theta)^2 + e^{iT\sqrt{p^2+m_2^2}}(\sin\theta)^2 \mid^2 \\ &= 1 - \left[\sin\left(\frac{T}{2}\left(\sqrt{p^2+m_1^2} - \sqrt{p^2+m_1^2}\right)\right)\right]^2 [\sin(2\theta)]^2 \end{aligned} \tag{5.32}$$

If we make the approximation that the neutrino mass is much smaller than its energy (or momentum), we get $\sqrt{p^2+m_2^2} - \sqrt{p^2+m_1^2} \approx (m_1^2 - m_2^2)/2E$ (with $E = p$, up to corrections of order m^2/E^2). Finally we express the result not in terms of the time of flight T of the neutrinos, but the distance L they travel. Since they are very relativistic we get $L = T$ (because $c = 1$). The final result is

$$P(\nu_e \to \nu_e) = 1 - \left[\sin(\Delta m^2 L/4E)\right]^2 [\sin(2\theta)]^2 \tag{5.33}$$

Since probability is conserved we must also have

$$P(\nu_e \to \nu_\mu) = \left[\sin(\Delta m^2 L/4E)\right]^2 [\sin(2\theta)]^2 \tag{5.34}$$

Note that the effect disappears if the neutrinos are degenerate in mass, or if the mixing angle θ vanishes.

Oscillations for three neutrino species. The three-family formula can be worked out along the same lines, and after a bit of work one obtains

$$\begin{aligned} P(\nu_\alpha \to \nu_\beta) = \delta_{\alpha\beta} \quad &- \quad 4\sum_{i>j}\mathrm{Re}(U^*_{\alpha i}U_{\beta i}U_{\alpha j}U^*_{\beta j})\sin^2\left(\frac{\Delta m^2_{ij}L}{4E}\right) \\ &\mp \quad 2\sum_{i>j}\mathrm{Im}(U^*_{\alpha i}U_{\beta i}U_{\alpha j}U^*_{\beta j})\sin\left(\frac{\Delta m^2_{ij}L}{2E}\right) \end{aligned}$$

where the upper sign is for neutrinos and the lower one for anti-neutrinos. One may verify that for two species we re-obtain Eqns. (5.33) and (5.34). In that case U is just

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

which is real, and hence the last term vanishes. In fact, in the two-family case there is an extra parameter, the Majorana phase. In the three family case there are two such phases, γ_1 and γ_2 in Eqn. (5.27). But such phases cancel out in the neutrino oscillation formula.

Note that because of the last term we can in principle measure the sign of Δm_{ij}^2, unless the prefactor vanishes.

5.2.7 Neutrino Experiments

There are three kinds of experiments that are sensitive to neutrino masses.

- Direct measurements
- Neutrino-less double-beta decay
- Neutrino oscillation experiments

In addition to this we have information from various astrophysical and cosmological sources (already briefly mentioned in section 5.2), such as the mass density of the universe, the effect of neutrino masses on Big Bang nucleosynthesis and on the cosmic microwave background and the travel time of neutrinos produced in supernova explosions. This is an exciting field with many opportunities for new results, but we will focus here on the three classes listed above, that are not affected by cosmological and astrophysical assumptions.

The first class of measurements amount to checking energy and momentum conservation in interactions where a neutrino has been produced. For the electron neutrino the standard experiment is tritium β-decay. The tritium nucleus decays to helium-3, and electron and an anti-neutrino:

$${}^3_1\mathrm{H} \rightarrow {}^3_2\mathrm{He} + e^- + \bar{\nu}_e + 18.6 \text{ KeV} .$$

If the latter has a mass, less energy is available for the electron. Hence the experiments try to determine the maximum energy of the decay electrons. So far no indication for non-vanishing mass was found, which implies an upper limit of about 2 eV for the mass of the electron neutrino. The masses of the other two neutrino combinations can be determined from energy-momentum conservation in accelerator experiments. The current limits are 190 KeV and 18 MeV for ν_μ and ν_τ respectively [8]. These are maxima on the missing mass in experiments, just as the β-decay limits. These experiments precede the observation of neutrino oscillations in 1998 by many years. Based on what we have learned meanwhile about neutrino oscillations it seems clear that the actual neutrino masses are much, much smaller than any of these limits.

Neutrino-less double beta decay is sensitive to the Majorana mass. This may either be a pure Majorana mass, or the Majorana component in the more complicated situation where also Dirac masses are present. If lepton number is violated (which is always the case if one introduces a Majorana mass, for the left- or for the right-handed neutrino), neutrino-less double beta decay is allowed, and may be observable. Several experiments are looking for it, but so far without undisputed results.

Neutrino oscillation experiments are sensitive to differences of mass-squares. A positive result proves that at least one neutrino must have non-zero mass, but unfortunately this does not tell us anything about the masses themselves. The oscillation experiments fall into several classes. First of all one can distinguish appearance and disappearance experiments. The first checks oscillation from species a to a different species b, and the second checks whether the total flux of species a is preserved. The experiments can also be subdivided according to the origin of the neutrinos: solar, reactor, accelerator, atmospheric or cosmic sources (*e.g.* supernovae).

Solar neutrino experiments. Solar neutrino experiments measure the number of electron neutrinos observed on earth that are produced in nuclear reactions in the sun. Initially these experiments were merely intended for finding solar neutrinos. They did indeed find them, but already since the 60's these experiments reported a shortage, finding only about one third of what was expected. The expectations depend on solar models, which were during many years seen as the main culprit of the shortage, but over the years the solar models became so robust that this became unlikely. Most of these experiments look at so-called "charged-current" interactions involving a W boson. The reaction is $\nu_e + n \rightarrow e^- + p$, where a neutron in a nucleus is converted into a proton. The difficulty is finding a few of these converted nuclei (*e.g.* Germanium) within a huge quantity of detector material (*e.g.* Gallium). This reaction is only sensitive to electron neutrinos since there is not enough energy available to produce muons or taus. The Sudbury Neutrino Observatory (SNO), was able to look in addition to neutral current interactions (involving the Z boson). In these interactions the final state lepton is also a neutrino, and interactions of all three neutrino species are observable. This experiment found in addition to the factor three deficit in charged current interactions, precisely the expected solar neutrino flux in neutral current interactions. This is very strong evidence that on their way from the sun to earth a substantial fraction of neutrinos have oscillated to other species. This was announced in 2001, and produced the decisive clue in the decade-old solar neutrino puzzle. In 2002 Raymond Davis received the Nobel prize for his pioneering work on detecting solar neutrinos. His detector used about 600 tonnes of Chlorine, which in a rare neutrino interaction gets converted to Argon. They captured about 2000 neutrinos over a period of thirty years!

It turns out that solar neutrino oscillations are only partly due to oscillations in vacuum. For high energy neutrinos (energies of about 5-20 MeV) there is a second oscillation effect due to oscillations in matter, *i.e.* the sun. This is called the Mikheyev-Smirnov-Wolfenstein (MSW) effect. The formulas we gave above are for oscillations in vacuum, and are not valid for oscillations in matter. Taking this effect into account one gets the survival probability of about 1/3 observed in the early solar neutrino experiments. It turns out that the MSW effect is sensitive to the sign of the mass difference. The data are consistent with a mass difference $\Delta m^2 \approx 7.5 \times 10^{-5}$ eV^2 and an angle θ_{12} of about 33^o. The MSW effect is less important for low energy (a few MeV) neutrinos, and in this case we can directly compute the neutrino oscillations using the $P(\nu_e \to \nu_e)$ formula given above. Plugging in the typical energy of the neutrinos and the distance to the sun one finds that we are in a region where the factor $\sin^2(\Delta m^2_{12} L/4E)$ is fluctuating rapidly. Hence this factor averages out to $\frac{1}{2}$. Now we use the value for θ_{12} and we get a survival probability of about 60%, which is indeed what is observed for low energy neutrinos. Of course historically θ_{12} and Δm^2_{12} were output, not input.

Atmospheric neutrino oscillations. Atmospheric neutrinos result from the decay of pions and Kaons, which in their turn are produced if cosmic ray particles (mainly protons) collide with the earth's atmosphere. The decays yield mainly muon neutrinos. One hopes to find signs of neutrino oscillations by comparing the neutrino flux from above and below (in which case the neutrinos have the pass through the earth, and hence travel different distances). These experiments (the best-known one is "(Super)Kamiokande", 1998) indicate that indeed such oscillations occur. In 2015 T. Kajita from Kamiokande and A. McDonald from SNO received the Nobel prize for their work on discovering neutrino oscillations.

Reactor experiments. Reactor experiments look for neutrinos from nuclear reactors. In 2005 the first experiment of this kind (KamLAND) reported evidence for oscillations. Previous experiments sensitive to smaller values of L showed no effect. The Daya Bay reactor experiment in China was the first to determine that $\theta_{13} \neq 0$ in a significant way.

Accelerator experiments. Accelerator experiments use neutrinos produced by colliding particle beams with targets. More than ten such experiments have been done or are being planned.

The present values for Δm^2 are about 7.5×10^{-5} eV^2 (for solar oscillations) and 2.5×10^{-3} eV^2 (for atmospheric oscillations). Since this concerns different processes, these results are not inconsistent with each other, nor with the existence of three neutrino species.

Hierarchy ambiguities. There is in the current data still an ambiguity in the ordering of the three mass eigenstates. As we have seen above, using the MSW effect one could determine the sign of the mass difference of the two mass eigenstates involved in solar oscillations. We do not have matter oscillations at our disposal for atmospheric neutrinos to determine the sign of the other mass difference. Hence we are left with an ambiguity. In the future, one may be able to use the fact that the full three-family oscillation formula has a term that is sensitive to the sign, but this is not yet possible. Hence we have two possible mass hierarchies. Either the two mass states whose mass difference agrees with solar oscillations are the lightest ("normal hierarchy"), or they are the heaviest ("inverted hierarchy"). The labelling convention is to number them 1,2 and 3 in increasing order of mass in the normal hierarchy, and 3,1, 2 in the inverted hierarchy. These are the labels used in the PMNS matrix above for the columns; the rows are labelled e, μ, τ. With this convention, the mixing angles are the same for both hierarchies. In particular, ν_e, ν_μ and ν_τ have the same decomposition in terms of mass eigenstates for the normal and the inverted hierarchy.

5.3 C,P and CP

It is shown in Appendix C, the kinetic terms in the fermion action $-\bar{\psi}_L\gamma^\mu D_\mu\psi_L$ transform into themselves under P and C up to a chirality flip. If we work in a left-handed representation, we must transform back to right-handed fields to see if parity is a symmetry. But this conjugates the representation matrices. Hence parity reversal and charge conjugation are symmetries of the kinetic terms, including minimal coupling to the gauge fields, if and only if the representation is self-conjugate. If it is not, then we see that C and P transform these terms in exactly the same way, so that in any case the combination CP is a good symmetry. Note that it is quite tricky to check C and P in a mixed left and right representation, and that statements like "P is broken because the left- and right-handed fields are in different representations of the gauge group" are simply not correct.

To discuss the symmetries of the Yukawa couplings there is no advantage in using the left-handed representation, so we use L-R notation instead. One should consider each term in combination with its Hermitean conjugate. A typical pair of such terms will have the form

$$\mathcal{L}_Y = g\bar{\psi}_L\chi_R\phi + g^*\bar{\chi}_R\psi_L\phi^* \tag{5.35}$$

Under parity the first term transforms to

$$g\bar{\psi}_R\chi_L\phi \ , \tag{5.36}$$

and under charge conjugation to

$$g\bar{\chi}_L\psi_R\phi^* \ . \tag{5.37}$$

If C and P are already broken by the kinetic terms, there is not much reason to expect them to be a symmetry now. Indeed, in general the fields χ_L and ψ_R (the parity conjugates of χ_R and ψ_L) need not even exist. If they do exist, but P or C are not a symmetry, the coupling constants in Eqs. (5.36) and (5.37) are not related to those of Eq. (5.35). However, since P and C separately are already broken by the minimal couplings to the gauge fields, our real interest is in their product, CP. It is should be clear that CP transforms the first term in Eq. (5.35) into a term that has the same structure as the second term in Eq. (5.35), but with a coupling constant g instead of g^*, since coupling constants do not transform under P or C. Hence we conclude that CP is broken if g is complex.

This argument is too simplistic, however. What really matters is whether CP is violated in a physical process. If there were just one set of fermions, this is certainly not true, since one can make g real by a phase rotation of, for example χ_R. In the Standard

Model the only place where an observable CP violation can occur is in the CKM matrix. It is easy to show that for one and two families this matrix can always be made real by suitable fermion rotations, but that this cannot be done for three families or more, provide all quarks are massive. Hence the fact that there are three families in nature gives us a natural mechanism for CP-violation.

The origin of CP-violation is however not well-established in the standard model, and is one of the areas where one should be prepared for deviations. It is quite easy to make slight changes to the model which provide additional sources of CP-violation. The best-known one is to extend the Higgs sector. An example will be discussed later.

5.4 Continuous Global Symmetries

For every set of N degenerate real fields in the same representation of the gauge group, the kinetic terms in the action have a global $O(N)$ symmetry. Similarly for complex fields there is a $U(N)$ symmetry. These symmetries might be broken explicitly by interaction terms. If they are not broken explicitly, but are broken spontaneously, the spectrum will contain massless Goldstone bosons corresponding to these symmetries. This is an important constraint on Beyond the Standard Model phenomenology.

In the Standard Model the gauge bosons are all in different multiplets, so there is at most a $O(1)$ symmetry for each, a sign change. But the minimal couplings forbid this also (for the $U(1)_Y$ gauge boson this symmetry symmetry becomes part of the charge conjugation symmetry of the complete Lagrangian).

The matter consists of three copies each of the five $SU(3) \times SU(2) \times U(1)$ fermion representations, plus a complex Higgs. The kinetic terms have thus a $U(3)^5 \times U(1)$ global symmetry. The Yukawa couplings break this symmetry. Any off-diagonal $U(3)$ transformations are destroyed because the eigenvalues of the three Yukawa coupling matrices are all different.

To see if any $U(1)$ transformations are preserved one can try diagonalize these matrices. As we already know, this cannot be done in the quark sector: one may diagonalize $g_{\mathcal{U}}$ using unitary matrices $U_{\mathcal{U}}$ and $V_{\mathcal{U}}$, but to diagonalize $g_{\mathcal{D}}$ we would need matrices $U_{\mathcal{D}}$ and $V_{\mathcal{D}}$, with $U_{\mathcal{U}} \neq U_{\mathcal{D}}$. Note that both both $U_{\mathcal{U}}$ and $U_{\mathcal{D}}$ both act on components of the left-handed quark doublet, and since they are different one cannot simultaneously diagonalize all Yukawa couplings. The matrices $U_{\mathcal{U}}$, $V_{\mathcal{U}}$, $U_{\mathcal{D}}$ and $V_{\mathcal{U}}$ are usable after Higgs symmetry breaking, because then it is meaningful to act on the separate components of the weak doublet. The fact that $U_{\mathcal{U}} \neq U_{\mathcal{D}}$ leads to a non-trivial CKM matrix, so we know experimentally that these matrices are indeed different. Therefore if we transform any quark field by a phase, and we want the quark Yukawa couplings to be invariant, we must transform all quarks by the same phase, and anti-quarks by the opposite phase. This surviving $U(1)$ symmetry is Baryon number (B), and it is normalized in such a way that all quarks have $B = \frac{1}{3}$.

We have previously identified four mechanisms for breaking the $U(6) \times U(6)$ chiral symmetries. They are broken to $U(1)^6$ (the separate flavor numbers) by QCD and QED. The weak interactions, and in particular the fact the the CKM matrix is non-trivial, break this global symmetry to just a single $U(1)$, thus adding a fifth origin of $U(6)\times U(6)$ breaking.

In the lepton sector the situation is more or less the same as in the quark sector. If we start with left-handed lepton doublets, plus right-handed charged leptons and neutrinos, then the quark sector and the lepton sector both have an $U(6)\times U(6)$ in the limit of zero fermion masses and if electroweak interactions are switched off. QCD chiral symmetry breaking only affects the quarks, but if we treat the quark and lepton sector otherwise equally, we also end up with just a single conserved quantity, lepton number. Since neutrinos oscillate into each other, we know that separate electron, muon and tau lepton numbers are not conserved. Apart from the absence of chiral symmetry breaking, the other novelty in the lepton sector is the possibility of introducing Majorana masses. This would break lepton number completely.

Finally we may transform the Higgs field by a phase. This is automatically a symmetry of the Higgs potential, but it is a symmetry of the Yukawa couplings only if the quarks and leptons transform with compensating phases. If one solves the conditions for invariance of the Yukawa couplings, one finds only one solution, namely the gauged $U(1)_Y$ symmetry of the Standard Model gauge group. So the single Higgs field of the Standard model does not introduce new global symmetries.

5.5 Anomalies

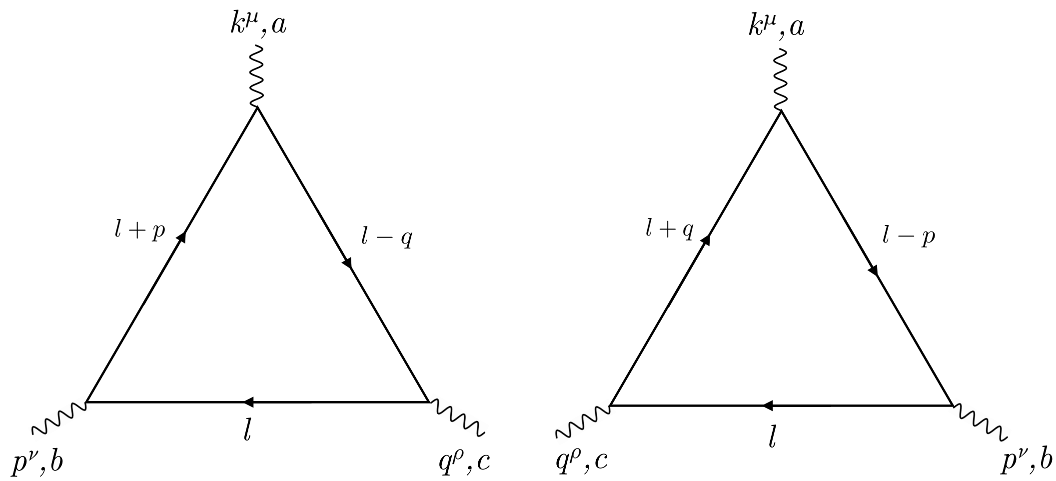
All symmetries we discussed so far were good symmetries classically, but quantum corrections break some of them. The Feynman diagrams responsible for this breaking are fermion triangles with external (axial) vector currents (in D space-time dimension anomalies originate from fermion polygons with $\frac{1}{2}D+1$ sides; chiral anomalies exist only if D is even). The problem occurs only if the fermion trace contains the matrix γ_5. In purely vector-like theories, where all couplings to the vector bosons are only via the Dirac matrix γ^μ, the problem does not occur. But as soon as there is a coupling via $\gamma^\mu\gamma_5$ some classical symmetries must be broken in the quantum theory. Such couplings typically arise if vector bosons only couple to left- or right-handed fermions, as in the weak interactions.

The triangle diagrams contribute to the amplitude $\epsilon^a_\mu(k)\epsilon^b_\nu(p)\epsilon^c_\rho(q)V^{\mu\nu\rho}(k,p,q)$, where $\epsilon^a_\mu(k)$ is a polarization vector of a vector boson. Here all three polarization tensors could be different, *i.e* they may belong to different vector bosons. If the classical symmetry is respected, then the amplitude must vanish if we replace any of the polarization tensors by the momentum of the vector bosons. This follows from the momentum space version of current conservation, $\partial_\mu J^\mu=0$. Hence the Green's function $V^{\mu\nu\rho}(k,p,q)$ should satisfy

$$k_\mu V^{\mu\nu\rho}(k,p,q)=0\ , \tag{5.38}$$

where k is one of the external momenta. An analogous relation should hold for the other two external momenta if the symmetry is to hold quantum mechanically.

The relevant triangle diagrams are:



This turns into an integral over a trace of the fermions, with three propagators and three vertices coupling to vector bosons, which can be either γ^μ or $\gamma^\mu\gamma_5$ (and similarly for ν, ρ). It turns out that if the fermionic trace contains a γ_5, no regularization of the diagram preserves the classical symmetry. One can impose conservation of two of the three currents, for example the ones coupling to the vertices labeled ν and ρ, but then one gets for the third current

$$ik_\mu V^{\mu\nu\rho}(k,p,q) = \frac{1}{2\pi^2}\epsilon^{\nu\rho\alpha\beta}p_\alpha q_\beta \ . \tag{5.39}$$

This result holds for a single fermion with axial vector couplings $i\bar{\psi}\gamma^\mu\gamma_5\psi$ to the external currents. The anomaly can be shifted to any of the three vertices, but cannot be removed.

If none of the currents in the diagram is gauged there is no problem, since then the diagram will never contribute to any Green's function. The same is true if only one vertex is a gauge current. If two or three external lines are gauge bosons there are important consequences, however.

5.5.1 Feynman Diagram Computation

The computation of these diagrams goes as follows. We take all momenta in the graphs as incoming, and we will choose the (k^μ, a) vertex (the top one) to be the one we act on with k^μ. We choose a loop momentum l, and the fermion propagators are assigned as indicated in the figure. The Feynman rule for a fermion propagator for a fermion of mass m and momentum k^μ is

$$\frac{i(\not{k}+m)}{k^2-m^2+i\epsilon} \tag{5.40}$$

The $i\epsilon$ is needed for making a correct Wick rotation later on, but for the moment we will just drop it to keep the notation simple. We use vertices of the form introduced in section 2.6, and in particular we allow for a non-abelian generator T^a at every vertex. We will set $m = 0$, but later we will need the Dirac propagator with non-vanishing mass.

Then the expression to be computed is

$$iV^{\mu\nu\rho}_{abc}(p,q) = \\ -\int \frac{d^4l}{(2\pi)^4}\,\mathrm{Tr}\left[(i\gamma^\mu\gamma_5 T_a)\left(\frac{i(\not{l}+\not{p})}{(l+p)^2}\right)(i\gamma^\nu T_b)\left(\frac{i\not{l}}{l^2}\right)(i\gamma^\rho T_c)\left(\frac{i(\not{l}-\not{q})}{(l-q)^2}\right)\right]$$

To this we have to add the same expression with (p, b, ν) simultaneously interchanged with (q, c, ρ). Note the overall minus sign due to the fact that we have a fermion loop. The trace is over the gamma matrices as well as the gauge generators. Collecting all factors and separating the traces we get

$$iV^{\mu\nu\rho}_{abc}(p,q) = \\ -\int \frac{d^4l}{(2\pi)^4}\,\mathrm{Tr}\left[\gamma_5\gamma^\mu\left(\frac{\not{l}+\not{p}}{(l+p)^2}\right)\gamma^\nu\left(\frac{\not{l}}{l^2}\right)\gamma^\rho\left(\frac{\not{l}-\not{q}}{(l-q)^2}\right)\right]\ \mathrm{Tr}\left[T_aT_bT_c\right]$$

The integral is linearly divergent: in Euclidean space we observe that the leading terms behave as $\int d^4l(l^3/l^6) \approx \int dl$. Finding divergent integrals is quite customary in quantum field theory, and how to deal with these divergences correctly is a long story. But in any case the first step in that process is to *regularize* the integral. This means that we write it as the limit of a convergent expression. A rather brutal way of doing that is to simply introduce a momentum cutoff. But this is not even well-defined, because it depends on how we define the loop momentum in the first place; note that we can shift l by some fixed amount. Generally one prefers regularization methods that can be applied directly to the Lagrangian, rather than manipulating individual diagrams. With such a prescription at least there is a relation between the ways different diagrams are regularized. A popular method in gauge theories is *dimensional regularization*. One simply treats the number of space-time dimensions as a variable, and sets it equal to 4 in the end. With proper care, this can be done in a continuous way. But the presence of a γ_5 in the trace makes proper care very tricky. This matrix is proportional to the product of $\gamma^0, \gamma^1, \gamma^2$ and γ^3, and this is a definition that does not extend smoothly to other dimensions.

For this reason another method is often used, called *Pauli-Villars regularization*. One introduces a new particle with the same spin as the fermion going around in the loop, but with opposite statistics. This particle is given a mass M, and in the end of the calculation M is taken to infinity. This means that we go back to the original Lagrangian in that limit, because particles with infinite mass can be ignored (they "decouple"). The idea is that by having opposite statistics the auxiliary particle makes exactly the same contribution as the fermion loop, but with opposite sign. Hence for $M = 0$ is cancels the entire diagram, and for nonzero M at least it cancels the divergence. Of course the auxilliary particle violates the spin-statistics theorem, but in the infinite mass limit it is not really there, so this should not matter. If we include the auxiliary particle loops the result for the two diagrams now becomes ("Reg." stands for "Regularized")

$$iV^{\mu\nu\rho}_{abc,\mathrm{Reg.}}(p,q) = \quad -\int \frac{d^4l}{(2\pi)^4}\left[I_0^{\mu\nu\rho}(l,p,q) - I_M^{\mu\nu\rho}(l,p,q)\right]\mathrm{Tr}\left[T_aT_bT_c\right]$$

$$-\int \frac{d^4l}{(2\pi)^4}\left[I_0^{\mu\rho\nu}(l,q,p) - I_M^{\mu\rho\nu}(l,q,p)\right] \mathrm{Tr}\,[T_a T_c T_b]$$

where I_0 is the integrand shown above, and I_M is the same one with a mass M in all fermion propagators.

The auxiliary particle mass M has potentially implications for the problem we are considering. This is because conservation laws for currents are as follows

$$\begin{aligned}\partial_\mu(\bar\psi\gamma^\mu\psi) &= 0\\ \partial_\mu(\bar\psi\gamma^\mu\gamma_5\psi) &= 2iM(\bar\psi\gamma_5\psi)\end{aligned}$$

Therefore the current of the auxiliary field is not conserved. This will be the reason we find an anomaly. At this point one may raise the question if perhaps another regularization method can be found that preserves the symmetry explicitly. But this is not possible. The most convincing way of seeing that is by analyzing the problem in terms of path integral quantization, but that is beyond the scope of these lectures. The answer is that no matter how one approaches the problem, one always ends up with the same anomaly.

It turns out that after regularization the linear divergence of the integral cancels, but in the triangle diagram without γ_5 there is still a logarithmic divergence that contributes to the renormalization of the three-point coupling. But this is not what we are interested in. We are interested in the contraction of the vertex $V^{\mu\nu\rho}$ with $k_\mu = -p_\mu - q_\mu$. Note that because of current conservation for $M = 0$ the contraction of the terms with I_0 with k_μ yields exactly zero, so the entire contribution will come from the I_M terms. However, without these terms the integral is not defined, so one cannot prove anything by sending M to infinity prematurely.

To work out the contraction with $(p+q)^\mu$ we use the manifest identity

$$\gamma_5(\not p + \not q) = \gamma_5(\not l + \not p - M) + (\not l - \not q - M)\gamma_5 + 2M\gamma_5 \tag{5.41}$$

The factors $(\not l + \not p - M)$ and $(\not l - \not p - M)$ combine nicely with the propagators, *e.g.*

$$(\not l + \not p - M)\frac{\not l + \not p + M}{(l+p)^2 - M^2} = 1 \tag{5.42}$$

We use this both in the terms with $M \neq 0$ as in the ones with $M = 0$. Let us first deal with the first two terms in Eqn. (5.41). The discussion for these two terms is identical. With one propagator cancelled, we are left with a trace over two propagators and two γ matrices from the vertices. We use the identity

$$\mathrm{Tr}\ \gamma_5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma = 4i\epsilon^{\mu\nu\rho\sigma} \tag{5.43}$$

Furthermore, the trace of γ_5 with fewer than four γ matrices vanishes. Then what is left is only

$$-\int \frac{d^4l}{(2\pi)^4}\left[\frac{-4i\epsilon^{\nu\sigma\rho\alpha}l_\sigma q_\alpha}{l^2(l-q)^2} - \frac{-4i\epsilon^{\nu\sigma\rho\alpha}l_\sigma q_\alpha}{[l^2-M^2][(l-q)^2-M^2]}\right] \tag{5.44}$$

This is a convergent integral. It must yield something of the form $\epsilon^{\nu\sigma\rho\alpha}q_\alpha X_\sigma$, where X_σ is a four-vector that results from the integral. But such a four-vector must point in some direction in four-space, that must be some linear combination of vectors appearing in the integrand. The only such vector is q, and hence X must be proportional to q, and then the whole expression vanishes. Note that for this conclusion it is important that the second term makes the integral finite. Without that second term, one might also think that the first term must necessarily be proportional to q_σ. But this conclusion would be wrong. Note that we could shift the integration variable from l to $l+t$, where t is an arbitrary vector. But if we do that, the integral will be proportional to a linear combination of q_σ and t_σ. The term proportional to t_σ does not vanish and is in fact logarithmically divergent, so clearly the conclusion that the integral can only be proportional to q_σ makes no sense. By contrast, if we make a shift of integration variable in the full expression Eqn (5.44) it has no effect, because it is merely a change of variables in a finite integral.

Having eliminated the contributions from the first two terms in Eqn (5.41) we now have only the last one to deal with. This yields the following γ-matrix trace

$$2M\mathrm{Tr}\ \gamma_5(\not{l}+\not{p}+M)\gamma^\nu(\not{l}+M)\gamma^\rho(\not{l}-\not{q}+M) \tag{5.45}$$

A trace of γ_5 with five γ^μ matrices always vanishes (at least two of the γ^μ have the same index, they can be anti-commuted to be next to each other, where they square to the identity. Then we have only three gamma matrices left). Hence we only get a contribution from the terms with four γ matrices, which yields $-8iM^2\epsilon^{\alpha\nu\rho\beta}p_\alpha q_\beta$.

This trace does not depend on l, so all that is left to is a scalar integral involving the three propagator denominators

$$S=\int\frac{d^4l}{(2\pi)^4}\frac{1}{[(l+p)^2-M^2]\,[l^2-M^2]\,[(l-q)^2-M^2]} \tag{5.46}$$

To make this well-defined we go to Euclidean space using a Wick rotation. Note that momenta $l_\mu l^\mu=l_0^2-\vec{l}^2$ are transformed to $-l^2$ in Euclidean space, if we replace l_0 by il_4. To keep track of the proper integration contours we first re-introduce the $i\epsilon$ terms in the propagators. We are interested in the limiting behavior of the expression for $M\to\infty$. In that limit the dependence on p and q can be ignored. After going to polar coordinates in four-dimensional Euclidean space we get

$$S=-i\frac{1}{(2\pi)^4}\int d\Omega_3\int_0^\infty dl\frac{l^3}{(l^2+M^2)^3}=-\frac{i}{8\pi^2}\frac{1}{M^2}\int_0^\infty dx\frac{x^3}{(x^2+1)^3}=-\frac{i}{32\pi^2M^2} \tag{5.47}$$

where the $-i$ comes from the Wick rotation and the three signs from the propagator denominators, and the $d\Omega_3$ integration is over the polar angles; this integral yields the surface area of a unit 3-sphere and is equal to $2\pi^2$. In the second step the integration variable was changed as $l=xM$. The indefinite integral is

$$\int dx\frac{x^3}{(x^2+1)^3}=-\frac{2x^2+1}{4(x^2+1)^2} \tag{5.48}$$

Hence we find

$$\begin{aligned}
ik_\mu V^{\mu\nu\rho}_{abc} &= -(p_\mu+q_\mu)\int \frac{d^4l}{(2\pi)^4} I^{\mu\nu\rho}_M(l,p,q)\text{Tr }T_aT_bT_c + (p,\nu,b)\leftrightarrow(q,\rho,c) \\
&= -(-8iM^2\epsilon^{\alpha\nu\rho\beta}p_\alpha q_\beta)S\text{ Tr }T_aT_bT_c + (p,\nu,b)\leftrightarrow(q,\rho,c) \\
&= \frac{1}{4\pi^2}\epsilon^{\nu\rho\alpha\beta}p_\alpha q_\beta\text{Tr }T_aT_bT_c + (p,\nu,b)\leftrightarrow(q,\rho,c) \\
&= \frac{1}{4\pi^2}\epsilon^{\nu\rho\alpha\beta}p_\alpha q_\beta\text{Tr }T_a\left\{T_b,T_c\right\}
\end{aligned}$$

If we set T^a, T^b and T^c equal to the identity matrix this yields Eqn. (5.39).

5.5.2 Anomalous Local Symmetries

If all three external lines are gauge bosons, one of the gauge symmetries cannot be an exact symmetry of the Lagrangian, because the triangle diagram is incompatible with three independent gauge transformations. Since gauge invariance is a crucial ingredient in the proof of renormalizability of gauge theories this is unacceptable, and hence we have to require that anomalies are absent or that they cancel. Cancellation is possible since the fermion trace is a sum over all fermions that couple to the external gauge bosons.

The currents we consider are of the form $i\bar\psi\gamma_\mu PT^a\psi$, where P is a linear combination of the identity matrix and γ_5. The trace over the Dirac indices splits thus into two terms, one without any γ_5 matrices, and one with a single γ_5. As indicated in the figure, there are two diagrams contributing to the amplitude under consideration. It is not hard to see that for the diagrams without a γ_5 the trace over the group representations is proportional to $\text{Tr}[T^a,T^b]T^c \propto f^{abc}$, whereas, as we have seen above, for the trace with a γ_5 the trace is proportional to $\text{Tr}\{T^a,T^b\}T^c$, which due to the cyclic properties of the trace is completely anti-symmetric in a, b and c. The terms proportional to f^{abc} contain infinities, which fortunately can be subtracted since the Lagrangian contains terms of this form as well. The symmetric terms are finite, but they do not satisfy the Ward identity Eq. (5.39) in all three indices simultaneously.

If we split all fermions in left and right-handed ones, their contribution to the anomaly will be with opposite sign if they are in the same representation. It is more convenient to assume that all fermions are left-handed. Then the complete group theory factor in the anomaly is proportional to $\text{Tr}\{T^a,T^b\}T^c$, where the trace is over the complete set of fermions. In the following all fermions are assumed to be left-handed. Writing all fermions in terms of left-handed components is another way of seeing that all anomalies cancel if there are only vector couplings: their left and right components are converted into two left-handed components with opposite charges. Therefore QED is safe. Furthermore, QCD is safe as well, because triplets (with representation matrix T^a) and anti-triplets (with representation matrix $-(T^a)^*$) have opposite contributions to the anomaly:

$$\begin{aligned}
\text{Tr}\{-(T^a)^*,-(T^b)^*\}[-(T^c)^*] &= -\text{ Tr}\{(T^a)^T,(T^b)^T\}(T^c)^T) \\
&= -\left[\text{ Tr}T^c\{T^b,T^a\}\right]^T = -\text{ Tr}\{T^a,T^b\}T^c
\end{aligned}$$

Let us first consider the situation that all three generators T^a, T^b and T^c are generators of the same simple factor G of the gauge group. The Lie algebras trace equals (see Appendix B)

$$\mathrm{Tr}\{T^a, T^b\}T^c = 2\ \mathrm{Str}\ T^a T^b T^c = 2I_3(R)d^{abc}\ , \tag{5.49}$$

where d^{abc} is a real tensor that is symmetric in three adjoint indices, and Str stand for the "symmetrized trace", defined in appendix (B). In general this is a trace over a reducible representation, in other words a sum over the traces contributed by each fermion in the problem. If a fermion is in a non-trivial representation of some other group G', the dimension of that representation should be taken into account as a multiplicity. The G-anomalies may cancel for two reasons: either $I_3(R)$ vanishes or the symmetric tensor d^{abc} does not exist for the group G. The vanishing of $I_3(R)$ can be a consequence of a non-trivial cancellation among several fermions, or it could happen that each fermion separately contributes zero. Note that in particular any real representation contributes zero, since the right-hand side of Eq. (5.49) is real, and on the left-hand side one may use that – in a suitable basis – $T^a = -(T^a)^* = -(T^a)^T$. The same is true for pseudo-real representations, satisfying $(T^a)^* = -CT^aC^\dagger$ for some unitary matrix C. Thus in particular the singlet and adjoint representations do not contribute to the anomaly.

If G has no symmetric tensor in three adjoint indices, there are no G-anomalies at all, for any fermion representation. This is automatically true if all G-representations are real or pseudo-real. This is the case for the gauge groups $SU(2), Sp(N)$, all exceptional groups except E_6 and all $SO(N)$ groups except $SO(4n+2), n \in \mathbf{Z}$. Most of the groups with complex representations do indeed have a non-vanishing tensor d^{abc}. This is true for all $SU(N)$ groups, $SO(2)$ and $SO(6)$. The remaining groups, E_6 and $SO(4n+2)$ for $n \geq 2$ have complex representations, but are nevertheless anomaly-free (*i.e.* $d^{abc} = 0$). Finally $U(1)$ groups have non-trivial anomalies, which are equal to the third power of the charge for each fermion (with the appropriate multiplicity as explained above).

The anomaly coefficients $I_3(R)$ are integers (provided d^{abc} is normalized in a reasonable way) which can have either sign. They can be looked up in tables (see *e.g.* [29]). If the group is not anomaly-free these coefficients are usually non-zero for any complex irreducible representation, with very few exceptions.

If T^a and T^b belong to the same factor G_1 of the gauge group, and T^c to a different one, G_2, then $\mathrm{Tr}\{T^a, T^b\}T^c = 2\ \mathrm{Tr}\ T^aT^b\ \mathrm{Tr}\ T^c$.

This relation holds for each irreducible representation (R_1, R_2) of $G_1 \times G_2$, and one sums over all irreducible components of the complete fermion representation at the end. Since $\mathrm{Tr}\ T^c = 0$ for simple Lie algebras, there can only be an anomaly if T^c is a $U(1)$ generator. If the full left-handed fermion representation is $\sum_i (R_i, q_i)$ the anomaly in the $U(1)$ current is thus proportional to $\sum_i I_2(R_i)q_i$.

Finally, if all three group generators belong to different gauge groups, there is only a contribution if all three are $U(1)$ generators, not embedded in a simple algebra.

To illustrate all this, let us see how it works for the Standard Model. The pure $SU(3)$ anomalies cancel because each family contains 2 triplets and 2 anti-triplets, and complex conjugate representations contribute with opposite signs. Cancellation of the pure $SU(2)$ anomalies is trivial, since $SU(2)$ is anomaly-free. The cancellation of the pure $U(1)$

anomalies is more interesting:

$$3.2.\left(\frac{1}{6}\right)^3+3\left(-\frac{2}{3}\right)^3+3\left(\frac{1}{3}\right)^3+2\left(-\frac{1}{2}\right)^3+1=0 \tag{5.50}$$

Note that the multiplicities due to the dimensions of $SU(3)$ and $SU(2)$ representations must (of course) be taken into account. The cancellation of the mixed $SU(3)$, $U(1)$ anomalies and $SU(2)$, $U(1)$ is also non-trivial. It is a simple exercise to check that these anomalies do indeed cancel.

5.5.3 Anomalous Global Symmetries

A generic gauge theory has a gauge group $G_{\text{gauge}} = G_1 \times G_2 \times \ldots \times G_n$, and fermions in representations $N^i(R_1^i, R_2^i, \ldots, R_n^i)$. Here R_j^i is an irreducible representation of the group labeled "j", i labels different representations, and N^i is the number of times such a representation occurs. We assume that the fermion representations were chosen so that all anomalies in local symmetries cancel. The global symmetry group of such a theory (if we consider only fermions and gauge fields, and no other couplings) is $G_{\text{global}} = U(N^1) \times U(N^2) \ldots U(N^k)$, where k is the number of distinct representations. All fermions are here assumed to be left-handed. If one does not do that, one would arrive at a smaller group, since one would overlook transformations between "L" and "R" fermions. For convenience we have assumed all fermions to be Weyl fermions in complex representations; if there are also Majorana fermions in real representations one will get orthogonal symmetries among them.

A natural question to ask now is if these global symmetries are preserved in the quantum theory. It turns out that they are in general affected by anomalies due to the same diagrams we have already computed. To see that think of global symmetry currents as vertices $c_\mu \bar{\psi}\gamma^\mu PT\psi$ added to the Lagrangian. Here P is some combination of the identity and γ_5 and T is some symmetry generator. The coefficients c_μ may be thought of as coupling constants. These terms in the Lagrangian then generate two-point vertices with two fermionic external lines, and combining these vertices with gauge boson-fermion three point couplings one can obtain triangle diagrams.

If one of the currents in the anomaly triangle represents a global symmetry, and the other two are local, we are forced to preserve the local symmetries (to maintain consistency) and choose the regularization of the diagram in such a way that the entire anomaly is in the conservation of the current of the global symmetry. Group theoretically these anomalies work exactly as the ones discussed above, but the interpretation is quite different. Anomalous global symmetries are acceptable, and in fact totally unavoidable. The only consequence is that a global symmetry of the classical action turns out not to be a symmetry quantum mechanically. Another way of saying this is: would it be possible to consistently gauge the global symmetry. If the answer is negative because of anomalies, then the global symmetry is not a symmetry of the quantum theory.

Hence triangle diagrams involving two generators of G_{gauge} and one of G_{global} will break part of the global symmetries. Since non-abelian generators are traceless, only $U(1)$'s can be broken in this way. In principle each non-abelian factor in G gauge is responsible for one anomaly. Furthermore, if there are m $U(1)$ factors in $G_{\rm gauge}$, they yield an additional $\frac{1}{2}m(m+1)$ in principle independent anomalies, since a triangle diagram can have two different $U(1)$ gauge generators. Hence in general one may expect $n-m+\frac{1}{2}m(m+1)=n+\frac{1}{2}m(m-1)$ global $U(1)$'s to be broken by anomalies. In practice there may be fewer, since the set of anomalous $U(1)$'s need not be independent. If this does not exhaust the set of available $U(1)$ symmetries, the remaining ones may be linearly combined into non-anomalous symmetries.

Even though a global current may be anomalous, the classical global symmetry means that at every vertex the charge is conserved. Hence an anomalous global symmetry is not broken to arbitrary order in perturbation theory since one can simply follow the charges through the diagram. However, the effects of the anomaly do appear non-perturbatively.

5.5.4 Global Anomalies in Field-Theoretic Form

The anomaly can be represented by a local counter-term involving the gauge fields

$$\partial_\mu J^\mu = \frac{g^2}{8\pi^2}\,{\rm Tr} F_{\mu\nu}\tilde{F}_{\mu\nu}\ , \tag{5.51}$$

where $\tilde{F}_{\mu\nu}=\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F_{\rho\sigma}$. The fields $F_{\mu\nu}\equiv F^a_{\mu\nu}T^a$ are of course in the representation of the fermions in the loop. The left-hand side of this divergence reproduces precisely Eq. (5.39) when written in momentum space. Since the left-hand side is itself the divergence of a current (see Eq. (4.9)) one can define a new current $\tilde{J}^\mu = J^\mu - \frac{g^2}{4\pi^2}K^\mu$ that is conserved. However, this does not change the fact that J^μ is not conserved, and furthermore K^μ is not gauge invariant: it is invariant under "small" gauge transformations, but not under certain "large" ones that cannot be continuously deformed to zero.

5.5.5 Global Anomalies in QCD $\times$ QED

Since the low-energy Standard Model $SU(3)\times U(1)$ has two gauge groups, one may expect two global $U(1)$'s to be affected. This is indeed true, although by totally different mechanisms. The oldest known example is an anomaly due to a triangle diagram with two external photons and an axial current. The axial current is $J^{3,{\rm A}}_\mu = \bar{u}\gamma_\mu\gamma_5 u - \bar{d}\gamma_\mu\gamma_5 d$ (the superscript "3" refers to the generator T^3 of $SU(2)$). Here we work in the approximation that the u and d quarks are massless, and all others are ignored. If we also ignore electromagnetism the QCD Lagrangian has an exact $U(2)\times U(2)$ chiral symmetry, as we have seen in section 4.1. The presence of distinct charges for the u and d quarks forbids rotations of these quarks into each other, but there is still an exact $U(1)^2\times U(1)^2$ chiral symmetry. These symmetries consist of the left and right phase rotations of the u and d

quarks separately. Out of these four we can make four independent linear combinations

$$\begin{aligned}
\tfrac{1}{3}\bar{u}\gamma_\mu\, u + \tfrac{1}{3}\bar{d}\gamma_\mu d &= J_\mu^{\mathrm{B}} \;\; \text{(Baryon number)} \\
\tfrac{2}{3}\bar{u}\gamma_\mu u - \tfrac{1}{3}\bar{d}\gamma_\mu d &= J_\mu^{\mathrm{em}} \;\; \text{(Electromagnetic current)} \\
\bar{u}\gamma_\mu\gamma_5 u - \bar{d}\gamma_\mu\gamma_5 d &= J_\mu^{3,\mathrm{A}} \\
\bar{u}\gamma_\mu\gamma_5 u + \bar{d}\gamma_\mu\gamma_5 d &= J_\mu^{\mathrm{A}}
\end{aligned}$$

The two vector currents, baryon number and electromagnetism, are conserved, because neither QCD nor QED has couplings with a γ_5. Note that if we insert the current $J_\mu^{3,\mathrm{A}}$ into a triangle diagram with two gluons, the contributions of the two terms cancel, because the u and d quark have the same couplings to the gluon. But their couplings to the photon are different, so the diagram with the current $J_\mu^{3,\mathrm{A}}$ and two photons is anomalous. Note that for the divergence of J_μ^{A} does get anomalous contributions with two gluons. This is why we choose these linear combinations. The effect of anomalies due to QCD is much stronger than those of QCD, so we look at a combination that is only affected by QED anomalies.

5.5.6 The $\pi^0 \to \gamma\gamma$ Decay Width

The symmetry corresponding to $J_\mu^{3,\mathrm{A}}$ is part of the axial $SU(2)$ symmetries that are spontaneously broken by QCD. This spontaneous breaking produces three pions as Goldstone bosons. In the limit of vanishing quarks masses and QED coupling all three pions are massless. If the QED coupling does not vanish, only $J_\mu^{3,\mathrm{A}}$ correspond still to an exact symmetry, so one would expect the corresponding Goldstone boson, π^0 to be exactly massless, while $\pi^\pm$ are slightly heavier due to electromagnetism. In the real world the quarks have a mass, lifting the pion masses to about 135 MeV, with π^0 slightly lighter than $\pi^\pm$.

The most interesting effect of the anomaly is not on the masses, but on the decay of the π^0. If this symmetry is exact, it would forbid the decay of $\pi^0 \to \gamma\gamma$ which is observed experimentally. This is a consequence of the Goldstone theorem. The pion field has the same matrix element with the two photon state as the divergence $\partial^\mu J_\mu^{3,\mathrm{A}}$ of the axial vector current, since the pion is the Goldstone boson of the axial symmetry. If the current is conserved the matrix element vanishes. If one includes the quark masses that break the chiral symmetry one gets a non-zero prediction for the decay width for $\pi^0 \to \gamma\gamma$ that is however much to small. The correct answer is that $\partial^\mu J_\mu^{3,\mathrm{A}}$ is not zero, but equal to an anomaly term involving the photon field, generated by a triangle diagram with an external axial vector current and two photons. Now the decay rate can be computed using the anomaly, whose normalization is known. The result is

$$\Gamma(\pi^0 \to \gamma\gamma) = \frac{\alpha^2 m_\pi^3 N_c^2}{576\pi^3 f_\pi^2} = 7.73 \text{ eV} \,, \tag{5.52}$$

where $f_\pi \approx 130$ MeV is the pion decay constant and N_c is the number of colors. The pion decay constant can be measured from the decay width of the charged pions to leptons. Hence the anomaly gives a parameter-free prediction of the $\pi^0 \to \gamma\gamma$ decay width. The agreement with the observed decay rate, 7.8 ± 0.2 eV is very good, which may be viewed as direct experimental evidence for the anomaly. Not only that, but the decay width is sensitive to the properties of the quarks in the loop. Originally, these computations were done with protons and neutrons instead of quarks. This gives the wrong answer. In QCD, the amplitude is – obviously– proportional to the number of quark colors, so that the width is proporional to N_c^2. Historically, this is one of the first ways it was discovered that there have to be tree distinct species of u and d quarks.

5.5.7 The Axial $U(1)$ Symmetry

The second anomaly is due to QCD, breaking the symmetry $U(1)_{\rm A}$ (generated by the current $J_\mu^{\rm A}$). This axial symmetry is spontaneously broken in the QCD vacuum, and hence one would expect an extra Goldstone boson with a mass close to that of the pions. However, there is no such boson. The anomaly by itself is not enough to explain this, since even in the presence of the anomaly one can define a conserved current, as we have seen above. The corresponding symmetry is spontaneously broken, and hence one would still expect a massless Goldstone boson. However, it is now essential that the new current is not gauge invariant. This allows non-perturbative instanton effects to break the symmetry explicitly, and remove any argument for the existence of a massless Goldstone boson.

5.5.8 Baryon and Lepton Number Anomalies

The complete Standard Model has one extra gauge group, $SU(2)$, and hence we may expect an additional independent anomalous current. This third anomaly is that of baryon number (or lepton number) with respect to the $SU(2)$ factor in the Standard Model gauge group. Baryon number (B) and lepton number (L) are global $U(1)$ symmetries of the classical Standard Model action. The values of these charges are $B = \frac{1}{3}$ for quarks, $-\frac{1}{3}$ for anti-quarks and 0 for all other particles, while $L = 1$ for leptons and -1 for anti-leptons. These symmetries are "vector-like", *i.e.* the current is $\bar{\psi}\gamma_\mu B\psi$ (or $\bar{\psi}\gamma_\mu L\psi$), in other words the charges are the same for left and right-handed fields. Since $SU(2)_{\rm w}$ has γ_5 couplings these symmetries are anomalous.

5.5.9 Proton decay by Instantons and Sphalerons

Also in this case instantons are responsible for the resulting global symmetry violation, but the effect is suppressed by a tunneling amplitude to a different vacuum, which is $\propto \exp\left(-4\pi \sin^2\theta_{\rm w}/\alpha\right) \approx 10^{-172}$. Such a e^{-1/g^2} behavior is typical for non-perturbative effects. Clearly there is no need to worry about the coefficient or the conversion to years. [The same tunneling amplitude also "suppresses" the $U(1)_A$ breaking discussed above, but this time with g_3 instead of g_2. Furthermore g_3 is to be evaluated at low energies, where it is not small at all.]

Baryon number can also be violated in high energy scattering, provided that the energy is large enough to build up field configurations ("sphalerons") that can go over the barrier (instead of tunneling through it). This mechanism also uses the $SU(2)$ anomaly.

The existence of the process itself is not terribly controversial, but the cross section is very hard to compute even approximately. Estimates vary between the unitarity bound and 10^{-70} times the unitarity bound.

One important thing to remember is that in any case in the standard model the proton is not a stable particle.

5.5.10 Anomaly-free Global Symmetries

Here we have chosen all three anomalous currents to lie in the quark sector of the Standard Model. This means that they contribute to the breaking of the QCD $U(6) \times U(6)$ chiral symmetries, just as the five mechanisms we discussed already. Those mechanisms had already broken the chiral symmetry group to just baryon number, and with the addition of three more breakings due to anomalies finally nothing is left. Of course there is a considerable amount of "overkill", but all mechanisms have their own specific consequences, and furthermore it is often instructive to study what happens if some of the origins of symmetry breaking are removed.

All other anomalies can now be removed by subtracting a suitable anomalous one. For example by making a linear combination of baryon number and lepton number we end up with the anomaly free combination $B - L$. This is an exact global symmetry of the Standard Model if there are no Majorana neutrino masses.

5.5.11 Mixed Gauge and Gravitational Anomalies

Instead of gauge bosons one may also have gravitons coupling to the external legs of the triangle diagram. In any even dimension such diagrams have anomalies if an even number of external lines is a graviton. In four dimensions the only such diagram has two external gravitons; the third vertex is then a vector current of either a global or a local symmetry with generator Q. The group theory trace is simply $\mathrm{Tr}Q$. One might expect the symmetry to be broken if this trace does not vanish.

If the symmetry is global this means that it might be broken by gravitational non-perturbative effects (which are probably completely negligible, except near black holes). If the symmetry is local one has a choice of giving up gauge invariance or general covariance. Of course one cannot argue that the latter should not be given up because one needs it to prove renormalizability of gravity, but on the other hand general covariance is not something one would give up easily. Note that although gravity is involved, there are no graviton loops, just loops of massless chiral fermions. The most likely conclusion is that theories with mixed gauge and gravitational are inconsistent and should be rejected. This means that the trace of all gauged $U(1)$'s in a sensible theory should vanish. Our only experimental test of this prediction, the Standard Model, does indeed meet that requirement.

One of the remaining exact global symmetries *is* broken by this anomaly, namely $B-L$. The anomaly in this symmetry can be canceled by adding right-handed neutrinos.

5.5.12 Other Anomalous Diagrams

In theories with non-abelian vector bosons there can also be anomalies due in box and pentagon diagrams. These anomalies play the rôle of establishing the correct non-abelian structure of the anomaly. For example the expression (5.51), when expanded in A_μ, yields terms of third and quartic order in the non-abelian fields. Furthermore one should consider the covariant derivative $D_\mu J^\mu$ instead of the ordinary derivative. None of this matters if the triangle anomalies cancel: then all these box and pentagon diagrams cancel as well. The study of the structure of anomalies is interesting in its own right, but mainly as a subject in mathematical physics.

One may also worry about higher loop diagrams. Fortunately there is an important theorem, the Adler-Bardeen theorem, that guarantees that higher loop diagrams do not contribute additional anomalies. Hence in the end only the triangle anomalies have to be considered.

5.5.13 Symplectic Anomalies

There is yet another kind of anomaly [35]. In some theories there are global gauge transformations (gauge transformations that cannot be connected to the identity in a continuous way) that change the sign of the path integral. This sign flip is always due to a fermion determinant changing sign. The most likely conclusion is that such theories are ill-defined, and hence not acceptable as a theory. The conditions for absence of such global anomalies are known. They are related to the fourth homotopy group of the gauge group, and this homotopy group is non-trivial only for $SU(2)$ and symplectic groups $Sp(N)$. Symplectic gauge groups are not encountered often in the literature, so only $SU(2)$ is really of interest to us. Since it occurs in the Standard Model we have to worry about non-trivial global anomalies. These anomalies are absent if the number of Weyl fermions in half-integral spin representations is even. An even number of fermions leads to an even number of sign changes, so that the anomaly cancels. The Standard Model respects this condition, and it does so within each family: there are four $SU(2)$ doublets per family.

5.6 Axions

Let us now return to the QCD θ-parameter discussed in section 4.1.2. We have already seen that it should be almost zero, and that within QCD alone it can simply be set equal to zero by imposing CP. However, since CP is not a symmetry of nature, this cannot really be justified. Furthermore, even if we put it equal to zero, non-vanishing corrections to θ are to be expected.

In fact, there is an effect which is not even small. To see why, we have to examine more carefully how we obtained the diagonal quark masses. In the Standard Model the only possible sources of CP violation are the CKM matrix for quarks and the PMNS matrix for leptons (see sections 4.3.3 and 5.2.6). Since θ is a strong interaction parameter the CKM matrix is most directly relevant. There is a CP-violating parameter in the CKM matrix if the number of families is three or larger. CP violation has been observed by

Cronin and Fitch in 1964 in the $K_0 - \bar{K}_0$ system, and more recently it has also been found in $B\bar{B}$ systems. Hence we know that the CP-violating parameter is non-zero. For this to work the Yukawa coupling matrices $g_{\mathcal{U}}$ and $g_{\mathcal{D}}$ cannot be real (if they are real the Lagrangian is manifestly CP invariant). Hence one expects the quark masses produced by the Higgs mechanism to be complex numbers. In section 4.3.3 we have made symmetry transformation to make the masses real, but the existence of anomalies in some symmetries forces us to verify if all those transformations were legitimate.

5.6.1 Phases in Quark Masses

Consider first a simpler example, namely a single quark with a complex mass, coupled only to QCD. We do not need to be specific here about the origin of the complex mass, but one may thing about complex, but diagonal, Yukawa couplings multiplied with a Higgs vev. The Lagrangian, including the θ-term is

$$\mathcal{L} = -\tfrac{1}{4} G^a_{\mu\nu} G^{\mu\nu,a} + \theta \frac{g_3^2}{16\pi^2} \mathrm{Tr}\ G_{\mu\nu} \tilde{G}^{\mu\nu} + i\bar{\psi} D_\mu \gamma^\mu \psi + m \bar{\psi}_L \psi_R + m^* \bar{\psi}_R \psi_L \tag{5.53}$$

Note that the last two terms are each other's conjugate, and hence the Lagrangian is real, even if m is complex*.

Let us write the mass as $m = |m| e^{i\alpha}$. In classical field theory one can make the mass real by means of the transformation

$$\begin{aligned} \psi_L &\rightarrow e^{i\alpha/2} \psi_L \\ \psi_R &\rightarrow e^{-i\alpha/2} \psi_R \end{aligned} \tag{5.54}$$

Note that there are other phase choices that achieve this, because simultaneous phase rotations of ψ_L and ψ_R have no effect at all. But whatever we choose, it is clear that we will have to transform ψ_L and ψ_R with different phases to make m real.

Of course this only works if the rest of the Lagrangian is invariant under this transformation. The fermion kinetic terms can be written as

$$i\bar{\psi} D_\mu \gamma^\mu \psi = i\bar{\psi}_L D_\mu \gamma^\mu \psi_L + i\bar{\psi}_R D_\mu \gamma^\mu \psi_R$$

and are manifestly invariant. The gauge kinetic terms do not even depend on ψ. So clearly the aforementioned phase transformation is a symmetry of the classical action. It is called an axial symmetry. As usual, there is a charge that generates the symmetry transformation, and the charge is related to a current, the axial vector current

$$J_5^\mu = \bar{\psi} \gamma^\mu \gamma_5 \psi$$

Without the γ_5 this is called the vector current:

$$J^\mu = \bar{\psi} \gamma^\mu \psi$$

* Complex masses are also used in the discussion of unstable particles. Then the real part is the mass and the imaginary part the decay width. But this has nothing to do with the present case. We have just a single quark that has nothing it can decay to. Hence its mass must be real.

If we assign electric charge to the fermion, the electromagnetic current is proportional to this vector current. Left- and right-handed components have the same electric charge, and hence the symmetry transformation acts in the same way on both of them. We can also write these transformations as

$$\begin{aligned} \psi &\rightarrow e^{i\alpha/2}\psi \qquad \text{(vector transformation)} \\ \psi &\rightarrow e^{i\alpha\gamma_5/2}\psi \quad \text{(axial vector transformation)}\ , \end{aligned} \tag{5.55}$$

where the last one is equal to (5.54).

Chiral Anomalies. But the axial symmetry is broken in the quantum theory, because there are one loop diagrams (the triangle diagrams computed in section 5.5.1) that do not satisfy axial current conservation. The way this symmetry is violated is given by Eq. (5.51). There it was written for an arbitrary current coupling to a triangle with two gauge bosos, without specifying where the γ_5's are in the triangle. In this case, we have a triangle with two gluons, which do not have γ_5 couplings, and hence the γ_5 can only come from the current itself. The relevant expressons are

$$\begin{aligned} \partial_\mu J_5^\mu &= \frac{g_3^2}{8\pi^2}\ \mathrm{Tr} G_{\mu\nu}\tilde{G}_{\mu\nu} \\ \partial_\mu J^\mu &= 0 \end{aligned}$$

In sections 2.2 and 2.3 we have seen how variations corresponding to currents J_μ affect the action. Let us adapt that discussion to the present case. Consider an x-dependent variation

$$\psi \rightarrow e^{i\alpha(x)\gamma_5/2}\psi$$

Now the action is not invariant, because the kinetic term of the fermions are not. We find

$$i\bar{\psi}\partial_\mu\gamma^\mu\psi \rightarrow i\bar{\psi}\partial_\mu\gamma^\mu\psi + i\bar{\psi}\partial_\mu[i\alpha\,(x)\,/2]\gamma^\mu\gamma_5\psi = \mathcal{L}_{\rm kin} - \tfrac{1}{2}\partial_\mu\alpha(x)J_5^\mu$$

Hence the change in the action is

$$\delta S = -\int d^4x \tfrac{1}{2}\partial_\mu\alpha(x)J_5^\mu = \tfrac{1}{2}\int d^4x\alpha(x)\partial_\mu J_5^\mu = \int d^4x\alpha\frac{g_3^2}{16\pi^2}\ \mathrm{Tr} G_{\mu\nu}\tilde{G}_{\mu\nu}$$

The x-dependence of α was just used here as a trick to do the computation. The final result also holds for constant α. Note that this term is of exactly the same type as the CP-violating θ term. So we discover that the operations needed to make the mass real leads to a shift in the value of θ to the value

$$\bar{\theta} = \theta + \alpha$$

Hence we can never observe θ and α separately, only the linear combination $\bar{\theta}$.

At first sight this make the problem worse. The mass terms seem *a priori* unrelated to the value of θ. So if we thought that we could solve the problem by finding an argument why $\theta = 0$, we just learned that we also need an argument why the masses are real. This might still look possible in this simple example, but – as already stated above – in the real world the masses are obtained from Yukawa couplings, that *must* be complex matrices in order to get CP-violation in the CKM matrix.

A massless up quark? But although this seems to make the problem worse, it also offers a first glimpse at possible ways out. On possibility is that $m = 0$. If the mass is zero, we can multiply it with an arbitrary phase. This phase then just shifts θ, and we can shift it to zero without encountering any change in the quark mass. It is sufficient to have just one such massless quark, because there is just one parameter θ to shift. Note that the electric dipole moment of the neutron, Eq. (4.11) vanishes if one of the light quark masses is zero (this formula was derived under the assumption that all other quarks are heavy, otherwise it would have been proportional to *all* quark masses).

But is there a massless quark in the real world? The lightest quark is the up quark and its mass is $m_u = 2.2^{+.6}_{-.4}$ MeV [8]. This is more than five standard deviations away from zero. Nothing about QCD would change qualitatively if $m_u = 0$, but it just does not seem to be true. Furthermore, if indeed m_u were to vanish this just leads to a problem that at first sight is as puzzling as $\theta = 0$: why would just one of the quark masses vanish exactly? Of course it is also possible that m_u is not exactly zero, but just small. It should then be small enough that the electric dipole moment of the neutron is below the current limit, with θ of order 1. This requires m_u of order 10^{-9} MeV. This is not only statistically very unlikely in view of the aforementioned experimental results, but it also looks theoretically very implausible (although that has not stopped people from pursuing this option).

5.6.2 The Peccei-Quinn Mechanism

A second way out suggests itself if we replace the complex mass by a vacuum expectation value of a complex field σ. This means that we consider the action

$$\mathcal{L} = -\tfrac{1}{4}G^a_{\mu\nu}G^{\mu\nu,a} + \theta\frac{g_3^2}{16\pi^2}\text{Tr}\ G_{\mu\nu}\tilde{G}^{\mu\nu} + i\bar{\psi}D_\mu\gamma^\mu\psi + g\sigma\bar{\psi}_L\psi_R + g^*\sigma^*\bar{\psi}_R\psi_L \tag{5.56}$$

Here g is a complex coupling constant and σ a complex field. In order to discuss the vacuum expectation value of σ we first need the potential. We choose

$$V(\sigma) = \mu^2|\sigma|^2 + \lambda|\sigma|^4 \tag{5.57}$$

Now let us assume that $\mu^2 < 0$ so that σ gets a non-trivial vacuum expectation value. The bottom of the "mexican hat" is at a value $se^{i\beta}$ for an arbitrary phase β, but we just make a convenient choice. Note that g is already complex, so we gain nothing by allowing yet another phase from the vacuum expectation value. So we set $\langle\sigma\rangle = s$, with s real. The value of s is of course determined by μ^2 and λ. Now we expand σ around the vacuum. One possible parametrization would be

$$\sigma = \frac{1}{\sqrt{2}}(s + \eta + ia)$$

where a and η are real fields. But it is more convenient to expand in the following way

$$\sigma = \frac{1}{\sqrt{2}}(s + \eta)\,e^{ia/s} \tag{5.58}$$

which is the same to first order in the fields, but this expansion makes the higher order terms come out in a nicer way: most of them disappear. This is possible because the action has a global continuous symmetry called a *Peccei-Quinn symmetry.* This is a shift symmetry of a due to a combined phase symmetry of σ, ψ_L and ψ_R. Its action on the fields is characterized by Peccei-Quinn charges. In this case these charges are $1, \frac{1}{2}$ and $-\frac{1}{2}$ for σ, ψ_L and ψ_R respectively.

Note the similarity with the expansion we made in the discussion of the Higgs mechanism, Eq. (3.17). It was used there to show that the phase degree of freedom disappears from the action. There is also a very essential difference: in the Higgs mechanism the phase degree of freedom becomes the longitudinal component of a massive gauge boson. Here that is not the case; indeed, there is no gauge boson coupling the σ. But the parametrization of σ is useful for the same reason. One immediate advantage is that the potential $V(\sigma)$ is manifestly independent of the field a. We can also make the field a disappear in the coupling to fermions. This requires to make a field transformation of the fermions

$$\begin{aligned} \psi_L &\rightarrow e^{i\alpha(x)/2s}\psi_L \\ \psi_R &\rightarrow e^{-i\alpha(x)/2s}\psi_R \end{aligned} \tag{5.59}$$

But we cannot make $a(x)$ disappear from the action completely, for two reasons. First of all $a(x)$ depends on x, and hence if we substitute our parametrization in the kinetic terms we will get terms proportional to ∂_μ. In the Higgs mechanism we make a gauge transformation (3.18) to remove such terms, but we do not have such transformations at our disposal here. The second reason is that the transformation (5.59) is anomalous. Hence it cannot be turned into a gauge transformation anyway. The result of the anomaly is that the transformation generates an additional term in the action

$$\delta S = \int d^4x \frac{a(x)}{s}\frac{g_3^2}{16\pi^2}\ \mathrm{Tr} G_{\mu\nu}\tilde{G}^{\mu\nu}\ , \tag{5.60}$$

Note that this is a dimension-5 operator: $a(x)$ is a boson field, and has dimension 1, and $G\tilde{G}$ has dimension 4. This is why the coupling constant in this term is proportional to s^{-1}. We should not be surprised to get dimension 5 operators, because we made a non-linear (exponential) field transformation. If one does that with any other bosonic field one gets an infinity of operators with dimension higher than 4.

Taking everything together we see now that $G\tilde{G}$ appears with a factor

$$\frac{a(x)}{s} + \theta + \alpha = \frac{a(x)}{s} + \bar{\theta}$$

This means that we have turned $\bar{\theta}$ into a dynamical variable. Rather than just a parameter in the Lagrangian, θ has become a field $a(x)$, and the value we observe is the vacuum expectation of that field. So if we could think of some dynamics that could fix the vacuum expectation value to a definite value, then we have determined the observed value of $\bar{\theta}$ dynamically. The field $a(x)$ is called the *axion* (the origin of the name will be explained below).

Multiple Quarks. The example discussed above is unrealistic in several ways. First of all we considered only one quark. This is easy to fix. We may generalize (5.56) to N fermions

$$i\sum_{i=1}^{N}\bar{\psi}^i D_\mu\gamma^\mu\psi^i + \sigma\sum_{i,j=1}^{N} g_{ij}\bar{\psi}^i_L\psi^j_R + \sigma^*\sum_{i,j=1}^{N} g^*_{ij}\bar{\psi}^i_R\psi^j_L \tag{5.61}$$

When σ acquires a vev s, this gives rise to mass matrices $M_{ij} = sg_{ij}$. The first thing to do is to diagonalize these mass matrices sg_{ij} using $SU(N)_L \times SU(N)_R$ transformations. Furthermore we can use diagonal $SU(N)$ transformations to make sure that all the eigenvalues have a common phase $e^{i\alpha}$. Since $SU(N)$ transformations have no overall phases, these transformations are not anomalous.

In the final step we remove the common phase, and then we encounter the anomaly, as discussed above. The phase is given by the determinant of the matrix M_{ij}, and is denoted as $\alpha = \arg\det M$. Then the Peccei-Quinn symmetry acts with charge $\frac{1}{2}$ on all N ψ^i_L and with charge $-\frac{1}{2}$ on all N ψ^i_R, and as before with charge 1 on σ. The triangle diagram now has N quarks contributing, so the anomaly will be N times as large. Hence the contribution of $a(x)$ to the action now becomes

$$\delta S = \int d^4x N\frac{a(x)}{s}\frac{g_3^2}{16\pi^2}\,\mathrm{Tr}G_{\mu\nu}\tilde{G}^{\mu\nu} \equiv \int d^4x \frac{a(x)}{f_a}\frac{g_3^2}{16\pi^2}\,\mathrm{Tr}G_{\mu\nu}\tilde{G}^{\mu\nu}\ , \tag{5.62}$$

5.6.3 General Axion Models

The Axion Decay Constant. Here we introduced a constant, usually called the *axion decay constant*. In the example discussed above it has the value $f_a = s/N$. The right-hand side of Eq. (5.62) is the canonical form of the axion-gluon-gluon coupling. In any axion model there will be such a term, and we define f_a so that with a canonically normalized kinetic term for a, $\frac{1}{2}\partial_\mu a\partial^\mu a$, the axion-gluon-gluon coupling has this form. This constant owes its name to an analogous constant in the pion effective Lagrangian. The pion has a similar interaction Lagrangian for the pion-photon-photon coupling, with a coefficient f_π called the pion decay constant. This constant determines the main decay mode of the π^0, the decay to two photons (see Eq. (5.52)).

The axion decay constant is the essential parameter of axion physics. In the concrete model discussed above f_a was related to the vacuum expectation value of a scalar by a definite numerical factor. But in a more general description we just treat f_a as a free parameter. It sets the scale for all axion physics. We will see that the mass of the axion and all of its couplings are proportional to $1/f_a$.

Axion Effective Action. In order to discuss axion physics without having to worry about specific models one uses the following effective action

$$\mathcal{L}_a = \frac{1}{2}\partial_\mu a\partial^\mu a + \frac{a}{f_a}\frac{g_3^2}{16\pi^2}\mathrm{Tr}G_{\mu\nu}\tilde{G}^{\mu\nu}, \tag{5.63}$$

This is the action for a free, massless boson with a non-renormalizable (dimension five) coupling to the gauge bosons. No matter which axion model one considers, one always ends up with an action of this form. The interaction term is generated by the anomaly of the axial current, from a triangle diagram with two gluons. In addition to these terms involving the axion field a there are the other terms involving $G\tilde{G}$, already mentioned above

$$\mathcal{L}_\theta = (\theta + \arg \det M)\frac{g_3^2}{16\pi^2}\mathrm{Tr}G_{\mu\nu}\tilde{G}^{\mu\nu}, \tag{5.64}$$

From this effective action one can see immediately how shifting the axion field by a constant can change the value of θ in the strong CP term.

The QCD-generated Axion Potential. Up to now it may have seemed that the potential of the field a is completely flat. It appears in the action only in the form of derivatives $\partial_\mu a$, plus the coupling to $G\tilde{G}$. But also in this coupling the dependence on a is through $\partial_\mu a$, because $G\tilde{G}$ is a total derivative, and we can move the derivative to $a(x)$ by partial integration. Hence classically the theory is invariant under shifts $a(x) \rightarrow a(x) + c$, for any real c. But this is not going to be a symmetry of the full quantum theory, because we also know that shifting the value of $a(x)$ changes θ. For different values of θ we will measure a different value of quantities like the electric dipole moment d_n of the neutron, so we really have different physics. And if the physics is different, the vacuum energy must be different as well. Hence somehow QCD creates a non-trivial potential $V(a)$ on top of the flat background we started with. This must be a non-perturbative effect, because perturbatively the shift symmetry is exact.

We also know that the non-perturbative physics is periodic in θ with periodicity 2π, so clearly the potential must be periodic as well. Furthermore $\langle\theta\rangle = \langle\frac{a(x)}{s} + \bar{\theta}\rangle = 0$ is a special point. In that point $d_n = 0$, and for small fluctuations around that point $d_n \propto \langle\theta\rangle$. Since we would not expect the vacuum energy to depend on the *sign* of $\langle\theta\rangle$, we expect that $\langle\theta\rangle$ is either a local minimum or a local maximum of the vacuum energy. More detailed arguments are needed to show that it is indeed a minimum [30]. The expansion of the potential around the minimum gives rise to a mass for the axion.

A real computation of the axion mass requires non-perturbative QCD physics but there is already something we can say simply because the potential is a periodic function of the dimensionless combination a/f_a. The simplest possibility is then

$$V(a) = \mathcal{F}\left[1 - \cos(a/f_a + \bar{\theta})\right]. \tag{5.65}$$

and this is indeed what one gets from computations. For dimensional reasons, there must be a pre-factor $\mathcal{F}$ of with dimension $[\text{mass}]^4$. This factor depends on the QCD scale and the quark masses. Using current algebra techniques (not discussed in these lecture notes) one can show that it is in fact equal to $(m_\pi f_\pi)^2$ times dimensionless ratios of quark masses. The latter ratios vanish in the limit where one quark mass goes to zero, because – as we have seen – in that limit QCD becomes invariant under shifts of the axion field.

If $V(a)$ is a function of a/f_a, even if it is not exactly a cosine, if follows that if we expand it around its minimum the first term is proportional to $(a/f_a)^2$. It follows that a rough approximation of the axion mass in terms of f_a is

$$m_a \approx m_\pi \frac{f_\pi}{f_a}$$

Problems with global symmetries. Since obviously flatness of the potential <u>without</u> non-perturbative QCD effects is essential, we need to rethink the potential (5.57). This looks like the most general potential of a complex field, but of course we do not really know if we should regard as a complex field, or two real fields σ_1 and σ_2, with $\sigma = \sigma_1 + i\sigma_2$. In the latter case, more general potentials are possible. Indeed, even the mass term could take the form $m_1^2\sigma_1^2 + m_2^2\sigma_2^2$. This would ruin the entire argument. The only way to justify the potential (5.57) is to insist on the phase symmetry $\sigma \to e^{i\alpha}\sigma$. But here we run into a potential contradiction with "folk theorems" in theories of gravity. It is generally believed that a theory of quantum gravity does not allow continuous global symmetries. The argument goes like this: a continuous global symmetry gives rise to exactly conserved charges. But if you throw such a charge into a black hole it is gone, and hence apparently not conserved. If the symmetry is local, *i.e.* a gauge symmetry, then each charge comes with an "electric" field that stretches out to infinity, and provides a permanent record of what went into the black hole.

The way out is that $\sigma \to e^{i\alpha}\sigma$ is not really a global symmetry. When combined with the action on fermions, it has an anomaly. But if one thinks in terms of general scalar fields in a theory of quantum gravity, this implies that there must be a somewhat mysterious feedback. Somehow the scalar potential "knows" that a certain global symmetry is allowed, because the phase symmetry must be realized on fermions to keep Yukawa couplings invariant, and the action on fermions is anomalous. This is generally considered to be the weakest point of the Peccei-Quinn mechanism. We can simply postulate a potential (5.57), but is this really consistent with a fundamental theory of quantum gravity? If quantum gravity abhors all continuous global symmetries, does it make sense to postulate such a potential at all? These are questions we cannot address without a concrete theory of quantum gravity.

5.6.4 Axions in the Standard Model

An important difference with the previous discussion is that in the Standard Model the Higgs is not a singlet, but an $SU(2) \times U(1)$ doublet. This makes little difference as far as mass diagonalization is concerned. Since quarks with charges $-\frac{1}{3}$ and $\frac{2}{3}$ do not mix, we do not get a generic 6×6 complex matrix, but it splits into two 3×3 blocks. Instead of $SU(6)_L \times SU(6)_R$ we have a symmetry $SU(3)_{\mathcal{U},L} \times SU(3)_{\mathcal{U},R} \times SU(3)_{\mathcal{D},R} \times SU(3)_{\mathcal{D},R}$ at our disposal. It acts exactly as in Eq. (4.28), except that there we used $U(3)_{\mathcal{U},L} \times U(3)_{\mathcal{U},R} \times U(3)_{\mathcal{D},R} \times U(3)_{\mathcal{D},R}$. Without using the four phase rotations, we can still bring the mass matrices in diagonal form with common phases:

$$\hat{m}_{\mathcal{U}} = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix} e^{i\xi_{\mathcal{U}}} \quad \hat{m}_{\mathcal{D}} = \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix} e^{i\xi_{\mathcal{D}}} \tag{5.66}$$

To get rid of the phases in the mass matrix we may use these extra phase transformations, but now we have to be careful with anomalies. We denote the four phases as $e^{i\alpha}, e^{i\beta}, e^{i\gamma}$ and $e^{i\delta}$ respectively. The phase transformations have the following effect on the quark masses

$$\begin{aligned}\hat{m}_{\mathcal{U}} &\rightarrow e^{-i\alpha}\hat{m}_{\mathcal{U}}e^{i\beta}\\ \hat{m}_{\mathcal{U}} &\rightarrow e^{-i\gamma}\hat{m}_{\mathcal{U}}e^{i\delta}\end{aligned}$$

Clearly the result depends only on $\beta-\alpha$ and $\delta-\gamma$. The transformation is anomalous if $\alpha-\beta+\gamma-\delta\neq 0$ (note that the transformations on opposite-handed fields contribute with opposite signs). It is easy to see that with non-anomalous transformations we can arrange that $\xi_{\mathcal{U}}=\xi_{\mathcal{D}}$, but in general we need an anomalous transformation to remove the overall phase. This transformation will generate, as in the example in section 5.6.2, a shift in θ. The shift is proportional to the phase of the determinant of the full 6×6 quark mass matrix M, $\arg\det M$. Hence we get

$$\bar{\theta}=\theta+\arg\det\, M=\theta+\arg(\det\,\hat{m}_{\mathcal{U}})(\det\,\hat{m}_{\mathcal{D}})$$

The implications are the same as before. It is hard to reconcile vanishing $\bar{\theta}$ with a complex CKM matrix. It is not impossible, and in particular it is possibly that the CKM matrix is complex and that det M is nevertheless real, but it is not clear how to arrange that in a fundamental way. There exist ideas in the literature exploring this route, but the resulting models look rather contrived.

No Peccei-Quinn Symmetry in the Standard Model. Now that we have formulated the problem in the full Standard Model we come back to the Peccei-Quinn mechanism explained above. Can it be realized in the Standard Model? What we need is a Peccei-Quinn symmetry: a continuous global symmetry that is only broken by an $SU(3)$-color anomaly. It must be global, and cannot be local, because it *must* be anomalous.

There is no such symmetry in the Standard Model. The fields involved in a putative PQ-mechanism are ϕ, $\psi_L^{\mathcal{Q},\alpha}, \psi_R^{\mathcal{U},\beta}$ and $\psi_R^{\mathcal{D},\beta}$ (see Eq. (4.24)). Let us assign these fields PQ-charges $q, p, r_{\mathcal{U}}$ and $r_{\mathcal{D}}$ respective; since all the quarks mix we cannot assign different charges to different families. Invariance of the Yukawa couplings requires that $-q-p+r_{\mathcal{U}}=0$ and $q-p+r_{\mathcal{D}}=0$. Hence we see that $2p-r_{\mathcal{U}}-r_{\mathcal{D}}=0$, but this combination is precisely the anomaly with respect to $SU(3)$. Note that if the left-handed field $\psi_L^{\mathcal{Q},\alpha}$ has PQ charge p, it contributes $2p$ to the anomaly because it is left-handed, and is a doublet, hence the factor 2. On the other hand $\psi_R^{\mathcal{U},\beta}$ and $\psi_R^{\mathcal{D},\beta}$ are right-handed, so they contribute to the anomaly with a minus sign, and without a factor 2. Hence we see that out candidate Peccei-Quinn symmetry has no anomaly with respect to $SU(3)$, and hence it cannot possible rotate θ. So there is no Peccei-Quinn symmetry.

There is a different way of arriving at the same conclusion. In the Standard Model the phase variation of the Higgs field ϕ is eaten by the Z-boson, so there is no axion left after Standard Model symmetry breaking. Indeed, in the Standard Model only one physical scalar is left, the Higgs scalar, which is massive and certainly not an axion. Hence if we want to solve the strong CP problem in this way (and in fact in any other known way) we have to extend the Standard Model.

Axions in Extensions of the Standard Model. One very simple solution is suggested by the previous discussion. In section 5.6.1 we saw that a single quark coupling to a complex singlet Higgs can do the job. This cannot be one of the known quarks, because we already know experimentally that they *must* get their mass from a doublet Higgs. But we can postulate a new quark χ with left- and right-handed components that couple to $SU(3)$ in the usual way, and that gets its mass from from a new scalar Higgs field σ, exactly as in (5.56). The PQ charges of σ, χ_L and χ_R are respectively $1, \frac{1}{2}, -\frac{1}{2}$, just as in the example discussed above. The new quark must be heavy enough to have escaped observation so far, but this by itself is not a big challenge, because it gets its mass from a different Higgs boson than the known quarks.

But adding extra weak-singlet quarks may look a bit awkward. An example that works without adding extra quarks is the two-Higgs model. Instead of the single Higgs of the Standard Model one introduces two Higgses, one that couples to the down quarks (and the leptons), and one that couples to the up quarks. That this should work is already clear from the previous section, because now we can assign different PQ-charges to $\phi_{\rm u}$ and $\phi_{\rm u}$, whereas with a single Higgs field the charges of ϕ and $\mathbf{C}\phi^*$ must be opposite.

Considering only the quark sector, the Yukawa couplings are thus

$$\mathcal{L}_Y = -g_{\mathcal{U}}^{\alpha\beta}\bar{\psi}_L^{\mathcal{Q},\alpha}\ \phi_{\rm u}\psi_R^{\mathcal{U},\beta} - g_{\mathcal{D}}^{\alpha\beta}\bar{\psi}_L^{\mathcal{Q},\alpha}\phi_{\rm d}\psi_R^{\mathcal{D},\beta} + \text{c.c.}\ , \tag{5.67}$$

where $\phi_{\rm u}$ takes over the rôle of $C\phi^*$. Hence $\phi_{\rm d}$ is in the usual Higgs boson representation $(1,2,\frac{1}{2})$ whereas $\phi_{\rm u}$ is in the representation $(1,2,-\frac{1}{2})$. The new element is that now we can rotate the phases of the up and down mass matrices independently, by phase rotations of $\phi_{\rm d}$ and $\phi_{\rm u}$, whereas previously we could only rotate the Higgs field ϕ. In the Standard Model the down quarks couple to ϕ and the up quarks to $C\phi^*$; therefore any phase rotation of ϕ cancels in $\arg\det M$. Hence in the Standard Model $\arg\det M$ is fully determined by the Yukawa couplings, but in the Peccei-Quinn model it is not.

To define the Peccei-Quinn symmetry of this theory we can choose charges 1 for both $\phi_{\rm d}$ and $\phi_{\rm u}$, charges $\frac{1}{2}$ for all components ψ_L and charges $-\frac{1}{2}$ for all components ψ_R. This rotates left and right components of the quarks with opposite phases, and hence it is anomalous and can rotate the θ-angle away. For this to be a symmetry of the entire Lagrangian, it must be a symmetry of the Higgs potential. Terms like $\phi_{\rm d}\phi_{\rm u}$, $(\phi_{\rm d}\phi_{\rm u})^2$ or $\phi_i^\dagger\phi_i\phi_{\rm d}\phi_{\rm u}$ (which are allowed by $SU(2)\times U(1)$) must be absent, since they are not invariant under this symmetry. Let us assume that the Higgs potential has that property. This can be imposed by requiring that the Peccei-Quinn symmetry is an exact global symmetry of the classical Lagrangian. This means that it is preserved by all vertices, and hence it will be preserved by all loop diagrams;* so if the unwanted terms in the potential are absent at tree level, they will not be generated by loop diagrams.

* This includes the anomalous triangle diagram, because this has no external line attached to the top of the triangle, and hence cannot contribute to any perturbative amplitude computation.

The Axion in the Two-Higgs Model. Let us see how the axion appears in the two-Higgs model introduced above. Weak symmetry breaking in this two-Higgs model occurs analogously to the one-Higgs model with fields ϕ and $C\phi^*$. However, since these are now two unrelated fields, their absolute value of their vevs are now unrelated as well:

$$\langle\phi_{\rm d}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v_d\end{pmatrix} \quad ; \quad \langle\phi_{\rm u}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}v_u\\ 0\end{pmatrix} \tag{5.68}$$

In combination with the aforementioned chiral phase transformations of the fermions the theory has a global symmetry which acts non-trivially on the vacuum. Hence it is spontaneously broken, and one gets a massless Goldstone bosons corresponding to the symmetry, the axion. It is easy to see which linear combination of the variations of $\phi_{\rm u}$ and $\phi_{\rm d}$ around the the vacuum (5.68) is the axion field. One expands around the vacuum as

$$\langle\phi_{\rm d}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}\sigma_d + i\rho_d\\ v_d+\eta_d+i\xi_d\end{pmatrix} \quad ; \quad \langle\phi_{\rm u}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}v_u+\eta_u+i\xi_u\\ \sigma_u+i\rho_u\end{pmatrix} \tag{5.69}$$

One linear combination of the phase fluctuations ξ_u and ξ_d is eaten by the Z-boson, namely $(v_u\xi_u - v_d\xi_d)$. However, because of the Peccei-Quinn symmetry the combination $a = (v_d\xi_u + v_u\xi_d)$ remains massless. This is the axion field. Note that the two complex doublet fields ϕ_u and ϕ_d have eight real components. Three of these are eaten by the Z and $W^{\pm}$ field, and hence five physical fields are left. They consist of two neutral massive scalars (one of which should correspond to the observed Higgs boson), a massive charged scalar, and a massless scalar, the axion.

However, as in the example we discussed earlier, it is more convenient to make a non-linear expansion similar to (5.58).

$$\langle\phi_{\rm d}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}\sigma_d + i\rho_d\\ v_d+\eta_d-i\rho_d\end{pmatrix}e^{ia(x)/v_a} \quad ; \quad \langle\phi_{\rm u}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}v_u+\eta_u-i\rho_u\\ \sigma_u+i\rho_u\end{pmatrix}e^{ia(x)/v_a} \tag{5.70}$$

The rest of the discussion then goes as before. We can remove the dependence of the Lagrangian on a apart from derivatives and a coupling to $F\tilde{F}$. Of course the value of v_a is related to v_u and v_d. One can determine this relation by expanding the kinetic terms of $\phi_{\rm u}$ and $\phi_{\rm d}$ and requiring that the resulting kinetic terms for the axion field have the canonical form $\frac{1}{2}\partial_\mu a \partial^\mu a$. It turns out that $v_a = \sqrt{v_u^2+v_d^2}$. Then v_a is related to f_a by a numerical factor that depends on the fermions in the anomaly triangle, because ultimately the coupling to QCD must be brought to the canonical form (5.62).

5.6.5 The Mass of the Original QCD Axion

The two-Higgs model is the simplest and arguably the most natural way to realize Peccei-Quinn symmetry in the Standard Model. The first calculation of the axion mass was done by Weinberg [32] and Wilczek [34]. The results presented here are from the first of these papers, and were computed using the two-Higgs model. An approximation was used where the u, d and s quarks are light. One has to consider the mass matrix of the bosons

$\bar{u}_R u_L, \bar{d}_R d_L, \bar{s}_R s_L$ as well as the phases of ϕ_d and ϕ_u. In the absence of instantons (non-perturbative QCD-contributions), quark masses and W, Z-bosons all these five particles are massless Goldstone bosons. Due to instantons one combination, the η', is not a Goldstone boson in any reasonable approximation (see section 5.5.7); one combination, to first approximation the relative phase of ϕ_d and ϕ_u, is eaten by the Z; one combination becomes the π^0 and another one the η; and finally the fifth linear combination, essentially the common phase of ϕ_d and ϕ_u, is the axion. For its mass Weinberg finds

$$m_A = \frac{N m_\pi f_\pi}{2\sqrt{m_u + m_d}} \left[\frac{m_u m_d m_s}{m_u m_d + m_u m_s + m_d m_s}\right]^{1/2} \frac{2^{1/4} G_F^{1/2}}{\sin 2\alpha} \ . \tag{5.71}$$

Here N_f is the number of flavors (meanwhile known to be six), G_F the Fermi constant ($G_F = \frac{1}{8}\sqrt{2}\frac{g_2^2}{M_\mathrm{W}^2}$), f_π the pion decay constant measurable in pion decay, and α parametrizes the ratio of the v.e.v's of the two scalars ϕ_d and ϕ_u in Eqn. (5.68): $\tan\alpha = v_d/v_u$. Apart from this, all parameters in the formula are known, and in particular there is no unknown QCD instanton-generated matrix element appearing. All QCD effects are encapsulated in the pion mass and the pion decay constant. This is possible because the pion is a pseudo-scalar pseudo goldstone boson, just like the axion. Hence measured parameters of the pion and its properties can be used in the computation of the axion mass.

Numerically one finds for the mass of the axion (for $N_f = 6$).

$$m_A \approx \frac{(140 \text{ keV})}{\sin 2\alpha} \ . \tag{5.72}$$

Note that the axion mass is proportional to m_π so that it vanishes in the chiral limit, and to the masses of the three "light" quarks. The latter dependence is a consequence of the fact that if one of the quarks is massless the theory becomes independent of θ (and hence $\bar{\theta}$) as discussed before. Then the potential is flat in the axion direction, and hence the axion is massless. One cannot take the other quark masses to zero in this formula because the calculation was done in the limit where their masses are much larger than the QCD scale.

This axion is not stable and would decay into two photons with a lifetime of about 10^{-2} second. Experimentally such a particle has not been seen. This was the situation in 1978. But as we will see in a moment, that is not quite the end of the story. This is why we called this the "original QCD axion" in the title of this subsection.

Historical Remarks. Historically, a series of interesting mistakes was made concerning the axion and its mass. Peccei and Quinn overlooked the axion completely. They discovered the mechanism, but did not realize that it always predicts a light scalar. Weinberg [32] and Wilczek [34] did realize that there had to be a light scalar (and Wilczek gave the axion its name, which he chose because it is a pseudo-Goldstone boson of an axial symmetry) and estimated its mass. But these authors then made the mistake of assuming that the value of f_a is somehow related to the weak scale. This is an easy mistake to make, because in the example of section 5.6.2 the value of f_a is related to s, the vacuum expectation value of σ, which determines the mass of one or more quarks. In the two-Higgs model, f_a is directly related to the Higgs vev. If one makes the assumption that f_a is related to the weak scale, f_a could be about 100 GeV, while f_π is about 130 MeV. This agrees with Weinberg's formula, because in that case $f_a \approx v \approx (G_F)^{-\frac{1}{2}}$. This would make the axion about a factor 1000 lighter than the pion. This gives a mass of ≈ 100 KeV, in agreement with the more precise calculation of Weinberg.

However, several authors [19, 28, 36, 6] realized a few years later that the axion scale does not have to be related to the weak scale at all. Indeed, the scalar σ introduced above is *not* the Standard Model Higgs field, as we have seen. If σ is just an additional scalar field, its vacuum expectation value can be increased so that the axion mass and couplings go to zero. In this way one can hope to make the axion "invisible".

5.6.6 Invisible Axions

So how can we get a value of f_a that we can adjust as we like? Of course we can just postulate such a coupling in an effective field theory, but it would be much more satisfactory to have a concrete renormalizable model (*i.e.* a model with only operators of dimension equal to four or less, that generates the required dimension five operator via anomalies).

We have already seen an example earlier, namely adding a new heavy quark coupling to a new singlet Higgs σ. Since σ gets a vev that is unrelated to the Standard Model Higgs vev, we can give this extra quark any mass we want. In particular we can make it very heavy, so that it escapes all experimental bounds, and simultaneously we can make the axion very light and weakly coupled, by making f_a large. A model of this kind was first proposed by Kim [19], and soon thereafter in [28], and this class is known as KSVZ models. A different approach, proposed in [36] and [6], is to add an additional scalar singlet to $\phi_{\rm u}$ and $\phi_{\rm d}$. This new scalar gets a large vev. We will not discuss this in detail, since the main features of both models are captured by the axion effective action. But it is important to know that this can be realized.

It is possible to make the axion completely "invisible", both to (current) experiments on earth as well as with respect to cosmological implications. For this to happen its mass must lie in a fairly narrow window, $6\ \mu\text{eV} < m_a < 6$ meV, corresponding to an axion scale f_a smaller than about 10^{12} GeV and larger than 10^9 GeV. Below this window on the f_a scale the axion would be observable, for example because stars and the sun would loose tot much energy by axion emission. For f_a too large the axion would give too large a contribution to cold dark matter. This sounds counterintuitive, because in that limit the axion is extremely light. Hence one one expect them to produce a gas of relativistic particles who contribute to hot dark matter, and whose contribution decreases with mass. But it is just the other way around. This is because the contribution to dark matter should not be thought of in terms of axions as particles, but as axion field oscillating coherently around the minimum of the potential.

Despite the discouraging name "invisible axion", several experiments are underway to try and find axions. This usually involves the axion coupling to two photons, which arises from an anomaly diagram, analogous to the π_0 coupling to two photons (see next subsection). The hope is to use strong electromagnetic fields to make the axion collide with a photon, and emit another photon.

5.6.7 Two-photon coupling

An allowed term in the effective action is the coupling to two photons

$$\mathcal{L}_{a\gamma\gamma} = \frac{a}{f_a'} \frac{e^2}{16\pi^2} F_{\mu\nu} \tilde{F}^{\mu\nu}, \tag{5.73}$$

where $F_{\mu\nu}$ is the electromagnetic field strength tensor. In specific models such a term is generated by anomaly triangles with two photons, coupling to the axial current times the Peccei-Quinn charge of the fermion. If there are no fermions with electric charge coupling to the axion, this could in principle vanish. However, first of all the axion *must* couple to quarks (either the ones we know or additional, heavy ones) and the standard rules of Standard Model charge quantization make it very difficult to make an electrically neutral quark. Secondly, the axion mixes with mesons, for example the π_0, and hence it will have a two-photon coupling via this mixing. In an effective theory there is therefore no good argument to set this coupling to zero, but it could happen to be small. This two-photon coupling is important in attempts to observe axions. The vast majority of axion search experiments rely on it.

There is indeed a large number of such experiments underway. Most use the fact that via the two photon coupling, an axion in a strong magnetic field can convert to a photon or vice-versa (think of the magnetic field as one of the two photons). An intriguing example is the “shining through a wall” phenomenon. One aims photons at a wall with strong magnetic fields on both sides of that wall. One magnetic field may convert the photon to an axion; this axion can pass through the wall because it barely interacts with matter; and on the other side of the wall the axion is reconverted into a photon, giving the impression that the photon passed through the wall. This process depends on two unlikely events (two photon-axion conversions) but certain regions of the parameter space of f_a and the axion-photon coupling can be explored. Other types of experiments are helioscopes, looking for axions coming from the sun, and haloscopes, looking for axion dark matter in the halo of our galaxy.

5.6.8 Axion-electron coupling

An axion electron coupling (and more generally the axion fermion coupling) is generated if the left- and right-moving components of a fermion have different Peccei-Quinn charges. The most important of these couplings is the coupling to electrons. It enters in certain constraints on axions, such as cooling of white dwarfs. Global Peccei-Quinn tranformations leave the action invariant, as they should, but if we try to remove the axion field by making a field-dependent transformation the kinetic terms are not invariant. Concretely, suppose

$$\begin{aligned} \psi_L &\to e^{iQ_L a(x)/2s}\psi_L \\ \psi_R &\to e^{iQ_R a(x)/2s}\psi_R \end{aligned}$$

Then we find

$$i\bar{\psi}_L\gamma_\mu\partial^\mu\psi_L + i\bar{\psi}_R\gamma_\mu\partial^\mu\psi_R \to -\tfrac{1}{2}\partial_\mu a(x)\left[Q_L\bar{\psi}_L\gamma^\mu\psi_L + Q_R\bar{\psi}_R\gamma^\mu\psi_R\right]$$

This can be written as

$$\tfrac{1}{4}\partial_\mu a(x)\left[(Q_L - Q_R)\bar{\psi}\gamma^\mu\gamma_5\psi + (Q_L + Q_R)\bar{\psi}\gamma^\mu\psi\right]$$

Conservation of the vector current implies that the second term vanishes upon partial integration. But for a massive particle the axial vector current is not conserved:

$$\partial_\mu\bar{\psi}\gamma^\mu\gamma_5\psi = 2m\bar{\psi}\gamma_5\psi$$

Note that this is simply due to the mass, and not the chiral anomaly. We want to apply this result to the electron, which does not couple to QCD. One can do the same computation for quarks, but then there will be two terms, one proportional to the quark mass, plus the coupling to $G\tilde{G}$ generated by the anomaly.

Hence the final result for the electron-axion coupling is

$$-\frac{1}{2}m_e(Q_L - Q_R)a(x)\bar{\psi}\gamma_5\psi$$

Note that this depends on the Peccei-Quinn charges of the electron, which could be zero. If they are zero, the fact that the axion mixes with mesons does not change the result, because mesons do not have a direct coupling to electrons. Hence it is not guaranteed that axions couple to leptons.

5.6.9 Generic Axions

The name axion is also used in a more general sense for pseudo-scalar pseudo-Goldstone bosons. The two "pseudo"'s here have a different meaning. A pseudo-scalar is a scalar that is odd under parity. Such a particle may have couplings to gauge field combinations of the form $aG\tilde{G}$. Another generic feature of axions is an approximate shift symmetry, $a \to a + x$. This would imply the existence of massless Goldstone bosons if the symmetry were exact. But folk theorems about gravity suggest that exact global symmetries cannot exist. Hence one would expect these symmetries to be broken by some non-perturbative effect, as QCD does for the Peccei-Quinn axion. Then the Goldstone boson acquires a mass and becomes a pseudo-Goldstone boson. Particles of this type exist in abundance in some realizations of string theory. The axion discussed so far, introduced to solve the strong CP problem, is usually called the QCD axion in order to distinguish it from generic axions. The constraints on QCD axions are more severe than on generic axions.

If there is more than one axion, their action and coupling to QCD will take the form

$$\mathcal{L}_a = \frac{1}{2}\partial_\mu a^i \partial^\mu a^i + \frac{a^i}{f_a^i}\frac{g_3^2}{16\pi^2}\mathrm{Tr}G_{\mu\nu}\tilde{G}^{\mu\nu}, \tag{5.74}$$

In principle each action might have a coupling to QCD, and this coupling depends on the way the axion interacts with colored particles. However, in this situation we may define the QCD axion as

$$\frac{a}{f_a} = \sum_i \frac{a^i}{f_a^i} \ , \tag{5.75}$$

and we choose for all other axions a basis that is orthogonal to this. Then only the QCD axion gets a mass from QCD effects, and all other remain massless.

The other axions must however somehow get a mass in another way, or we would end up with an ungauged continuous symmetry. So one may ask what happens if we assume that all axions have an additional explicit mass term, $\frac{1}{2}m_i^2 a_i^2$. In principle this does not have to be diagonal, but we may assume that the coupling to QCD we wrote above is in terms of eigenstates a_i of the explicit masses. The QCD generated axion potential plus the mass terms is in this case (we drop the "axion" label a on f_a^i)

$$V(a_i) = \mathcal{F}\left[1 - \cos\left(\sum_i \frac{a_i}{f_i} + \bar{\theta}\right)\right] + \tfrac{1}{2}\sum_i m_i^2 a_i^2. \tag{5.76}$$

Here $\mathcal{F}$ is the parameter introduced in (5.65); its value is roughly $m_\pi^2 f_\pi^2$. The equations of motion determining the minimum of the combined potential are

$$\frac{\mathcal{F}}{f_j}\sin\left(\sum_i \frac{a_i}{f_i} + \bar{\theta}\right) + m_j^2 a_j = 0$$

Multiplying with f_j and subtracting the equations from each other we find

$$f_j m_j^2 a_j = f_k m_k^2 a_k = -\mathcal{F}\sin\left(\sum_i \frac{a_i}{f_i} + \bar{\theta}\right)$$

We can express all other a_j in terms of a_1. Then the equation becomes

$$f_1 m_1^2 a_1 = -\mathcal{F}\sin\left(\frac{a_1}{f_1}\sum_i \frac{m_1^2 f_1^2}{m_i^2 f_i^2} + \bar{\theta}\right)$$

Define $x = f_1 m_1^2 a_1$ and

$$R = \sum_i (m_i^2 f_i^2)^{-1}$$

Note that $\mathcal{F}$ and x have dimension 4 and R has dimension -4. Then the equation reads

$$x = -\mathcal{F}\sin\left(xR + \bar{\theta}\right)$$

We are only going to solve the problem if the argument of the sine is small. So let us assume that it is, and check afterwords if this assumption is correct. If the argument is small, we can expand the sine, and we find

$$x = -\frac{\mathcal{F}}{1 + \mathcal{F}R}\bar{\theta}$$

Hence the argument of the sine is

$$xR + \bar{\theta} = \frac{\bar{\theta}}{1 + \mathcal{F}R}$$

This is of course also the value of the argument of the sine at the minimum, or in other words the physical θ angle we would observe. We must assume that $\bar{\theta}$ is of order 1, so the condition for the argument of the sine being small is that $\mathcal{F}R \gg 1$. Assume the axion with minimal value of $m_i^2 f_i^2$ has label k. Then $\mathcal{F} \gg m_k^2 f_k^2$. Now $\mathcal{F}$ is a QCD parameter approximately equal to $m_\pi^2 f_\pi^2$, so we see that the condition for the validity of the approximation is that there is one axion label k so that for that axion $m_\pi^2 f_\pi^2 \gg m_k^2 f_k^2$. Furthermore the observed θ-angle is

$$\theta_{\rm phys} = \frac{\bar{\theta}}{1 + \mathcal{F}R} \approx \frac{m_k^2 f_k^2}{m_\pi^2 f_\pi^2}\bar{\theta}$$

The approximation is valid if there is just one lightest axion separated by a substantial gap from the next-to-lightest one. If there are more light ones the value of $\theta_{\rm phys}$ becomes smaller; for example for M degenerate lightest axions with the same values of both m_i and f_i the result is reduced by a factor M. Observe that $\theta_{\rm phys}$ approaches zero if m_k goes to zero. The reason for this is clear: in that limit we obtain an axion with an exact shift symmetry. The dependence on f_i is also clear: for $f_i \to \infty$ the axion decouples from QCD, so then even a very light axion becomes useless. If we make the axion scale very large, we know that that QCD generated mass of a single axion is $M_{a_i} \approx m_\pi f_\pi / f_a^i$. The condition for small $\theta_{\rm phys}$ reads then

$$m_k \ll M_{a_k}$$

which must hold for at least one axion. To get the required tuning of $\bar{\theta}$ to a value smaller than 10^{-10} we need $m_k < 10^{-10} M_{a_k}$. In words, there must be at least one axion whose explicit mass is ten orders of magnitude smaller than its QCD-generated mass. The existence of axions with intermediate masses is irrelevant.

Finally we compute the masses of the axions. The mass matrix is given by the second derivative matrix at the minimum. The result is

$$M_{ij}^2 = \frac{\mathcal{F}}{f_i f_j}\cos\left(\theta_{\rm phys}\right) + m_i^2 \delta_{ij} \approx C m_\pi^2 \frac{f_\pi^2}{f_i f_j} + m_i^2 \delta_{ij}$$

where C is a factor of order 1, proportional to the quark masses. To simplify the notation, define

$$M_i = \frac{\sqrt{\mathcal{F}}}{f_i}$$

The first term is a matrix of the form $M_i M_j$, and has only one non-zero eigenvalue, with eigenvector M_i, and eigenvalue proportional to $\vec{M}^2$. All vectors orthogonal to $\vec{M}$ have eigenvalue 0. Hence without the mass terms m_i there is one axion with mass $\sum_i M_i^2$, and all the others have zero mass. For m_i large, we can ignore the first term. Hence heavy axions just keep their explicit mass m_i, and do not contribute to the Peccei-Quinn mechanism.

Let us analyse this explicitly for 2 axions. The QCD-generated mass contribution plus the explicit mass term together generate a mass matrix of the following form

$$\begin{pmatrix} M_1^2 + m_1^2 & M_1 M_2 \\ M_1 M_2 & M_2^2 + m_2^2 \end{pmatrix}$$

The eigenvalues are

$$\tfrac{1}{2}\left(M_1^2 + M_2^2 + m_1^2 + m_2^2 \pm \sqrt{(M_1^2+M_2^2)^2 + (m_1^2 - m_2^2)^2 + 2(M_1^2 - M_2^2)(m_1^2 - m_2^2)}\right)$$

There are two cases of special interest: if one of the m_i is large with respect to the QCD-generated axion masses M_i, and if both m_i are small.

Case 1: one heavy axion. We see that if m_2 is much larger than all other masses the dominant term in the argument of the square root is m_2^4, and hence we can expand it as follows

$$m_2^2 \sqrt{1 - 2\frac{m_1^2}{m_2^2} - 2\frac{M_1^2 - M_2^2}{m_2^2}} \approx m_2^2 - m_1^2 - M_1^2 + M_2^2$$

Hence the eigenvalues are m_2^2 and $M_1^2 + m_1^2$. The latter is the QCD axion mass. The PQ mechanism will only work if $m_1 < 10^{-10} M_1$, as seen above. The value of M_2 is irrelevant. It is determined by the axion coupling of a_2, but a_2 decouples and does not participate in the PQ-mechanism.

Case 2: two light axions. If both axions are light, the dominant term in the square root is $M_1^2 + M_2^2$, and we get the approximation

$$(M_1^2 + M_2^2)\sqrt{1 + 2(m_1^2 - m_2^2)\frac{M_1^2 - M_2^2}{(M_1^2 + M_2^2)^2}} \approx M_1^2 + M_2^2 + (m_1^2 - m_2^2)\frac{M_1^2 - M_2^2}{M_1^2 + M_2^2}$$

Then one eigenvalue is $M_1^2 + M_2^2$ and the other is

$$\frac{m_1^2 M_2^2 + m_2^2 M_1^2}{M_1^2 + M_2^2} = \frac{m_1^2 f_1^2 + m_2^2 f_2^2}{f_1^2 + f_2^2}$$

Note that now the large eigenvalue is the one of the QCD axion, and it has the expected mass, $M_1^2 + M_2^2$. The light particle has a mass much less than the axion mass. Note that we need $m_i \ll M_j$ only to make the square root expansion valid. It is sufficient that only one of the two axions has a mass $\ll M_i$ for the PQ-mechanism to work, as we have seen above.

5.6.10 Multiple gauge group factors

One may wonder why we only consider the $SU(3)$ factor here. The discussion of axion started with the desire to solve the strong CP problem, which concerns the $G\tilde{G}$ term in QCD. But we could have introduced a similar term in the weak interactions and also a term $F\tilde{F}$ in the Y-factor of the standard model. So what happens to the corresponding θ-parameters?

The answer is different in these two cases. In a $U(1)$ theory there are no non-perturbative effects, and $F\tilde{F}$ is a total derivative of a gauge invariant current. This can easily be checked explicitly in Eq. (4.10). In $SU(2)$ the θ_{weak} term is potentially physical, but it can be rotated to zero using a global symmetry. This symmetry is baryon number (or lepton number). Baryon number has an anomaly with respect to $SU(2)$, and hence a baryon number phase rotation changes θ_{weak}. Furthermore, apart from the $SU(2)$ anomaly, baryon number is an exact symmetry of the Standard Model. It is instructive to compare this with the analogous phase rotation of θ_{strong}. Here the global transformation is the axial rotation $\psi \to \exp(i\alpha\gamma_5)\psi$. But this symmetry is broken not just by the anomaly, but also explicitly by the Yukawa coupling terms. This led to a link between the rotation of θ_{strong} and the phases in the quark masses. If we bring the latter in the physically preferred form (real masses) we cannot simultaneously bring θ_{strong} in the preferred form (zero).

It may well be that baryon number is broken not just by the anomaly with the weak interactions, but also explicitly. This happens in Grand Unified Theories (discussed in chapter 8) and many people expect gravity to break baryon number as well. But for now we have not seen protons decay. The anomaly-generated decay is too weak to observe anyway, but decay by other mechanisms may be observable in the future. If these have been observed, there will be additional Standard Model parameters we can measure, and some of those will be modified by baryon number rotations. If there is a natural canonical basis for these proton decay parameters, then we can in principle measure θ_{weak} with respect to that basis. But it will probably be simpler to define $\theta_{\text{weak}} = 0$, and define the proton decay parameters with respect to that choice. But before we can even discuss that, we need to find evidence of proton decay, and then measure CP-violating phases in such processes. This is not going to happen anytime soon.

A related problem is that of the QCD axion coupling to other gauge groups. These could be $SU(2)_{\text{weak}}$ or $U(1)_Y$. We have already seen such couplings. They give rise to the two-photon coupling of the axion, and also to $a \to W^+W^{-1}$ or $a \to ZZ$. But will these couplings affect the PQ-mechanism or the axion mass? The answer is no: precisely for the reasons discussed above, there is no $U(1)$ generated contribution to the axion potential, and there is no $SU(2)_{\text{weak}}$ generated one either (except perhaps a tiny contribution if baryon number is broken by new physics).

But there might exist additional non-abelian interactions that we have not seen yet or cannot see at all. There can be an entire strong interaction sector acting only on Dark matter. There may exist non-perturbative effects that have no corresponding gauge group. Such phenomena have been found in string theory, and go by the name "exotic instantons". Indeed, such contributions may be needed to give mass to the large number of axions these theories sometimes produce, since exactly massless axions point to an inconsistency (see the discussion of global symmetries above). If there are multiple non-perturbative contributions to the axion potential, it will look like this

$$V(a_i) = \sum_\alpha \mathcal{F}_\alpha \left[1 - \cos\left(\sum_i \frac{a_i}{f_{\alpha i}} + \bar{\theta}_\alpha\right)\right]$$

with $i = 1, \ldots, N$ and $\alpha = 1, \ldots, M$. The equations of motion are

$$0 = \frac{\partial V}{\partial a_i} = \sum_\alpha \frac{\mathcal{F}_\alpha}{f_{\alpha i}} \sin\left(\sum_i \frac{a_i}{f_{\alpha i}} + \bar{\theta}_\alpha\right) \equiv \sum_\alpha \frac{\mathcal{F}_\alpha}{f_{\alpha i}} S_\alpha$$

If $N \geq M$ this generically implies $S_\alpha = 0$. Then all sines vanish, and all their arguments must be zero. This is true if the matrix $\mathcal{F}_\alpha / f_{\alpha i}$ is non-degenerate. For example, if there is a single axion and a single group, it still will reduce $\bar{\theta}$ to zero if it does not couple to $G\tilde{G}$ (*i.e.* $f_{11} \to \infty$).

If we ignore degeneracies, roughly the following will happen. Let us assume all $f_{\alpha i}$ are of the same order of magnitude (but not all equal), so that all the scale dependence comes from the $\mathcal{F}_\alpha$. This is the case if the $f_{\alpha i}$ are generated by some fundamental theory at some high scale M_X (for example the GUT scale or the Planck scale), and if the $\mathcal{F}_\alpha$ are generated by strong interaction dynamics. Strong interaction dynamics has scales of order $\exp(-1/g^2)M_X$, where g is a dimensionless number of order 1. An example of a strong dynamics scale is the QCD scale. We expect on the basis of the dependence on g that the scales $\mathcal{F}_\alpha$ can be distributed roughly logarithmically over a large range (*i.e.* roughly the same number of distinct $\mathcal{F}_\alpha$ values per decade of energy scale). Let us assume for concreteness that $\mathcal{F}_\alpha$ takes values $\mathcal{F}_n = e^{-xn} M_X$, $n = 1, \ldots, M$. Furthermore we define $f_{\alpha i} = 1/\beta_{\alpha i} M_X$. If the dimensionless number x is sufficiently large, $\mathcal{F}_{n+1}$ can be ignored in comparison to $\mathcal{F}_n$.

So start with $\mathcal{F}_1$. We get N equations for $\mathcal{F}_1$, which are all of the form

$$\beta_{1i} S_1 = -\beta_{2i} e^{-x} S_2 - \beta_{3i} e^{-2x} S_3 - \ldots$$

Without further information, we should conclude that all sines $S_2, \ldots S_3$ are of order 1. This clearly implies that S_1 is of order e^{-x}. But that would be true if there is just one axion. With N axions, we can make linear combinations of the equations, and eliminate S_2, S_3 etc, all the way to S_N. Hence if there are N axions, the first non-vanishing term on the right-hand side is

$$-\beta_{(N+1)i} e^{-Nx} S_{N+1}$$

and we conclude that S_1 is of order e^{-Nx}.

Having solved the equations for S_1 we now turn to S_2. The equations for S_2 read

$$\beta_{2i} S_2 = -\beta_{1i} e^{x} S_1 - \beta_{3i} e^{-x} S_3 - \ldots$$

By linear combinations we can eliminate the S_1 term and $N-1$ of the subsequent terms, or we can eliminate N of the subsequent terms and leave the S_1 term. In either case, the conclusion is the same: S_2 is of order $e^{-(N-1)x}$. We can continue doing this, and conclude that S_k is of order $e^{-(N-k+1)x}$. Hence the last S_k that is still reduced somewhat is S_N, but S_{N+1} is of order 1.

For the QCD axion this implies that the sine is reduced by a factor e^{-mx}, where m is the number of light axions. We need $e^{-mx} \approx 10^{-10}$. This can be achieved with one axion, and a next scale ten orders of magnitude below the QCD scale, or with ten axions, with strong interaction mass scales differing by a factor 10; of course there are many other possibilities. In either case the lowest value of $\mathcal{F}_\alpha$ is about $10^{-10}\Lambda_{\rm QCD}$. Anything less than that is of course also fine. An important point is that there can be any number of mass scales in between this lowest relevant scale and the QCD scale, as long as there are sufficiently many light axions available. This is the same conclusion we reached above when we introduced explicit axion masses.

This is arguably the most plausible realization of the PQ mechanism. One needs a theory producing a large number of axions, and a large number of interaction scales, so that all of theses axion acquire a dynamical mass. If these scales are distributed on the entire energy scale (for example between the Hubble scale of 10^{-42} GeV and the Planck scale of 10^{19} GeV) there will be a substantial number of light axions below the QCD scale. That is all that is needed. This is known as the *axiverse* [1], and it may be realized in string theory.

6 Loop Corrections of the Standard Model

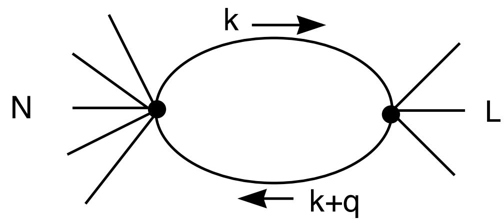
In this chapter we consider quantum corrections to the Standard Model due to loops of virtual particles. These corrections seem at first sight to be infinite, but on closer inspection these infinities can all be absorbed in the definition of the parameters. However, there is a remnant. It turns out that the value all parameters redefined in this way ("renormalized") now depends on the energy scale at which they are measured. This dependence gives us important information about the high-energy behavior of the Standard Model.

6.1 Divergences and Renormalization

The general idea of renormalization can be understood by considering scalar field theories. This has the advantage that many technicalities having to do with spin and gauge invariance can be left out of the discussion. These technicalities are not irrelevant. For example, gauge invariance plays an essential rôle in the renormalization theory of spin-1 fields. But for the aspects of interest here they are less important.

6.1.1 Ultraviolet Divergences

Consider a simple loop diagram in a scalar theory, such as



To be rather general we have left the number of external lines as a free parameter, and we have used an N+2-point vertex with coupling constant λ_{N+2} and an L+2-point vertex with coupling constant λ_{L+2}. The interaction Lagrangians are thus $\frac{1}{(N+2)!}\lambda_{N+2}\phi^{N+2}$ and the same with L instead of N. The loop integral is

$$(i\lambda_{N+2})(i\lambda_{L+2})\int \frac{d^4k}{(2\pi)^4}\left[\frac{i}{k^2-m^2}\right]\left[\frac{i}{(k+q)^2-m^2}\right] .$$

Here the integral is over all of momentum space.

For large k this integral behaves as (see section 6.2.1 for more details)

$$V = \int d^4k \frac{1}{k^2(q+k)^2} .$$

This integral diverges for the same reason why $\int dx(1/x)$ diverges. One calls this a logarithmic divergence. Since it occurs for large momenta it is also called an *ultraviolet divergence.*

6.1.2 Regularization

One can make the ultraviolet divergence explicit by cutting off the integral. Instead of integrating over all of momentum space, one integrates over a finite sphere of radius Λ^2, so that $k^2 < \Lambda^2$.* After introducing the cutoff the integral is finite, but now it depends on the cutoff,

$$V \propto \lambda_{N+2}\lambda_{L+2}\log(\frac{\Lambda^2}{q^2})$$

and we cannot take the cutoff to infinity.

The process of making the integral finite is called *regularization.* There are other ways of achieving this, and since it has no obvious physical meaning, all physical quantities one finally obtains should be independent of the regularization procedure. But first we have to get rid of the divergences.

* Since we are in Minkowski space this requires a bit more discussion, since it is not obvious what a "sphere" is. In fact all these manipulations are always done after one has analytically continued the integrand to Euclidean space using a Wick rotation.

6.1.3 The Origin of Ultraviolet Divergences

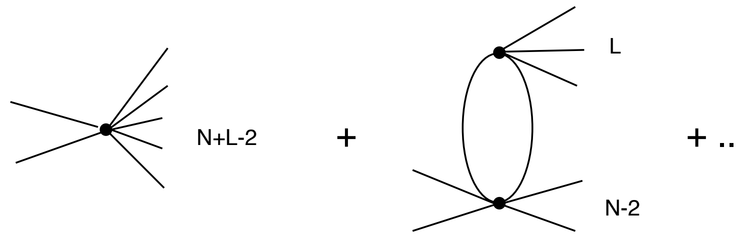
What is the reason for the infinity? Note that when we integrate over all of momentum space we are doing something that is physically ridiculous. Large momentum corresponds to large energies, and to short distances. Experimentally we have been able to explore nature up to several hundred GeV, and without doing further experiments we cannot pretend to know what happens at larger energies or shorter distances. Suppose that at shorter distances space-time has a crystalline structure. Then the inverse of the cell size would provide a maximum momentum, since wavelengths smaller than the cell size make no sense. In this situation the momentum cut off introduced above *would* have a physical meaning.

One may also envisage changes to the vertices that are small at low energies, but cut off the integral at large energies. For example, suppose the Feynman rule for a vertex was not λ_{L+2} but something like $\lambda_{L+2}\ [\Lambda^2/(P^2+\Lambda^2)]$, where P is the sum of the incoming momenta and Λ a large mass (larger than 1 TeV, say). A low energy observer would experimentally detect the existence of the $\lambda_{L+2}\phi^{L+2}$ vertex by scattering two ϕ particles, and measuring the probability that L such scalars come out. At low energies $P^2 \ll \Lambda^2$, and the correction factor is almost one. If Λ is large enough, it would be impossible to observe it. However, if we insert the same vertex in a loop diagram we integrate over all momenta, and we are sensitive to any such factor. Factors of this kind do indeed occur, for example if our ϕ particle were not elementary, but is in fact a bound state of two other particles. Then the interaction vertices are corrected by "form-factors". If the binding scale is sufficiently high, a low energy observer cannot resolve the sub-structure, and for all practical purposes sees the particle as elementary.

In other words, if we claim that Feynman diagrams are divergent for large momenta, we are simply making a completely unfounded extrapolation of known physics to extremely short distances. But that leaves us with the question what to do about these integrals.

6.1.4 Renormalization

Let us ask the question from the perspective of an experimentalist. Clearly the loop diagram contributes to processes with $N+L$ external lines. Suppose our theory has an additional vertex $\frac{1}{(N+L)!}\lambda_{N+L}\phi^{N+L}$. Suppose we do a scattering experiment to measure this vertex for example 2 ϕ particles to $N+L-2$ such particles. The amplitude, expanded to one-loop level has now schematically the following contributions



This is just intended schematically, and in particular we did not draw all diagrams here; there are others with one or both incoming lines attached to the other vertex. An experimentalist can only measure the sum of these diagrams.* The sum gives an expression like

$$\lambda_{L+N} + C\lambda_{N+2}\lambda_{L+2}\log(\frac{\Lambda^2}{q^2}) + \dots \ ,$$

where C is some numerical coefficient and q is some combination of the external momenta. The explicit form of both follows from the details of the computation, but is not relevant for our purpose. The dots indicate terms that are finite for $\Lambda \to \infty$ plus contributions of higher order diagrams.

The coupling constant λ_{L+N} is a physical parameter of the theory, that is not predicted by the theory itself, but must be measured. To measure it we must specify a physical process. In the present case, that physical process could be ϕ-ϕ scattering to $N + L - 2$ ϕ-particles with precisely specified external momenta. Let us call the value of q for those fixed momenta q_0. Then the physical value of the coupling constant is related in the following way to the parameters in the Lagrangian

$$\lambda^{\text{physical}}_{L+N} = \lambda_{L+N} + C\lambda_{N+2}\lambda_{L+2}\log(\frac{\Lambda^2}{q_0^2}) + \dots \ .$$

Experimentalists can only measure finite numbers, so clearly $\lambda^{\text{physical}}_{L+N}$ is finite, and independent of Λ. The terms in the linear combination appearing on the right hand side are physically irrelevant, because we can never measure them separately. If we now re-express the physical process for arbitrary momenta in terms of the physical, measured coupling constant, we get of course

$$\lambda^{\text{physical}}_{L+N} + C\lambda_{N+2}\lambda_{L+2}\log(\frac{q_0^2}{q^2}) + \dots \ ,$$

which is finite.

* In fact only the infinite sum of all diagrams is a measurable quantity. Here we work to second order in the coupling constants λ_K, which are assumed to be small. This may seem strange since the loop correction diverges as $\Lambda \to \infty$. But note that for any *finite* choice of Λ we can make the coupling constants small enough so that the next order can be ignored. After computing a physical cross section for small coupling, we continue the coupling to its physical value.

This process of absorbing short distance singularities into physical quantities is called *renormalization*.* The quantity $\lambda_{L+N}^{\text{physical}}$ is usually called the renormalized coupling constant, and the quantities that appeared in the Lagrangian are called *bare* coupling constant. They cannot be measured.

The crucial point is now the following. We can only give one definition of $\lambda_{L+N}^{\text{physical}}$, but of course this coupling constant appears in many different processes. Whenever λ_{L+N} (the bare coupling) appears, we replace it by $\lambda_{L+N}^{\text{physical}}$, using Eq. (6.1.4). If all goes well, this should remove all $\log \Lambda$ terms at the next order. For this to work, it should be true that λ_{L+N} *always* receives at the next order exactly the same loop corrections. To some extent one can see that intuitively, but actually proving it is quite hard.

The foregoing can be summarized by the following prescription:

1. Calculate some process to a given loop order in perturbation theory.

2. Introduce a prescription to cut off all the "divergent" integrals. (regularization).

3. For each physical parameter, choose one specific physical process to define and measure it.

4. Then use this definition in all other processes to substitute the bare parameters by the renormalized ones.

If all goes well, one now obtains for each process one computes a perturbative expansion in terms of physical, renormalized parameters, and all dependence on the regulator scale Λ has disappeared.

Note that it does not matter whether the momentum integrals are actually infinite or are cut off by some unknown short distance physics. All the unknown physics is absorbed in the renormalized parameters. These parameters depend on unknown physics and are therefore not determined theoretically.

However, in general the number of parameters one needs in this procedure is infinite. We can only absorb a $\log \Lambda$ in a physical parameter if that parameter actually exists. For a scalar theory the procedure outlined above will generate a vertex with $N + L$ lines if there exists a vertex with $N + 2$ and one with $L + 2$ lines. Suppose $N = L = 3$, *i.e.* we consider two five point vertices. Then $N + L = 6$, and to absorb the corresponding divergence we need a six-point vertex. Combining that with a five-point vertex gives a seven-point vertex, and clearly this never stops. Then the theory has an infinite number of parameters. To determine it *completely* one needs to do an *infinite* number of experiments.

6.1.5 Renormalizability

A theory is called *renormalizable* if all divergences can be absorbed into a *finite* number of parameters. This is a very strong restriction, but it makes the theory enormously

* The physical quantities meant here are all parameters in the Lagrangian, *i.e.* masses and coupling constants In addition to physical quantities, some singularities are absorbed in the normalization of the fields ("wave function renormalization"), which is not a physical quantity. For simplicity, we assume here that all divergences are absorbed in a single coupling constant.

more powerful. After the determination of a handful of physical parameters, one can make detailed predictions of all physical cross sections and decay rates! In our scalar theory example this allows only two vertices, ϕ^3 and ϕ^4. If there is just one scalar, the only parameters are the couplings λ_3 and λ_4 and the mass of the scalar. The mass is treated in a quite similar way: it also must be determined experimentally, and it is also renormalized.

Other examples of renormalizable theories are QED and QCD. Both have just one parameter, the coupling constants e and g respectively (if one ignores the fermion masses). The coupling constant of QED can be defined by means of the electron-photon coupling at zero photon momentum. For the QCD coupling constant the equivalent procedure cannot work, because we do not have free quarks and gluons, and furthermore because QCD perturbation theory does not work at zero gluon momentum. So one necessarily has to define g rather more indirectly, and at a non-vanishing momentum scale.

6.1.6 Dimensional Analysis

An important constraint on the allowed vertices comes from dimensional analysis. Since we have set $\hbar = 1$ and $c = 1$ there is just one physical dimension left, that of a mass. The number of powers of "mass" a physical quantity contains is called its *dimension*. Hence mass has dimension 1 and length has dimension -1; derivatives then have dimension 1.

Actions are dimensionless, and therefore Lagrangian densities must have dimension 4. From the kinetic terms we can then determine the dimensions of all fields. For example

$$\mathcal{L} = \tfrac{1}{2}\left[\partial_\mu \varphi \partial^\mu \varphi\right]$$

tells us that φ has dimension 1.* Similarly, fermion kinetic terms tell us that fermion fields have dimension $\frac{3}{2}$, and gauge kinetic terms require that gauge fields have dimension 1.

Now consider all other terms in the Lagrangian. They are all of the form

$$\text{(Coupling Constant)} \times \text{(Combination of fields and derivatives)} \tag{6.1}$$

Since the dimensions of the fields are fixed, dimensional analysis now fixes the dimensions of the coupling constants. Take for example $\lambda_N \varphi^N$. Obviously the dimension of λ_N is $4 - N$. If $N > 4$ the coupling constant has a negative dimension. This turns out to be the origin of non-renormalizability. Feynman diagrams with combinations of such coupling constants can have coefficients with arbitrarily negative dimensions, whose divergences correspond to terms with an arbitrarily large number of fields.

In the following we consider some theory with scalars, fermions and spin-1 fields. We denote these generically as φ, ψ and A. Without loss of generality, we may assume that ψ is left-handed. This eliminates the need for taking into account the matrix γ_5 in the discussion, because this matrix is just equal to 1 when acting on any left-handed spinor. This discussion is for general field theories, but we will comment on the most interesting

* Note that we could have allowed for a coefficient in front of the kinetic term, which could have its own dimension. However, we can always absorb such a coefficient by redefining φ. Any other term in the Lagrangian will always have a coefficient.

special case, the Standard Model. We order all terms in the Lagrangian according to the total dimension of the fields that appear in them. Of course all terms must respect the symmetries we require, which in any case includes Lorentz invariance and usually some local symmetries (gauge invariance). Note that the number of fermions in each term must always be even because of Lorentz invariance: an odd number of spinors can never produce a Lorentz singlet. Hence all terms in the Lagrangian have integer dimension.

The possible terms of dimensions 1,2 and 3 are

Dimension 1

- φ

Here we already used Lorentz invariance to eliminate A_μ. In the standard model a linear term in φ is not allowed because of $SU(2) \times U(1)$ invariance. In general such terms can be eliminated in the following way. If φ is allowed by all symmetries, so is φ^N, so we have no good argument to assume their absence. Then one can always shift φ by a suitable constant to eliminate the linear term.

Dimension 2

- φ^2
- A^2
- $\partial_\mu A^\mu$

Terms like A^2 can be made Lorentz invariant, but they are always forbidden by gauge invariance: they give mass to the gauge boson. Gauge invariance is required in order to make sense of the physical degrees of freedom of a vector boson. Hence there is no escape: A^2 terms are not allowed, unless they are generated by a Higgs mechanism. The term $\partial_\mu A^\mu$ is a total derivative, and hence not relevant. Hence the only option is φ^2, a scalar mass term. It has a coupling constant of dimension 2, and does indeed occur in the Standard Model.

Note that we will not have to discuss vector fields A at all, because as long as gauge symmetry is exact, and not spontaneously broken, all occurrences of A are via covariant derivatives $\partial_\mu + A_\mu$. Hence we only have to discuss derivatives, which must be contracted in a Lorentz invariant way. Hence they must be contracted together or with Dirac matrices γ^μ.

Dimension 3

- ψ^2
- φ^3
- $\partial^2\varphi$

The last term is always a total derivative, so we will never have such terms. The first two terms are allowed in principle, and have coupling constants of dimension 1. But in the Standard Model no such terms can occur, because they would violate gauge invariance.

Dimension 4

- φ^4
- $\bar{\psi}\psi\varphi$

Other allowed terms are of the form $\partial^2\varphi^2$ and $\partial\psi^2$, but when we make them Lorentz invariant they just yield the kinetic terms. The other two terms do indeed exist in the Standard Model, and they have dimensionless coupling constants.

If we continue to higher dimensional operators, we inevitably get coupling constants with negative mass dimension. These destroy renormalizability: if we allow one of them, we will have to allow arbitrarily many field combinations of arbitrary higher dimension. Symmetries may restrict that set, but there will always be infinitely many. Hence the definition of the Standard Model is that all such terms are absent. The scales they correspond to are sent to infinity, and hence their coupling constant is sent to zero. This, in combination we a choice of gauge symmetries and fermion and scalar representations, defines the Standard Model completely. We may well find experimental evidence for the existence of operators with dimension 5 or higher, but this implies physics beyond the Standard Model.

Nevertheless, we can already look ahead and construct the operators of higher dimension that are allowed to exist by Lorentz invariance and gauge invariance. It turns out that there is just one of dimension 5 (in the absence of right-handed neutrinos, and treating family indices with a coupling matrix, and not as separate operators), 63 of dimension 6 and 20 of dimension 7 [21]. The unique dimension-5 operator is the Weinberg operator for neutrino masses, Eqn. (5.17).

To understand the precise counting, let us do it for dimension 4. There are three distinct gauge kinetic terms; one scalar kinetic term; five fermion kinetic terms, a φ^4 interaction and three Yukawa terms, for a total of 13. The five fermion kinetic terms are those of the fields of type $\mathcal{Q}, \mathcal{U}, \mathcal{D}, \mathcal{L}$ and $\mathcal{E}$. Just as the three Yukawas, these have flavor indices, but we do not take these into account in the counting, because the group structure of all such terms is the same. Note that terms of the form $F\tilde{F}$ have not been taken into account in [21].

6.1.7 The Meaning of Renormalizability

There is another way of looking at the requirement of renormalizability. Mass scales in physics usually have a deeper meaning. If a coupling constant has a non-zero dimension, the corresponding mass scale must have a physical interpretation in terms of "new physics". Take for example a fermion four-point vertex $(\bar{\psi}\psi)^2$. This has dimension six, so the coupling constant has dimension -2. Fermi wrote down an interaction vertex of this type to understand the weak interactions, and this gives a very accurate description of weak interactions at low energies. However, now we know a more fundamental explanation for this four-fermi interaction. In the Standard Model, the interaction is attributed to exchange of a heavy W-boson, and takes the form*

$$\bar{\psi}\gamma^{\mu}\psi\frac{1}{k^2 - M_W^2}\bar{\psi}\gamma_{\mu}\psi \ .$$

For low values of k this looks like a four-fermi vertex with a coupling constant $\frac{1}{M_W^2}$, but at higher energies the effect of the propagator momentum k_μ becomes noticeable.

The four-fermion theory is not renormalizable. If one imagines a time before the weak interactions were discovered, but QED was known, then the physicists of that time could live happily with the knowledge that their theory was renormalizable. The discovery of the four-fermion interaction changed that. Its presence hints at new physics. That physics is described by the Standard Model, which again looks to us as a renormalizable theory. But future experiments may change that again. If evidence for new interactions with negative dimensional coupling constants is found, we may again expect new physics. Nobody knows where that may happen, but the point is that it does not matter as long as the scale of the new physics is large enough. Then the extra interactions are anyway invisible to us. In other words, renormalizability is not some deep property of nature, but rather an inevitable consequence of doing physics well below the next scale where interesting new phenomena occur.

In our description of nature both renormalizable and non-renormalizable theories play a role. For example the Standard Model of weak, electromagnetic and strong interactions is renormalizable, but the theory of pion-nucleon interactions is not. In the former case that means that we can predict scattering amplitudes of quarks and leptons with – in principle – unlimited accuracy in terms of only a few (about 27) parameters that must be determined experimentally. In the latter case we may be able to describe low-energy pion-nucleon interactions, but if we attempt to go to higher energies more and more parameters are required and finally the description becomes completely inadequate. At sufficiently short distances we have to take the quark substructure of pions and nucleons into account, and we cannot pretend that they are fundamental fields.

At some time in the future we may find ourselves in the same situation with the Standard Model, but only experiment can tell us if and when that happens.

6.2 Running Coupling Constants

Parameters in a Lagrangian are not physical quantities. The latter can only be defined by specifying some procedure for measuring them. This is also true for "coupling constants".

In QED there is a natural way of defining the coupling constant: the photon-electron coupling at zero photon momentum. This is a well-defined physical process, and measurement produces a value for coupling constant, $\alpha = \frac{g^2}{4\pi} \approx \frac{1}{137.04}$. When one tries to do the

* This is only schematic, and important details such as the left-handed nature of the currents are suppressed.

same in QCD one realizes that QCD at zero momentum is extremely non-perturbative. Instead one can then try to define the QCD coupling constant in processes at higher momentum, for example deep inelastic scattering, but then it becomes quickly clear that one gets a different answer if the momentum scale is changed.

This has to do with the logarithmic corrections that always accompany the infinities of quantum field theory. In the present case the relevant infinities are the ones that are removed by counter-terms that are absorbed by redefining the bare coupling constant in the Lagrangian. For example in ϕ^4 theory the first two contributions to the four point function are shown below in figs. 2 and 3.

Typically the computation of such a loop correction to one of the vertices yields an expression of the form*

$$V(Q) = g_{\rm bare} - g^n_{\rm bare} b_0 \log(\frac{\Lambda}{Q}) \ , \tag{6.2}$$

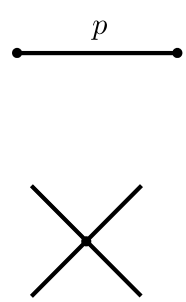
where Q is some invariant built out of the external momenta and Λ is the cut-off (cut-off regularization is not a good procedure for gauge theories, but is most suitable for explaining the main point); $g_{\rm bare}$ is the coupling constant appearing in the Lagrangian. In the example of ϕ^4 $n = 2$ and $g_{\rm bare} = \lambda$; for gauge theories $n = 3$. In more complicated situations (Yukawa couplings for example) the one-loop corrections involve more than one coupling constant. Here and in the following "log" denotes the logarithm on base e.

6.2.1 Example: Scalar Field Theories

Let us consider the simplest interacting scalar field theory, with Lagrangian

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \tfrac{1}{2}\mu^2\phi^2 - \tfrac{1}{24}\lambda\phi^4 \ . \tag{6.3}$$

This Lagrangian leads to the Feynman rules shown below. Note that the normalization of the ϕ^4 vertex is chosen in such a way that the vertex has a factor 1. The first loop



$$-\frac{i}{p^2 - m^2}$$

$$-i\lambda$$

Figure 2: Feynman rules for ϕ^4 theory

correction to the vertex comes from the diagram shown below; here $q = p_1 + p_2 = p_3 + p_4$ and all momenta flow from left to right. Actually, there are two more diagrams, distin-

* Here only the renormalization of the coupling constants is considered; for simplicity, wave-function and mass renormalizations are ignored.

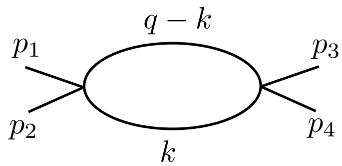


Figure 3: One of the three scalar one loop graphs.

guished only by connecting p_1 and p_3 (and p_2, p_4) to the same vertex and by connecting p_1 and p_4 (and p_2, p_3) to the same vertex. The computation of these other two diagrams is completely analogous to the one shown here, and the net result will just be a factor of three in the logarithm. The Feynman integral is

$$\frac{1}{2}(-i\lambda)^2 \int \frac{d^4k}{(2\pi)^4} \left(\frac{-i}{k^2-m^2}\right)\left(\frac{-i}{(k-q)^2-m^2}\right) \tag{6.4}$$

The factor $\frac{1}{2}$ is a symmetry factor, needed because there are two identical propagators between the two vertices. To evaluate the integral we use Feynman's trick

$$\frac{1}{AB} = \int_0^1 dx \, \frac{1}{(xA+(1-x)B)^2} \ , \tag{6.5}$$

where A and B are the two propagator denominators. We make a change of variable in the integration, defining $l^\mu = k^\mu - xq^\mu$. Then we get

$$\frac{1}{2}\frac{\lambda^2}{(2\pi)^4}\int d^4l \frac{1}{(l^2+x(1-x)q^2-m^2)^2} \tag{6.6}$$

Now we make a Wick rotation to Euclidean space, by defining $l_0 = il_4$. Doing this properly requires a definition of the location of the poles of the propagators in the complex plane, the "$i\epsilon$-prescription", which we will not discuss here in detail. One defines the propagators as

$$\lim_{\epsilon\to 0} \frac{1}{p^2-m^2+i\epsilon} \tag{6.7}$$

Then the integration contour is rotated to the imaginary axis in such a way that the poles are avoided. The net result is that the integral now becomes

$$\frac{1}{2}\frac{\lambda^2}{(2\pi)^4}\int_{-\infty}^{\infty} dl_4 \int d^3l \frac{1}{\left(-l_4^2-\vec{l}^2+x(1-x)q^2-m^2\right)^2} \tag{6.8}$$

We define a variable $Q^2 = -q^2$ in order to avoid branch cuts that exist for $q^2 > 0$. The integral over l can be written in terms of four-dimensional Euclidean polar angles

$$\int_{-\infty}^{\infty} dl_4 \int d^3l = \int d^4l = \int_0^\infty l^3 dl \int d\Omega_3 = 2\pi^2 \int_0^\infty l^3 dl \tag{6.9}$$

The integral over l can be done using

$$\int_0^\infty dl \frac{l^3}{(l^2+a)^2} = \frac{1}{2}\left[\log(l^2+a) + \frac{a}{l^2+a}\right]\Bigg|_0^\infty \tag{6.10}$$

Clearly, this integral is divergent. One can deal with this by introducing a cut-off Λ. By "cut-off" we mean the highest allowed loop-momentum. If we believe that physics remains unchanged up to arbitrary high momenta (*i.e.* arbitrarily short distances), Λ would be infinite, but this is clearly a preposterous assumption: we cannot possibly know this. If there are fundamental changes in the theory at short distances (as is the case in string theory or if space-time is discrete, just to mention a few possibilities) the momentum integral may be finite, and then Λ is simply a very large momentum scale. At this point we treat Λ as if it were just an additional parameter, representing our ignorance about physics at very short distances. This looks worrisome. We are just computing a single one-loop graph, and immediately we encounter a new parameter. This will happen in many other graphs, and hence if we go on to arbitrary loop order, we would encounter an infinite number of parameters, which have to be determined experimentally in order to use them. Fortunately, it turns out that these parameters are not all separately observable. Only certain combinations can be observed, and in the Standard Model these combinations correspond precisely to the parameters in the Lagrangian.

After replacing ∞ by Λ we get

$$\frac{1}{2}\frac{\lambda^2}{16\pi^2}\int_0^1 dx \left[\log\left(\frac{\Lambda^2 + x(1-x)Q^2 + m^2}{x(1-x)Q^2+m^2}\right) - 1\right] \tag{6.11}$$

Now we consider the limit $Q^2 \ll \Lambda^2$ and $Q^2 \gg m^2$. The former limit is simply the assumption that the point where momenta are cut-off is well beyond the energy scale of physical interest. If that is not the case, then surely we must know further details about how they are cut off. The other limit assumes that we are considering energy scales much larger than the particle masses. This is a good assumption in LHC physics, but of course this is not always true. If a mass is larger than Q, we may ignore the Q^2 dependence in the argument of the logarithm. We will return to that case later. Consider now the limiting case and add the result of this diagram to the tree amplitude. Then we get

$$-i\left(\lambda - \frac{\lambda^2}{32\pi^2}\log\left(\frac{\Lambda^2}{Q^2}\right) + \ldots\right) \tag{6.12}$$

where the dots represent finite terms. The other two diagrams give the same Λ dependence, but a different dependence on external momenta. We can define $Q_s^2 = -(p_1+p_2)^2$, $Q_t^2 = -(p_1-p_3)^2$ and $Q_u^2 = -(p_1-p_4)^2$, and then the sum can be written as

$$-i\left(\lambda - \frac{3\lambda^2}{32\pi^2}\log\left(\frac{\Lambda^2}{{Q_0}^2}\right) - \frac{\lambda^2}{32\pi^2}\left[\log\left(\frac{Q_0^2}{Q_s^2}\right) + \log\left(\frac{Q_0^2}{Q_t^2}\right) + \log\left(\frac{Q_0^2}{Q_u^2}\right)\right] + \ldots\right) \tag{6.13}$$

Here Q_0 is some common energy scale, chosen in such a way that the momentum dependent finite terms in the square brackets are small, so that they can be considered as part of the finite terms. This example illustrates how expressions like Eq. (6.2) come out of actual loop computations. Note that the coefficient $b_0 = 3/16\pi^2$ in this case.

6.2.2 The Renormalization Group Equation

If a theory is renormalizable all occurrences of the large scale Λ can be absorbed in the definition of a finite number of parameters (coupling constants, masses and field normalizations). In the case under consideration that works as follows. One can define $g_{\rm bare}$ so that in the limit $\Lambda \to \infty$ the physical coupling constant is finite, namely by defining

$$g_{\rm bare} = g_{\rm phys}(\mu) - g^n_{\rm phys} b_0 \log(\frac{\mu}{\Lambda}) \ . \tag{6.14}$$

This cancels all "infinities" up to order n, but it is clear that the definition of the coupling constant is now dependent on some arbitrary scale μ. Furthermore a finite term $b_0 \log(\mu/Q)$ remains. [Note that we work here in an expansion in the coupling constant, and that higher order terms are ignored. Thus in particular $g^n_{\rm bare} \approx g^n_{\rm phys}$ since the corrections are of higher order in g, even though their coefficients may involve $\log \Lambda$. They should be taken care of at the next order in g.]

Another way of saying this is to define the coupling constant as the value of $V(Q)$ (here we assume for the sake of the argument that $V(Q)$ is a directly measurable quantity which at tree level is equal to the coupling constant). Since no experiment can directly measure the coefficients in the Lagrangian this is the only thing we can do. It follows immediately that $V(Q)$ cannot be a constant. At best we can choose a reference scale $\mu = Q$ to define and measure it, and then calculate its value at any other scale. At present the most commonly used reference scale for the Standard Model couplings is $M_{\rm W}$. Note that the coupling constant increases with increasing Q if b_0 is positive.

One of the consequences of renormalizability is that the same redefinition removes the infinities associated with the coupling constant in all diagrams. This implies that in the finite result the same logarithmic corrections $-b_0 \log(\mu/Q)$ will always appear with any coupling constant, albeit with process dependent quantities Q.

If we measure $g(\mu)$ in one process we can now make predictions for all others, but what should we take for μ? The best choice would seem to be the one that minimizes the logarithmic corrections, *i.e.* $\mu = Q$. If we take μ very different from Q the convergence of the loop expansion becomes very bad, since at each order in g one encounters the large logarithmic correction $\log(\mu/Q)$ to the same power. By setting $\mu = Q$ we are effectively summing up these large logarithms. Consequently each process now has its own coupling constant $g(Q)$, and the coupling "constant" is not a constant anymore, but a function of the scale. This is called the *running coupling constant*.

Technically this is done by means of the renormalization group equation. We will show here how this equation is derived in the present, slightly simplified context. Consider the measurable quantity $V(Q)$ introduced in Eq. (6.2), and substituting for $g_{\rm bare}$ the physical coupling constant Eq. (6.14).

$$V(Q) = g_{\rm phys}(\mu) - g^n_{\rm phys}(\mu) b_0 \log(\frac{\mu}{Q}) + \text{higher order} \tag{6.15}$$

Now it seems that the physical quantity $V(Q)$ depends on μ, the energy scale at which we have decided to define and measure the coupling constant. But this is just a convention, on which no physical quantity should depend. Hence it must be true that

$$\mu \frac{d}{d\mu} V(Q) = 0 \tag{6.16}$$

This leads immediately to the equation

$$\mu \frac{d}{d\mu} g_{\rm phys}(\mu) - b_0 g^n_{\rm phys}(\mu) = 0 \tag{6.17}$$

Here we have ignored the derivative of $g^n_{\rm phys}(\mu)$ because this is of higher order in the coupling constant. If we define $\beta(g) = b_0 g^n$ (plus terms of higher order), we may write this as

$$\mu \frac{d}{d\mu} g_{\rm phys}(\mu) = \beta(g_{\rm phys}(\mu)) \tag{6.18}$$

On the other hand, if we view $V(Q)$ as a function of $g_{\rm phys}$, μ and Q (with an explicit dependence on μ through the logarithm and an implicit dependence via $g_{\rm phys}$) we may write the derivative Eq. (6.16) in terms of partial derivatives as

$$0 = \mu \frac{d}{d\mu} V(Q) = \left[\mu \frac{\partial}{\partial \mu} + \mu \frac{d g_{\rm phys}(\mu)}{d\mu} \frac{\partial}{\partial g_{\rm phys}} \right] V(g_{\rm phys}, \mu, Q) \ , \tag{6.19}$$

or in terms of the function $\beta(g)$ we just introduced

$$\left[\mu \frac{\partial}{\partial \mu} + \beta(g_{\rm phys}) \frac{\partial}{\partial g_{\rm phys}} \right] V(g_{\rm phys}, \mu, Q) = 0 \ . \tag{6.20}$$

This derivation can in fact be done to any order in g, and for any Green's function. The function $\beta(g)$ (called "the β-function") now becomes a polynomial in g rather than just a single term we found in the one-loop case. The general answer for a Greens' function G is (omitting again for simplicity the effects of masses and external lines, and renaming $g_{\rm phys}$ simply g)

$$\left[\mu \frac{\partial}{\partial \mu} + \beta(g) \frac{\partial}{\partial g} \right] G(g, \mu, Q) = 0 \ , \tag{6.21}$$

where $\beta(g)$ is the β-function and g denotes the physical (renormalized) coupling constant at the scale μ. The statement that this holds to arbitrary order in g should not be misinterpreted. Of course both $\beta(g)$ and G have an expansion in powers of g with coefficients we do not know, except for the lowest orders. However, if we just introduce parameters for these coefficients, then Eq. (6.21) holds to any order. It is simply a consequence of the requirement that physics should not depend on an arbitrary choice of reference scale μ.

6.2.3 Summing Leading Logarithms

The renormalization group equation can formally be solved in the following way

$$G(g, \mu, Q) = G(\bar{g}(\log(Q/\mu), Q, Q) \ , \tag{6.22}$$

where the function $\bar{g}$ is the solution to the differential equation

$$\frac{d}{dt}\bar{g}(t) = \beta(\bar{g}(t)) \;, \tag{6.23}$$

subject to the boundary condition $\bar{g}(0) = g$ (where, as above, g means $g_{\rm phys}$), so that we get the correct answer for $Q = \mu$ (here $t = \log(Q/\mu)$). To show that this is correct, observe that on the one hand

$$\begin{aligned} \mu\frac{\partial}{\partial\mu}G(\bar{g}(t), Q, Q) &= G'(\bar{g}(t), Q, Q)\mu\frac{\partial}{\partial\mu}\bar{g}(t) \\ = -G'(\bar{g}(t), Q, Q)\frac{d}{dt}\bar{g}(t) &= -G'(\bar{g}(t), Q, Q)\beta(\bar{g}(t)) \end{aligned}$$

where G' is the partial derivative of G with respect to the first variable. On the other hand

$$\beta(g)\frac{\partial}{\partial_g}G(\bar{g}(t), Q, Q) = G'(\bar{g}(t), Q, Q)\beta(g)\frac{\partial\bar{g}(t)}{\partial g}$$

Now we use the relation

$$\beta(\bar{g}(t)) = \beta(g)\frac{\partial\bar{g}(t)}{\partial g} \tag{6.24}$$

to show that Eqn (6.21) is indeed satisfied. To show that Eqn. (6.24) holds, define

$$F(t) = \beta(\bar{g}(t)) - \beta(g)\frac{\partial\bar{g}(t)}{\partial g} \;.$$

Using Eqn (6.23) it is easy to show that the derivative $\frac{\partial F}{\partial t}$ is proportional to $F(t)$:

$$\frac{\partial F}{\partial t} = X(t)F(t)$$

for some function $X(t)$. Then

$$\begin{aligned} \frac{\partial^2 F}{\partial^2 t} &= \frac{\partial}{\partial t}\left[X(t)F(t)\right] \\ &= \left(\frac{\partial}{\partial t}X(t)\right)F(t) + X(t)\frac{\partial F}{\partial t} \\ &= \left(\frac{\partial}{\partial t}X(t)\right)F(t) + X(t)^2F(t) = X'(t)F(t) \end{aligned}$$

This implies that *all* derivatives of $F(t)$ are proportional to $F(t)$. Furthermore due to the boundary condition we have $F(0) = 0$, and hence all derivatives vanish at $t = 0$. Then F must vanish for all t. Note that Eqn. (6.24) also implies that

$$\left[\mu\frac{\partial}{\partial\mu} + \beta(g)\frac{\partial}{\partial g}\right]\bar{g}(t)) = 0 \;, \tag{6.25}$$

which is a necessary condition for $\bar{g}(t)$ itself to be a physical quantity.

In the solution (6.22) all explicit dependence on μ via logarithms $\log(Q/\mu)$ is removed by setting $Q = \mu$. The entire dependence on both Q and μ is absorbed into the coupling "constant" (which is actually not a constant, but depends on Q; hence the somewhat contradictory name "running coupling constant").

At one loop the differential equation for the running coupling constant (6.23) can easily be solved, and the solution is

$$\bar{g}^{n-1}(t) = \frac{g^{n-1}}{(1-(n-1)b_0 t g^{n-1})} \tag{6.26}$$

If we expand this solution to order g^n we get precisely the one-loop contribution discussed above. However, even if we take for $\beta(g)$ just the one-loop expression $b_0 g^n$ we see that $\bar{g}$ contains an infinite number of terms. These correspond to the so-called "leading log" contributions to higher loop diagrams. Higher terms in $\beta(g)$ correspond to "next-to-leading logs", which are down by one or more powers of $\log(Q/\mu)$. This solution is valid only if g is small, since otherwise it is certainly not correct to ignore the higher order terms in the β function. If we extrapolate to higher energies ($t = \log(Q/\mu) \to \infty$) we observe that g^{n-1} becomes smaller and smaller if $b_0 < 0$. However, if $b_0 > 0$ the coupling constant increases until it becomes formally infinite for $t = 1/((n-1)b_0 g^{n-1})$ (we assume that $g > 0$). This is called a "Landau pole". Here perturbation theory breaks down, and hence one cannot conclude exactly what happens to the theory. Theories with $b_0 < 0$, which are well-behaved at higher energies, are called *asymptotically free*. This is a very desirable property since it makes it plausible that no new dynamics will appear at higher energy; in order words, if we understand the theory at low energies, we can be quite confident that it harbors no surprises when extrapolated to arbitrarily large energies. In practice, however, we still have to worry about interactions with other theories, most notably gravity, disturbing our extrapolations.

6.2.4 Asymptotic Freedom

To see which theories are asymptotically free we list here the values of b_0 for some popular theories. For non-abelian gauge theories coupled to Weyl fermions:

$$b_0 = -\frac{1}{96\pi^2}\left(11 I_2(A) - 2 I_2(R_f) - \frac{1}{2} I_2(R_s)\right) \ , \tag{6.27}$$

where $I_2(R)$ is defined as

$$\mathrm{Tr}_R T^a T^b = \frac{1}{2} I_2(R) \delta^{ab} \ , \tag{6.28}$$

for any representation R. The representations occurring in Eq. (6.27) are the adjoint A, the representation of the Weyl fermions R_f (left- and right handed fermions give the same contribution) and of the scalars R_s. The scalars are assumed to be real; if a scalar is complex, as it must be if it transforms in a complex representation, one gets an extra factor of 2. The term depending on A is due to the gauge bosons and the Fadeev-Popov ghost required by gauge fixing. Note that R_f and R_s may be reducible representations; the index I_2 is then simply the sum of the indices of each component. This formula is only correct for one choice of normalization of the generators, namely so that $I_2(A) = C_2(A)$, where C_2 denotes the quadratic Casimir eigenvalue. This is the normalization adopted throughout these lectures (see appendix B).

For the group $SU(N)$ one has $I_2(R) = 1$ in the fundamental representation, and $C_2(A) = 2N$. It follows that $SU(3)$ is asymptotically free if the number of Weyl fermions in the fundamental representation or its conjugate is less than 33. In the Standard Model there are 6 flavors of Dirac fermions, each with two Weyl components, so that QCD is asymptotically free. In $SU(2)$ one can accommodate 22 Weyl doublets, or 21 and one Higgs. The Standard Model has four per family, and hence a total of 12, so that $SU(2)$ is asymptotically free as well. The running coupling constant of these theories is given by Eq. (6.26) with $n = 3$.

The leading terms (*i.e.* the one loop contribution) of the β-functions in the Standard Model are then

$$\beta_3 = -\frac{42}{96\pi^2} g_3^3 + \dots \tag{6.29}$$

$$\beta_2 = -\frac{19}{96\pi^2} g_2^3 + \dots \tag{6.30}$$

These are the β-functions at high energies, where all $SU(3) \times SU(2) \times U(1)$ multiplets are participating. Since the heaviest Standard Model particle is the top quark, this means that these β functions are valid at energies above the top quark mass. In general, if one considers some energy scale Q, one should include in the β-function all particles with masses $m < Q$. The second term contains the contribution of the Higgs doublet ϕ in the representation $(1, 2, \frac{1}{2})$ Note that the Higgs field is complex, so that the $I_2(R_s)$ term in Eq. (6.27) must be multiplied by 2. Hence the Higgs field contributes $1/96\pi^2$, and the total contribution is $-44 + 2 \times 12 + 1 = -19$; the factor 12 in the second term is due to four Weyl doublets in each of the three families. For QCD the exact count is $(-11 \times 2 \times 3 + 2 \times 12) = -42$.

6.2.5 Abelian gauge theories

Now consider abelian gauge theories. There is no gauge boson and ghost contribution, and the matter contribution can be obtained directly from the non-abelian case, using $I_2(R) = 2\ \mathrm{Tr}T^2 = 2\sum q^2$. Now b_0 is always positive, and the coefficient is

$$b_0 = \frac{1}{96\pi^2}\left(4\sum q_f^2 + \sum q_s^2\right) , \tag{6.31}$$

where q_f and q_s are the Weyl fermion and real scalar charges. For QED coupled to a single Dirac electron we find thus $b_0 = 1/12\pi^2$.

The fine structure constant α increases from its low energy value of $1/137.04$ to a value of about $1/128$ at the weak scale Q_W. Beyond that we should evolve the coupling constants g_i of $SU(3)\times SU(2)\times U(1)$ rather than the QED coupling constant. The value of $96\pi^2 b_0$ for $U(1)_Y$ per family is

$$4\times\left[6\times\left(\tfrac{1}{6}\right)^2+3\times\left(\tfrac{1}{3}\right)^2+3\times\left(-\tfrac{2}{3}\right)^2+2\times\left(-\tfrac{1}{2}\right)^2+1\right]=\tfrac{40}{3} \tag{6.32}$$

The Higgs scalar contributes $\sum q_s^2 = 4\times(\frac{1}{2})^2=1$ (with a factor two for the dimension and another one for the complexity). This gives the following result for the g_1 β-function:

$$\beta_1=\frac{41}{96\pi^2}g_1^3\ +\dots \tag{6.33}$$

If we match it with the QED coupling at the weak scale, $Q_W\approx 100$ GeV, we get a boundary condition $g_1=e/\cos(\theta_w)$ at that scale. This yields

$$g_1(0)=\frac{\sqrt{4\pi\alpha(0)}}{\cos(\theta_w)}=0.357 \tag{6.34}$$

Here we are using Q_W as the reference scale μ, so that $t=\log(Q/Q_W)$, and $\alpha(0)=1/128$.

Now we can extrapolate g_1 to higher energies Q. It increases, until $g_1(t)$ reaches its Landau pole at for

$$t=\frac{1}{2b_0 g_1(0)^2}=\frac{48\pi^2}{41 g_1(0)^2}\approx 90.7 \tag{6.35}$$

i.e. $Q=e^{90.7}Q_W\approx 10^{39}Q_W\approx 10^{41}$ GeV. This is far beyond the Planck mass of $\approx 10^{19}$ GeV, and hence we are likely to encounter more serious difficulties before having to worry about it.

6.2.6 Yukawa Couplings

The one-loop β-function for a Yukawa coupling y is

$$\beta_y=\frac{1}{16\pi^2}\left[\frac{9}{2}y^3-(8g_3^2+\frac{9}{4}g_2^2+\frac{17}{12}g_1^2)y\right]+\dots \tag{6.36}$$

Here there is an interesting competition between the first term and the other three, of which the QCD contribution is the dominant one due to its larger coupling constant and numerical factor.

If y is small, the negative terms dominate, and the Yukawa couplings evolve to smaller values. If we ignore the running of g_3 the equation for y has the form $\frac{dy}{dt}=-Cy$, and the solution is a negative exponential. In this case there are no problems.

If y is large, the first term dominates and the coupling grows, and will become infinitely large. The border between these two cases is

$$y^2=\frac{64}{9}\pi\alpha_s \tag{6.37}$$

This can be converted to a quark or lepton mass of about 270 GeV. If y has exactly this value at M_{W}, the coupling constant initially does not move at all, but since α_s decreases the first term will eventually win, and y will increase. If one requires that M does not become infinite before the Planck scale, one gets an upper mass limit for quarks and leptons of about 200 GeV. The exact evolution of y, including the effect of running of the gauge coupling constants, is shown in fig. 4.

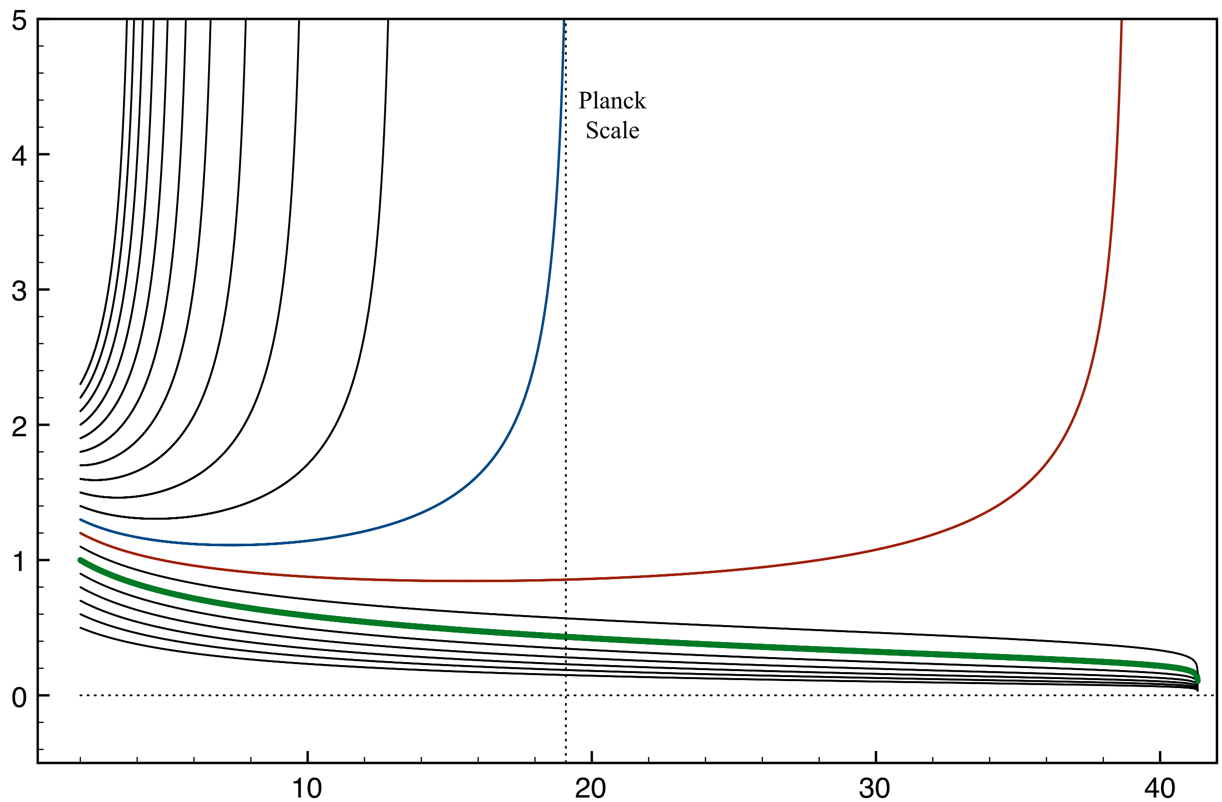


Figure 4: Running of the Yukawa coupling y for initial values $y = 0.5 + 0.1n$ (corresponding to a top-quark mass of $87 + 17.5n$ GeV). The horizontal axis is the energy in GeV, in powers of 10. The thick green curve represents the observed top quark mass, the red curve is for $m_t = 208$ GeV, and the blue one for $m_t = 226$ GeV. These are respectively the last one without a Landau pole below the Planck scale (in our discrete set) and the first one with a Landau pole below (actually almost exactly at) the Planck scale. The cutoff behavior above 10^{40} GeV is caused by the Landau pole of g_1.

The β function Eq. (6.36) is said to have an infrared fixed point. If we start at some high energy scale with a value y_0 and we allow the coupling constant to evolve to lower energies, it must end up at the fixed point value Eq. (6.37) (this assumes that higher order corrections to the β-function may be ignored). For the fermion masses this evolution to arbitrarily small energy scales is not really relevant though. They are determined by the value reached by y at the weak symmetry breaking scale.

In this situation choosing almost any value for y at, for example, the Planck scale yields a value of about 200 GeV for the fermion mass (if we blindly apply the one-loop β function even when the couplings are large), almost independent of the input value $y(M_{\rm Planck})$. Only for very small values of this parameter a significant reduction of the mass is found.

Interestingly, the top quark mass is about 175 GeV, just below the bound. This means that all Yukawa couplings decrease at shorter distances, and hence they do not cause any problems.

6.2.7 The Higgs Self-coupling

Now that the Higgs boson has been found and its mass determined, we also know the value of its self-coupling. But there has been a long period before 2012 when only the vacuum expectation was known, and the value of the Higgs mass was *completely* unknown. Indeed, expectations for that mass have varied between a few MeV and a TeV. In the 90's, precision experiments at LEP (an electron-positron collider at CERN) and other experiments started to constrain the range of masses. Furthermore, there were upper and lower bounds if one requires that the self-coupling remained finite and positive at higher energies. Even now that the Higgs mass is known, these bounds still have some relevance.

The Triviality Bound. The Higgs self-coupling is also not asymptotically free: $b_0 = 3/2\pi^2$ and $n = 2$. We can express the value of the Higgs self-coupling λ in terms of the Higgs boson mass ($\approx$ 125 GeV), the mass of the W boson and coupling constants, using the relations $v = 2\sqrt{-\mu^2/\lambda}$, $M_{\rm W} = \frac{1}{2}g_2 v$ and $M_{\rm H} = \sqrt{-2\mu^2}$ (here μ is the Higgs mass parameter, not the renormalization scale):

$$\lambda = \frac{M_{\rm H}^2 4\pi\alpha}{2M_{\rm W}^2 \sin^2\theta_{\rm w}} \approx .2\left(\frac{M_{\rm H}}{M_{\rm W}}\right)^2 \approx .51 \tag{6.38}$$

Here we defined all couplings, as before, at $Q_W \approx 100$ GeV. The Landau pole occurs at

$$t = \frac{1}{\lambda b_0} = \frac{2\pi^2}{3\lambda(0)} \tag{6.39}$$

This already tells us that the Higgs system becomes strongly coupled if $M_{\rm H} \gtrsim 3M_{\rm W} \approx 250$ GeV. The relevant scale in this computation is $M_{\rm W}$, and hence it is reasonable to assume that λ takes the value (6.38) at the scale $M_{\rm W}$. The scale Q_∞ at which the coupling formally becomes infinite is then given by

$$Q_\infty = M_{\rm W} e^{\frac{8\pi^2}{3\lambda(M_{\rm W})}} \approx M_{\rm W} e^{133.6(\frac{M_{\rm W}}{M_{\rm H}})^2} \ . \tag{6.40}$$

Beyond Q_∞ the Standard Model with a fundamental Higgs stops making sense. If $Q_\infty \geq M_{\rm Planck}$ this problem is hidden behind the Planck scale. This is true if $M_{\rm H} \lesssim 2M_{\rm W} = 164$ GeV. Of course we are using a very poor approximation, since we ignored all contributions of other particles to the β function, higher loops, and also because we trust the renormalization group equations all the way to the pole, which is certainly not correct. Nevertheless, it gives us a feeling that for some reasonable values of $M_{\rm H}$ the Standard Model can be extrapolated all the way to $M_{\rm Planck}$, but that for large values one will encounter the pole before $M_{\rm Planck}$.

If we increase $M_{\rm H}$ the scale Q_∞ will decrease, and at some point they meet. It is easy to solve Eq. (6.40) with $Q_\infty = M_{\rm H}$, and one finds $M_{\rm H} \approx 8M_{\rm W} \approx 650$ GeV. It does not make sense to increase $M_{\rm H}$ beyond this point, because then the mass of the scalar is larger than the scale up to which the theory makes sense.

All this was based on extrapolation of perturbation theory beyond its limits. It can be made more precise by putting the theory on a lattice to deal correctly with the non-perturbative physics. This confirms in a more rigorous way that there is an upper bound of about 700 GeV for the Higgs mass. For values of $M_{\rm H}$ below that bound, the theory should be viewed as an effective theory, valid only up to Q_∞. Sometimes this is also formulated in the following way: if we really want to make sense of the theory for arbitrarily large scales, we are forced to set the coupling constant to 0. Then the theory is "trivial", it is a free theory that is certainly valid for arbitrary scales, but not very interesting. The upper bound on $M_{\rm H}$ is usually referred to as the "triviality bound".

The Stability Bound. The expression for the ϕ^4 β-function given above ignored all other interactions. It is instructive to consider the complete β-function at one loop order:

$$\beta(\lambda) = \frac{1}{16\pi^2}\left[6\lambda^2 - 24y^4 + 12\lambda y^2 - \lambda(9g_2^2 + 3g_1^2) + \frac{9}{2}g_2^4 + 3g_2^2g_1^2 + \frac{3}{2}g_1^4\right] + \ldots \tag{6.41}$$

Here y can be any quark or lepton Yukawa coupling (leptons contribute with a relative factor $\frac{1}{3}$, since the quark contribution is enhanced by a color factor). In fact, each occurrence of y is an implicit sum over all quarks and leptons, but of course to a very good approximation only the top quark contributes.

If λ is small it is not the first term that dominates (as assumed earlier), but the second one. Then λ will decrease rather than increase, and one has to worry that it does not go through zero, since negative values of λ would correspond to an unstable Higgs potential. Requiring that this should not happen puts a lower bound on λ and hence on the Higgs mass. A detailed two-loop analysis of the coupled equations [12] (using the known top quark mass of about 175 GeV) gives a lower limit on the Standard Model Higgs mass of about 150 GeV. If we combine it with the upper bound coming from the requirement that λ remains finite below $M_{\rm Planck}$ we are left with a very small window between 150 and 160 GeV. Of course both bounds are different if we add extra particles to the Standard Model. But if we don't want to do that, and the Higgs is not found within this window, we can be pretty sure that the Standard Model must loose its validity in some way before the Planck mass is reached.

Note that even though λ may initially decrease with increasing energy scale, the Yukawa coupling decreases as well, and its contribution will eventually be smaller that the first term. Then at higher scales the value of λ starts to increase again, and hence the triviality problem is not solved by including the Yukawa coupling. Here we are using the fact that the top quark mass is still below the bound of 200 GeV mentioned in the previous section.

This presentation is actually a bit too naive. One should really use the full effective potential instead of the tree level potential. This has the effect of replacing $\lambda\phi^4$ by $\lambda(\phi)\phi^4$,

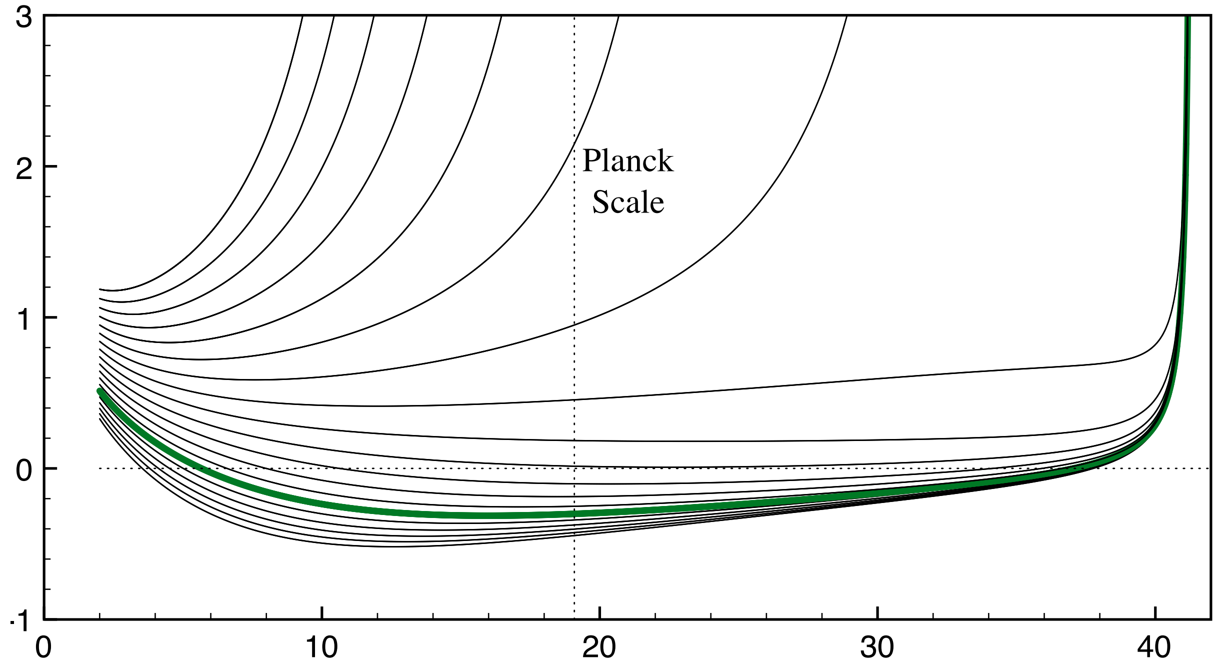


Figure 5: Running of the Higgs coupling constant λ using the one-loop β-function. The lines correspond to Higgs masses of $100 + 5n$ GeV, $n = 0, \ldots 19$. The green line corresponds to the observed Higgs mass of 125 GeV. The first line that does not cross zero is the one for a Higgs mass of 145 GeV, the first one with a Landau pole below the Planck scale is the one for a 170 GeV Higgs.

where $\lambda(\phi)$ is the running coupling constant evaluated at the scale ϕ. Since we have just argued that for large scales λ becomes positive again, it is clear that the potential is not really unbounded from below: for $\phi \to \infty$ the potential will eventually become positive. However, it is also clear that the potential develops a second minimum (in addition to the one that breaks $SU(3) \times SU(2) \times U(1)$) for a value of ϕ near the zero of λ. A problem arises then if that minimum is the global minimum of the potential, since one would then expect the Standard Model vacuum to be unstable (this is often called a "false vacuum"), and to decay to the true vacuum. In the true vacuum $SU(3) \times SU(2) \times U(1)$ would also break to $SU(3) \times U(1)$, but with a much larger Higgs v.e.v. and hence much larger W and Z masses. Note that the top Yukawa coupling only enters the β function for scales larger than the top quark mass; below that mass the top decouples. Hence if λ goes negative, this can only occur for scales much larger than $m_{\rm top}$, and hence inevitably the resulting W mass will be much than $m_{\rm top}$.

The lower limit of 150 GeV quoted above is based on an analysis of the effective potential, although it turns out that simply requiring that λ remains positive leads to essentially the same result. A recent two-loop analysis (see the second paper in [27]) yields a slightly smaller number, 140 GeV.

There is a further remark to be made here. We should probably not worry about the absolute stability of our vacuum, but rather about its lifetime. It could decay via tunneling or thermal fluctuations, or even as a result of high energy collisions in particle accelerators. All we need to require is that it lives longer than the age of the universe. This inevitably lowers the bound somewhat, but not by more than a few GeV [27].

7 Intermezzo: Standard Model problems

Despite its impressive successes many are not satisfied with the standard model. We have already mentioned the really serious problems: we don't know how to combine it consistently with gravity, and even if we could ignore gravity at least two of the coupling constants, the QED coupling and the Higgs self-coupling, blow up (have a Landau pole) if we start to probe arbitrarily small distances. These are not the most frequently mentioned problems of the Standard Model, though they are the only ones that show beyond any doubt that the Standard Model cannot be viewed as a fundamental theory for arbitrarily large energies. However, the Landau poles of the Higgs and QED lie beyond the Planck scale, so we will have to be able to deal with quantum gravity in order to address them.

The more common complaint can be summarized as follows. Given the rules we know for writing down consistent field theories with fermions, gauge bosons and scalars, there is an enormous number of theories we can write down. Which principle selects the Standard Model we observe, including its gauge group, the values of the gauge coupling constant, the fermion representation and their "triplication", the Yukawa couplings and last but not least the weak scale?

7.1 The Hierarchy Problem

In some cases there is a sharper way to ask the question. There are many dimensionless ratios between the parameters. Many of these ratios have extremely small values, for example $\frac{m_e}{m_t} \approx 10^{-6}$ or $\frac{M_{\rm W}}{M_{\rm Planck}} \approx 10^{-17}$. This does not look like the result of a random choice.

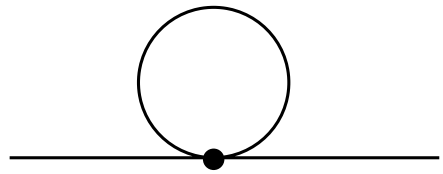
The problem with the second ratio is a little bit more serious than that of the first. This has to do with the differences in mass renormalization between fermions and scalars (which set the weak scale). For fermions one has

$$\delta_m \propto g^2 m \log(\Lambda/m) \ , \tag{7.1}$$

whereas for scalars

$$\delta_m = g\Lambda^2 \ , \tag{7.2}$$

where Λ is the cutoff and g the coupling constant. The latter contribution is due to the diagram



The mass renormalization is formally infinite in both cases, if we make the cut-off arbitrarily large. But the fermion mass correction has two positive features: it diverges only logarithmically, and it is proportional to the mass of the particle. Although $\log(\infty)$ is infinite, $\log(M)$ is a number of order 1 for any reasonable choice of M, such as the Planck scale. On the other hand, if we substitute the Planck mass for Λ in the scalar mass correction, the correction is 17 orders of magnitude large than the physical mass. In other words, there is no "protection mechanism" in the Standard Model to keep the scalar mass small.

This is not necessarily a fundamental problem. In both cases one can absorb the infinities into the bare mass and obtain any desired value for the physical mass. This requires huge readjustments at every loop order for the scalar mass, but one could retort that perturbation theory is a typically human activity, and that nature does not work order by order in perturbation theory. If there is some good reason why the physical parameter should be small, then it is not obvious that a protection mechanism is needed to keep it small in perturbation theory.

In other words, there are really two problems: why *is* a parameter small, and why does it *remain* small. To appreciate this, note that the scalar mass suffers from both problems, but that the electron mass is protected. We do not know why it is small, but at least all corrections are proportional to the mass itself. Another way to view the difference between these two cases is to check if one gains any new symmetries if the small parameter is put to zero. In the case of the electron that is true: one obtains a $U(1)$ chiral symmetry, which forbids any perturbative contributions to the electron mass. In terms of Feynman diagrams, one can view e_L and e_R as completely independent fermions, whose lines can be followed trough each diagram. If there is no $\bar{e}_L e_R$ vertex in the theory, it can never be generated. Note that this argument is strictly perturbative. For example the same reasoning could be used for quarks, but we know already that non-perturbative QCD effects break the chiral symmetry spontaneously.

On the other hand, one does not gain a symmetry if one puts a scalar mass to zero (naively one would expect to gain a scaling symmetry $\phi(x) \to l\phi(l^{-1}x)$, but this symmetry is explicitly broken by the mass scale introduced in any regularization procedure).

7.2 The Strong CP problem

There are many similar "small parameter" problems in the Standard Model. Another example is the θ parameter of QCD, which is extremely small experimentally. Setting it to zero enhances the symmetry of QCD by itself, by restoring P and CP, but unfortunately these are not symmetries of the full Standard Model. The CP violation which has been observed experimentally in the weak interactions (in the $K_0\bar{K}_0$ system) will via some higher loop corrections inevitably contribute to the θ parameter of QCD.

7.3 The Multiverse and Anthropic Reasoning

The trouble with all these small-parameter problems is that it is impossible to be certain that they are real problems. The Standard Model is perfectly consistent even if we never find a solution to these problems. It is possible that all theories we can write down are equally good, or that at least some infinite subset is, and that our universe was born with a given choice out of this parameter space. Another universe might have a different set of fields and parameters, which the inhabitants of that universe would call their Standard Model. In fact, most universes would presumably have no inhabitants at all, and what makes ours special is precisely that it does. This sort of reasoning leads to the "anthropic principle", and whatever one thinks of that, it is useful to remember that there might be parameters of the Standard Model that are part of the "boundary conditions" of our universe, and that can therefore never be determined from first principles.

It is certainly true that if we just change the parameter μ^2 and nothing else, there are drastic changes in our environment that would be fatal for our kind of life. For example, the amount of Carbon and Oxygen in our environment changes by an order of magnitude if μ^2 is changed by only about 10%. This is due to the fact (first pointed out by the famous astrophysicist Fred Hoyle) that the Carbon production rate is very sensitive to a resonance in the Carbon nucleus. But perhaps life can be formed in other ways if there is no Carbon. Even then, it is quite clear that a small value of the scale of the Standard Model (*i.e.* both the strong and the weak scale) in comparison to the Planck scale is important. This is the reason gravity is such a weak force. If we scale up the Standard Model so that $m_{\text{proton}}/M_{\text{Planck}} = 10^{-9}$ instead of 10^{-19}, gravity would crush any cluster of more than 10^{27} protons (such as a human being) into a black hole.* This argument does not rely on details like Carbon production, and would suggest that nothing of substantial complexity (and hence with a substantial number of protons) could exist. But is this, by itself, the reason that we observe such a small weak and strong scale? The answer to that question depends on the options that exist fundamentally, and how they are distributed and selected. This cannot be answered if all we know is the Standard Model. We need some "fundamental theory", that presumably must include gravity.

This just serves as an additional warning that some of the Standard Model problems we are trying to solve may not have any conventional solution at all. But not all small parameter problems are potentially "anthropic". For example, there is no such argument for the θ parameter of QCD.

* The maximal number of protons in a compact object is given roughly by $(M_{\text{Planck}}/m_{\text{proton}})^3$, and is about 10^{57} in our universe, the number of protons in a star. Indeed, stars have a broad range of values for brightness, but their masses are within one or two order of magnitude equal to 10^{57} proton masses.

7.4 Cosmological Problems

All the aforementioned problems are strictly theoretical ones. Some of these hint at potential inconsistencies, but at inaccessible energies, the others seem mostly easthetic, and we are not totally certain that they require a solution.

From the experimental point of view, some people see neutrino oscillations as a breakdown of the Standard Model, but it should be pointed out that oscillations are consistent with pure Dirac neutrino masses, and such masses fit in a very natural way in the structure of the Standard Model, apart from being unnaturally small.

For more serious threats to the Standard Model we have to look at cosmological data. For example, it is becoming clear that a large fraction of the matter in our universe is not built out of the quarks and leptons in the Standard Model. This so-called "dark matter" can be observed only because of its gravitational interactions, and several independent sources of information all point to its existence in abundance. This clearly implies that something is missing in the Standard Model. Another problem is the abundance of baryons over anti-baryons in our universe. There does not seem to be any way to explain this within the Standard Model. Not even putting it in as an initial condition of our universe would work, because one can show that any initial difference would be "washed out" by baryon number violating processes occurring at high energies and temperatures due to the baryon number anomaly. The Standard Model does contain all the ingredients to generate net baryon number (the so-called "Sacharov conditions", which require baryon number, C and CP violation, and a phase transition). However the phase transition at which baryon number is supposed to be generated, weak interaction symmetry breaking, turns out to be not strong enough. This also points at a need for additional physics beyond the Standard Model.

Another example, though strictly speaking not purely in the Standard Model, is the cosmological constant. This is a fine-tuning problem related to the vacuum expectation value of the potential.

The value of the potential is irrelevant if a theory not coupled to gravity. Changing the value simply shifts the Lagrangian by a constant, but this constant does not affect the equations of motion or the Feynman rules. If we couple a theory to gravity such a shift is no longer irrelevant, due to the factor $\sqrt{g}$ that appears as a factor in front of the action. A term $-\int d^4x\sqrt{g}V_0$ was first introduced (and again withdrawn) by Einstein as an addition to the action for gravity. Its presence implies a non-zero value for the cosmological constant

$$\Lambda_c = \frac{8\pi G_N}{c^2} V_0 \tag{7.3}$$

If $V_0 > 0$ the solution to the matter-free Einstein equations is not Minkowski space, but de-Sitter space; for $\Lambda < 0$ one gets anti de-Sitter space. For a long time there was a strongly held belief that its value would be exactly zero, but recent observations point to a non-zero and positive value of the order of 10^{-84} GeV^2, which implies that the expansion of our universe is accelerating.

Perhaps the "natural" scale for Λ would be the Planck scale, in which case the expected value is about 10^{38} GeV^2. But even if a Planck-scale contribution can be avoided, the shift in the Higgs potential due to weak interaction symmetry breaking is about $-10^{-33}M_{\rm H}^2$, where $M_{\rm H}$ is the Higgs mass. Clearly this is also much larger than the lower limit, and represents a fine-tuning problem that is much worse than that of the the weak interaction scale in comparison to the Planck scale.

It is common practice to ignore the cosmological constant problem when one tries to solve the other fine-tuning problems. One hopes that a full understanding of gravity will lead to an understanding of why the cosmological is so small. This may be correct, but it is also possible that all fine-tuning problems are related and have a common solution. If that is true, we would be wasting our time by trying to understand the the smallness of θ and $M_{\rm W}/M_{\rm Planck}$ while ignoring the smallness of Λ_c.

8 Grand Unification

In this chapter we discuss the idea of embedding the Standard Model in a larger gauge group. One of the motivations for doing that is the convergence of the coupling constants at high energies. We begin by examining this more closely.

8.1 Convergence of Standard Model Couplings

If one follows the evolution of the three Standard Model coupling constants one observes a tendency to converge. The $SU(3)$ coupling constant is largest and falls off fastest, the $SU(2)$ coupling constant is the next smallest and drops off more slowly, and the QED coupling constant is the smallest and increases. The simplest way to study this is to plot $\frac{1}{\bar{g}(t)^2}$ as a function of t. In the one-loop approximation, which is in fact accurate enough for most purposes, each function is a straight line. This follows from Eqn. (6.26). It was observed around 1975 that according to the data available at that time these three lines went through the same point at a scale of about 10^{15} GeV.

Plotting the three lines in one figure is a bit misleading, because it suggests that we know their relative normalization. This is certainly not true for $U(1)$ coupling constants, because the only measurable quantity is the product of charges and the coupling constant. Hence we can multiply the coupling constant by an arbitrary factor, and divide all charges by the same factor. This does not happen for non-abelian groups, because one can agree in advance on a normalization of the Lie algebra generators.

So we have a plot with three lines, one of which can be scaled by an arbitrary factor. Clearly there is always a factor so that the three lines go through one point, unless two of them are parallel. Indeed, the observed unification at 10^{15} GeV does not occur for the coupling constants g_1, g_2 and g_3 introduced in chapter 4 but for $1.291g_1$, g_2 and g_3. The excitement caused by this discovery had two reasons: first that the scale, 10^{15} was "reasonable", and second that $1.291 \approx \sqrt{5/3}$, a number that can be explained by group theory as we will see in a moment.

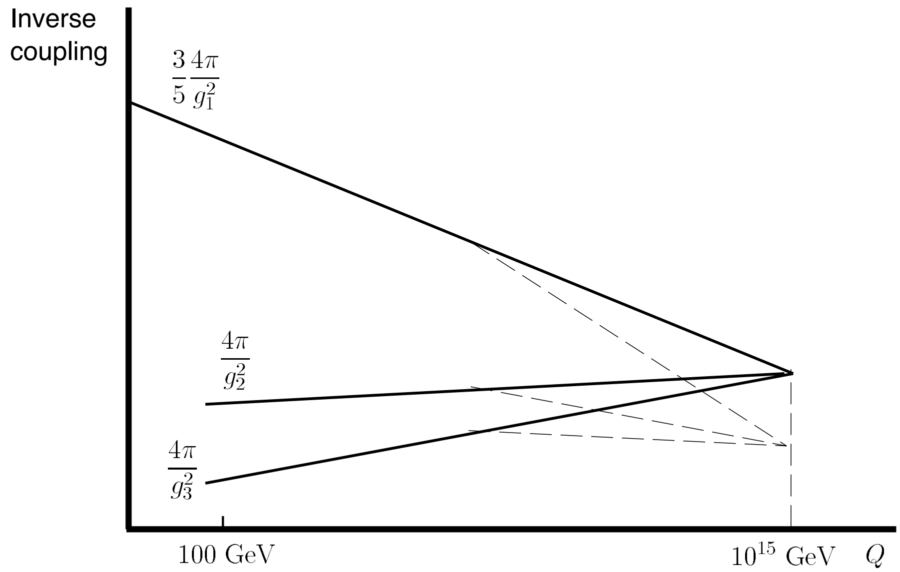


Figure 6: Unification of coupling constants. The dashed lines are explained in section 8.1.1

Any such statement is based on assumptions about the physics beyond the weak scale. Since any particle in $SU(3) \times SU(2) \times U(1)$ representations alters the β-functions, one is assuming that there are no (or very few) unobserved particles between 100 and 10^{15} GeV, except for $SU(3) \times SU(2) \times U(1)$ singlets. Any unknown massive particle changes the slope of one or more of the lines. Since it only has effects for scales larger than its own mass, the result would be a kink in the straight lines in the figure. Note that any additional matter affects all three lines by bending them in the same direction (namely downward, with increasing energy), since matter contributions to b_0 always have the same sign. We will see in a moment that it is not quite true that no matter is allowed in the "desert" between 100 and 10^{15} GeV, since there is a natural mechanism for bending all lines in exactly the right way so that they continue to merge, as shown by the dashed line in Fig. 6.

The fact that two coupling constants are equal at a certain scale need not have physical implications. They may just cross each other and continue. But one is tempted to conclude that it has a deeper meaning, namely that the three groups of the Standard Model somehow are combined into one "unified" theory.

8.1.1 Coupling Constant Unification: Generalities

One of the mechanisms to give a physical explanation for the apparent convergence of coupling constant unification is to assume that at the convergence scale the Standard Model gauge group enlarges to a group G containing it. The group G is assumed to break at that scale by means of the Higgs mechanism to a subgroup, which is either directly the Standard Model, or some intermediate group.

Here we want to discuss in general what happens to coupling constants if a simple group G breaks to a subgroup $H_1 \dots H_k$. We start with the unbroken G-theory. Its kinetic term is the canonical one,

$$-\tfrac{1}{4}F^a_{\mu\nu}F^{\mu\nu,a} \ . \tag{8.1}$$

The normalization guarantees that the propagator for each field A^a_μ is normalized in the standard way, namely as $-i\delta_{ab}g_{\mu\nu}/k^2$ (Feynman gauge). This is important here since only if the fields are normalized properly one can read off the coupling constant from the Lagrangian. The gauge coupling to all fields is governed by the covariant derivative

$$D_\mu = \partial_\mu - igA^a_\mu T^a \ , \tag{8.2}$$

where the generators are in the representation of the field they act on.

Now suppose by some mechanism one introduces a gauge boson mass matrix

$$+\tfrac{1}{2}M_{ab}A^a_\mu A^{\mu,b} \ . \tag{8.3}$$

The mass matrix is symmetric, and can be diagonalized by an orthogonal transformation S:

$$M_{\rm diag} = SMS^T \tag{8.4}$$

This introduces new mass eigenstate fields B^a_μ

$$B^a_\mu = \sum_b S_{ab}A^b_\mu \ . \tag{8.5}$$

Now we express the other terms in the action in terms of the new fields B. We note first of all that the quadratic terms in the kinetic action are not affected since S is orthogonal (we worry about the gauge self-couplings later). For the covariant derivatives we find

$$D_\mu = \partial_\mu - igB^a_\mu S_{ab}T^b \ . \tag{8.6}$$

Suppose now that some of the fields B remain massless, *i.e.* that M has some zero eigenvalues. Massless vector fields must necessarily be gauge bosons, and hence whatever mechanism we use to generate the mass matrix M, it must be such that the massless gauge bosons couple to a closed set of Lie algebra generators. From the form of the covariant derivative we read off these generators

$$U^{\hat a} = S_{\hat a\hat b}T^b \ . \tag{8.7}$$

Here the hat on the label a indicates those labels in the set for which $B^{\hat a}$ is massless.

In order to define a coupling constant, we have to fix not only a normalization of the gauge fields (as we have already done), but also for the generators. This works differently in abelian and in non-abelian theories. The canonical normalization for non-abelian gauge theories is given in appendix B.* For $SU(N)$ groups, it is $\mathrm{Tr}\ T^aT^b = \frac{1}{2}\delta^{ab}$ in the N dimensional representation (the vector representation). In general, one has

$$\mathrm{Tr}_R\ T^aT^b = I_2(R)\delta^{ab}\ , \tag{8.8}$$

where the subscript R indicates the representation of G under consideration, and I_2 is the second index, defined in appendix B. These are integers, whose normalization has been fixed *a priori* for all non-abelian Lie algebras. For abelian groups there is no "intrinsic" normalization. However, the electromagnetic coupling constant has a definite value because we have fixed the charge of the proton to be 1, thus fixing the overall normalization of the $U(1)_{\rm em}$ generator.

In the present case, orthogonality of S implies

$$\mathrm{Tr}_R U^{\hat{a}}U^{\hat{b}} = I_2^G(R)\delta^{\hat{a}\hat{b}}\ . \tag{8.9}$$

Here we have indicated that the second index of G appears on the right-hand side. The generators U form a sub-algebra H of G. For the time being, we will assume that H is a simple non-abelian Lie algebra. In terms of H-representations, the representation space of R decomposes (in general) to a direct sum (the notation $\sum_k \oplus$ indicates a direct sum over components labelled by k)

$$R \quad \longrightarrow \quad \sum_k \oplus\ r_k\ , \tag{8.10}$$

where r_i is an irreducible representation of H. Now define the correctly normalized generators $\hat{U}$, which differ from U by an overall factor, $\hat{U} = \lambda U$. These generators satisfy

$$\mathrm{Tr}_R \hat{U}^{\hat{a}}\hat{U}^{\hat{b}} = I_2^H(R)\delta^{\hat{a}\hat{b}}\ . \tag{8.11}$$

Note that the trace is still over the same space, the full representation space of the G-representation R. However, the index that appears is that of H, and not that of G as above [Here we use the symbol R for the H-representation $\sum_k \oplus r_k$]. Comparing the two expressions, we find that the normalization factor λ is given by

$$\lambda^2 = \frac{I_2^H(R)}{I_2^G(R)} \equiv I(G/H)\ . \tag{8.12}$$

Here $I(G/H)$ is called the embedding index for the embedding of H in G. Note that λ must be the same for all representations, so that the representation dependence must cancel in the ratio. It is then easy to see that $I(G/H)$ must be an integer for any algebra G with a representation of index 1 (such as $SU(N)$), since one can use that representation

* In addition to the normalization one should also make sure that the structure constants f^{abc} have some standard form. Once the normalization is fixed, this can always be achieved by choosing the orthogonal matrix S appropriately.

to define $I(G/H)$. In fact it can be shown that $I(G/H)$ is always an integer for simple non-abelian groups G and H.

Let us now return to the covariant derivative. When we express it in terms of the canonically normalized fields B and generators $\hat{U}$, we can now read off the correctly normalized coupling constant, which is the factor of these quantities:

$$g_H = \frac{1}{\sqrt{I(G/H)}} g \tag{8.13}$$

The discussion of the gauge-boson self-couplings is essentially the same, with the representation R equal to the adjoint representation. This decomposes into several irreducible components, one of which is the adjoint representation of H. To get the correctly normalized structure constants, we need the same normalization factor λ, which is absorbed in the coupling constant. The original gauge kinetic terms now split into gauge kinetic terms for the massless gauge bosons $B^{\hat{a}}$, plus minimal couplings of these bosons to the massive B-bosons.

If there is more than one simple factor in H, one defines a separate embedding index for each of them. For the coupling constant of the $i^{\rm th}$ factor one finds then

$$g_i = \frac{1}{\sqrt{I(G/H_i)}} g \tag{8.14}$$

In other words, at the unification scale (at which the symmetry breaks), the quantities $\sqrt{I(G/H_i)}g_i$ are equal to each other and to g.

In this situation the logarithm of the unification scale is determined by solving a linear equation of the form

$$\frac{1}{I_i}\left(\frac{1}{g_i^2} + b_0^i t\right) = \frac{1}{I_j}\left(\frac{1}{g_j^2} + b_0^j t\right) , \tag{8.15}$$

where i and j label two factors of the gauge group, and I_i is an abbreviation of $I(G/H_i)$.

It is interesting to examine the effect on convergence of coupling constants due to extra matter representations . Suppose two coupling constants converge at some large scale t_0 when they are naively extrapolated from some low scale. This means that Eq. (8.15) is satisfied for $t = t_0$. We use the word naively, because inevitably such a statement implies an assumption about the presence of matter between the low scale and t_0. Suppose we modify the evolution by adding matter in a representation R of the low energy gauge group $H_i \times H_j$. This changes b_0^i to $b_0^i - sI_2^i(R)$, where s is a spin-dependent factor, and similarly for j. Here I_2^i denotes of course the second index of the group H_i. At the scale t_0 the groups H_i and H_j are embedded in G. Suppose now that the $H_i \times H_j$ representation R forms exactly a representation of G. Then the modification of each side of Eq. (8.15) can be written as $\frac{I_2(R_i)}{I_i}t = I_2^G(R)t$ for i as well as j. The result is thus independent of i, and hence both sides of the equation change in exactly the same way. The solution t_0 remains thus unchanged (the value of the coupling constant at unification it does change, and is in fact increased.) This is illustrated by the dashed lines in fig. 6. Here the extra matter is assumed to have a mass somewhere in the desert between $M_{\rm W}$ and $M_{\rm GUT}$. This result is completely independent of the precise decomposition of R with respect to the subgroup, but does assume that all its components get roughly the same mass. Even if unification takes place, it need not always happen that a $H_i \times H_j$ representation R is exactly a G representation. It might happen that to complete it to a G representation additional particles with a mass near the unification scale are required. In that case t_0 *does* change, and furthermore simultaneous unification for three or more coupling constants may be affected. In any case, the lesson is that the fact that couplings unify, or the scale at which this happened, is less sensitive to intermediate matter than one might have expected.

8.2 Electric Charge Quantization

The electric charge of the proton is the exact opposite of the electron charge. This is known with enormous precision, because any small deviation would lead to small net charges of matter. This fact is not really explained in $SU(3) \times SU(2) \times U(1)$. Although anomaly cancellation and mass generation by the Higgs mechanism impose some constraints, the group $U(1)_Y$ allows representations with any real charge.

Furthermore we have only observed particles with charges that are a multiple of the electron charge. There have been many searches for fractional charges, starting with the famous Millikan experiment. Nothing has been found, but unfortunately we have no idea what masses such particles might have and how abundant they should be, if they exist at all.

What makes this even more puzzling is that the components of hadrons, the quarks, have third-integral charges, and that such charges also have never been observed. We have a qualitative understanding of this fact. It is a direct consequence of two features of $SU(3)_{\rm QCD} \times U(1)_{\rm QED}$ and the representations that occur in nature, namely

- QCD confines color
- The $SU(3) \times U(1)$ representations (R, q) that occur in nature satisfy an empirical constraint, namely $\frac{t}{3} + q = 0 \text{ mod } 1$.

Here t is the "triality" of the $SU(3)$ representation R. In terms of Young tableaux, t is equal to the number of boxes modulo 3; in terms of Dynkin labels (a_1, a_2) the triality is defined as $(a_1 + 2a_2)$, if $(1, 0)$ is the triplet and $(0, 1)$ the anti-triplet representation. Yet another equivalent way of introducing it is by defining the center element $z = \text{diag }(e^{2\pi i/3}, e^{2\pi i/3}, e^{2\pi i/3})$ of $SU(3)$ (a center element commutes with all other group elements). In any irreducible representation the matrix $\hat{z}$ representing z must satisfy the group property $\hat{z}^3 = 1$. Hence the eigenvalues of $\hat{z}$ must either be $e^{2\pi i/3}$, $e^{-2\pi i/3}$ or 1, and they must be all equal because otherwise $\hat{z}$ would not commute with all other group elements (according to Schur's lemma, the only matrix that commutes with all representation matrices of an irreducible representation is a multiple of the identity). Now define t by writing this eigenvalue as $e^{2\pi i t/3}$. Note that t is opposite for complex conjugate representations, and 0 for real ones, including the adjoint. Confinement allows only particles with total triality equal to zero in the spectrum, and then the observed charge quantization follows.

A similar relation holds for the known $SU(3)\times SU(2)\times U(1)$ representations, namely

$$t/3+s/2+Y=0 \text{ mod } 1\ , \tag{8.16}$$

where s is $SU(2)$ "duality" (equal to 1 for spinor representations and to 0 for vectors). Because the electromagnetic charge is $Q_{\rm em}=T_3+Y$ this leads automatically to the $SU(3)\times U(1)$ relation of the previous paragraph.

Mathematically this means that the Standard Model gauge group we have observed so far is not $SU(3)\times SU(2)\times U(1)$, but $S(U(3)\times U(2))$. The fundamental representation of this group consists of matrices of the form

$$U=\begin{pmatrix} U_3 & 0\\ 0 & U_2\end{pmatrix}\ , \tag{8.17}$$

where U_i is an element of $U(i)$, with the condition det $U=1$. The latter is precisely the charge quantization condition. The Lie-algebra of this group is exactly the same as that of $SU(3)\times SU(2)\times U(1)$, but the groups are globally different. A comparable situation occurs between the groups $SO(3)$ and $SU(2)$: they have the same Lie algebra, but the latter has spinor representations, and the former does not. It is precisely the absence of certain representations from the spectrum that leads us to conclude that the group is $S(U(3)\times U(2))$, and not $SU(3)\times SU(2)\times U(1)$.

To see how charge quantization arises, note that elements of $S(U(3)\times U(2))$ can be parametrized as

$$g(\hat U_3,\hat U_2,\phi)=\begin{pmatrix} \hat U_3 e^{-i\phi/3} & 0\\ 0 & \hat U_2 e^{i\phi/2}\end{pmatrix}\ , \tag{8.18}$$

where the hatted matrices are elements of the groups $SU(3)$ and $SU(2)$ respectively. This includes the group element $g(z,y,2\pi)$, where $z=$ diag $(e^{2\pi i/3},e^{2\pi i/3},e^{2\pi i/3})$ and $y=$ diag $(-1,-1)$. Note that $g(z,y,2\pi)=1$. This element can be obtained as sequence of products, and can be reached by multiplying a series of group elements that are close to the identity. Since products are preserved in any representation, by definition of the latter, this element must equal the identity in *any* representation. Consider then a representation that is trivial in $SU(3)$ (a color singlet) and in $SU(2)$. In the Lie-algebra $SU(3)\times SU(2)\times U(1)$, this could be the representation $R=(1,1,q)$, with representation matrices $g_R(\hat U_3,\hat U_2,\phi)=e^{iq\phi}$, for any real value of q. But in the group $S(U(3)\times U(2))$ most of these values of q do not give rise to valid representations, because the element $g_R(z,y,2\pi)=e^{2\pi iq}$ must be equal to the identity. Hence the charge q must be an integer. This shows that we get integral charges for $SU(3)\times SU(2)$ singlets, and hence we see that q is normalized in the right way to be interpreted as the Standard Model charge Y. Otherwise we would have had to introduce a normalization factor at this point.

Having normalized the charge correctly with respect to the Standard Model conventions, we may now consider other representations. For a general representation $g_R(\hat U_3,\hat U_2,\phi)$ the element $g(z,y,2\pi)$ is given by

$$e^{2\pi it/3}e^{2\pi is/2}e^{2\pi iY} \tag{8.19}$$

Since this must be one, we see that all representations of $S(U(3) \times U(2))$ indeed satisfy the observed quantization condition Eq. (8.16).

Of course all this depends on the fermion representations observed so far, and we do not know whether this is a fundamental property of nature, a coincidence, or something else. Note that the requirement of anomaly cancellation does not really give us much choice for the charge Y, at least not within one family of 15 fermions. However, this already changes if we add an $SU(3) \times SU(2)$ singlet of arbitrary charge. This gives us 16 fermions, the same as a Standard Model family with a right-handed neutrino. Even if we require that the Higgs boson can give a mass to all fermions (which implies a relation between the Y charges of the singlets and the doublets) there is a solution to anomaly cancellation with arbitrary real charge. We also add a massive fermion in the representation $(1, 1, q)_L + (1, 1, -q)_L$, where q is completely arbitrary, and in particular could be fractional.

8.3 Gauge Unification in $SU(5)$ GUTs.

We now put the two pieces of information discussed in the last two sections together. There is a rather obvious way of realizing the global group $S(U(3) \times U(2))$ discussed above, namely by embedding it in the simple group $SU(5)$. The word "simple" is used here in the mathematical sense, and refers to simple Lie algebras, see Appendix B.1. Just think of it as a group consisting of just a single factor, instead of the three factors of $SU(3) \times SU(2) \times U(1)$. Since it is a simple group, the normalization of all the generators of its algebra is fixed: there are no real charges that can take arbitrary values. Charge quantization is now automatic. As we have seen, charge quantization is already guaranteed by the smaller group $S(U(3) \times U(2))$, but this looks rather ad-hoc.

If we combine this group-theoretical fact with the approximate unification of coupling constants, a very compelling picture suggests itself: perhaps at high enough energies there is an enlarged gauge symmetry, $SU(5)$, and the Standard Model somehow emerges from that. In 1974, when unification was first discussed, the data on coupling constants were all in agreement with unification of the Standard Model into a simple group, $SU(5)$. Historically, the group-theoretical embedding was discovered first by Georgi and Glashow [13], and the coupling unification argument was first given about half a year later by Georgi, Quinn and Weinberg [14].

The present data, especially the high precision data from LEP, do not support the exact convergence anymore, at least not for "minimal $SU(5)$", without extra matter in the desert. Remarkably, the three lines go through a single point nearly exactly if we extend the Standard Model by making it supersymmetric. This will be discussed in chapter 9. Supersymmetry is a symmetry between bosons and fermions. However, it turns out that none of the Standard Model particles can be mapped into each other by this symmetry. This implies (at least) adding a boson for every Standard Model fermion and a fermion for each boson, always in the same $SU(3) \times SU(2) \times U(1)$ representations, and it also implies doubling the Higgs sector. If we assume that all this extra matter has a mass of around 1 TeV, it makes the three lines bend in precisely the right way. So far

LHC has not revealed any evidence for supersymmetric particles of masses near or below 1 TeV. Nevertheless, the concept of $SU(5)$ unification is still important enough to have a closer look.

The idea of $SU(5)$ gauge unification is extremely simple. One builds a gauge theory with an $SU(5)$ gauge group, and then one breaks this symmetry group spontaneously to its $SU(3) \times SU(2) \times U(1)$ subgroup. The spontaneous breaking is achieved by a new Higgs-like field that must be added to the theory. This Higgs field is assumed to get a vev of about 10^{15} GeV, the energy scale at which the three gauge coupling lines cross in fig. 6. This is called the GUT scale This gives a mass to all the gauge bosons that are in $SU(5)$, but not in $SU(3) \times SU(2) \times U(1)$. Below the GUT scale these extra gauge bosons do not contribute anymore to the running of the gauge couplings, as they run to lower energies. They go their own way, resulting in fig. 6. In this way the three Standard Model gauge groups are unified, and so are the gauge couplings.

The smallest simple group in which one can embed the Standard Model group is $SU(5)$. The gauge action is just the canonical one, with a coupling constant g_5. The fermions are minimally coupled in a way that depends only on their $SU(5)$ representations. This representation must be anomaly free, and we will need three copies to get three families. It must also be complex, since otherwise we would expect it to be massive, and furthermore the theory would be invariant under C and P, while the standard model is not. It must have at least 15 or 16 Weyl fermions per family, and preferably not more. This is just a rough guide towards the right answer; ultimately we must find the correct $SU(3)\times SU(2)\times U(1)$ representations by working out the breaking of $SU(5)$.

Note that we will use the left-handed representation for all the fermions. This allows us to transform them freely into each other. One cannot make internal rotations among fermions with different handedness.

8.4 Embedding the Standard Model Gauge Group.

Let us examine more closely how $SU(3) \times SU(2) \times U(1)$ is embedded in $SU(5)$. The group $SU(5)$ is defined as the set of 5×5 unitary matrices with determinant one. Now consider the subset of matrices of the form

$$U = \begin{pmatrix} U_3 & 0 \\ 0 & U_2 \end{pmatrix} , \tag{8.20}$$

where U_3 and U_2 are unitary 3×3 and 2×2 matrices satisfying the relation $\det U_3 \det U_2 = 1$. This is precisely the group $S(U(3)\times U(2))$ identified in section 8.2 as the global group of the Standard Model. If we write $U_3 = e^{i\phi}\hat{U}_3$ and $U_2 = e^{i\chi}\hat{U}_2$ where $\hat{U}_3$ and $\hat{U}_2$ have determinant 1, then we have identified the $SU(3)$ and $SU(2)$ subgroups. The phases must satisfy $3\phi + 2\chi = 0 \bmod 2\pi$. This leaves one independent phase, corresponding to the $U(1)$.

8.4.1 Decomposition of $SU(5)$ Representations

If the fundamental theory underlying the Standard Model is an $SU(5)$ GUT, all matter must belong to some $SU(5)$ representation. The available representations and their dimensions can be enumerated using Young tableaux, as discussed in appendix B. We will have to determine how these representations decompose if the symmetry breaks to $SU(3) \times SU(2) \times U(1)$. This is usually written as

$$R \to \sum_i (r_i, s_i, y_i) \ , \tag{8.21}$$

where R is an $SU(5)$ representation, r_i an $SU(3)$ representation, s_i an $SU(2)$ representation, and y_i the $U(1)$ charge. If this decomposition is known for the vector representation, then the subgroup embedding is completely fixed, and hence we can compute the decomposition for all other representations. There are several ways of doing that. One method is to construct the representation matrices in $SU(5)$ and decomposing the space on which they act into $SU(3) \times SU(2)$ blocks, as we did above for the vector representation. A more often used method is to build representations as a tensor product of vector representations, and then work out the tensor products of the $SU(3)$ and $SU(2)$ components. After doing that one works out the $U(1)$ charges for each component.

In the following we denote $SU(5)$ representations by their dimension in bold face, and the complex conjugate representation of an $SU(5)$ representation by an asterisk. In appendix B.6 we derive the decomposition of the representations of most interest, namely the **5**, the **10** and the **24**, the adjoint representation. Here we just summarize the result, including for completeness also the symmetric tensor **15**.

$$\mathbf{5} \to (3, 1, -\tfrac{1}{3}q) + (1, 2, \tfrac{1}{2}q) \tag{8.22}$$

$$\mathbf{24} \to (8, 1, 0) + (1, 3, 0) + (1, 1, 0) + (3, 2, -\tfrac{5}{6}q) + (3^*, 2, \tfrac{5}{6}q) \ . \tag{8.23}$$

$$\mathbf{10} \to (3^*, 1, -\frac{2}{3}q) + (1, 1, q) + (3, 2, \frac{1}{6}q) \ , \tag{8.24}$$

$$\mathbf{15} \to (6, 1, -\frac{2}{3}q) + (1, 3, q) + (3, 2, \frac{1}{6}q) \ , \tag{8.25}$$

Here we have allowed for an arbitrary real factor q since the normalization of $U(1)$ charges is not fixed by the algebra. The $SU(3)$ and $SU(2)$ generators can simply be taken as a subset of the $SU(5)$ generators.

8.4.2 Normalization of Generators.

From the point of view of $SU(5)$ there is a natural normalization for the $U(1)$ generator. We choose the canonical normalization for the vector representation of $SU(N)$, so that $\mathrm{Tr}\ T^a T^b = \frac{1}{2}\delta^{ab}$. It is important that this trace is proportional to δ_{ab}, since this was implicitly assumed in writing the gauge kinetic terms. Note that this normalization is indeed the one we used previously in $SU(2)$ to derive the relation between T_3, Y and the electric charge, and in the computation of the β function.

If we make sure that the $SU(3) \times SU(2) \times U(1)$ generators all have the same normalization, we can choose a basis for the 24 $SU(5)$ generators consisting of 12 $SU(3) \times SU(2) \times U(1)$ generators (numbered $1 \ldots 12$) and 12 remaining ones. Then

$$\sum_{a=1}^{24} A_\mu^a T^a = \sum_{a=1}^{12} A_\mu^a T^a + \quad \text{rest} \; . \tag{8.26}$$

(The terms denoted "rest" will be discussed later.) The properly normalized generators appear in the Lagrangian in combination with the unified coupling constant g_5. If we want to view our $U(1)$ generator directly as a properly normalized generator, we should choose $T_Y = \sqrt{3/5}$ diag $(-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, \frac{1}{2}, \frac{1}{2})$, which satisfies $\text{Tr} T_Y^2 = \frac{1}{2}$, in other words, the factor q introduced above equals $\sqrt{3/5}$.

If we now compare the $SU(5)$ minimal couplings with those of the Standard Model, we get immediately the relations $g_2 = g_3 = g_5, g_1 = \sqrt{3/5} g_5$. These are precisely the relations required for coupling constant unification (according to the pre-LEP data at least). From now on we will absorb the factor $\sqrt{3/5}$ in the definition of the coupling constant, so that content of the representation **5** is $(3, 1, -\frac{1}{3}) + (1, 2, \frac{1}{2})$, *i.e* we set $q = 1$ from here on. The last entry is now precisely the Y-charge as defined previously.

8.5 Fermion Representations

We now have to decide which representations of $SU(5)$ to use for the fermions, the quarks and leptons. The most important criterium is that they should contain at least the particles in one family, and preferably nothing more. However, we can already get a hint from anomaly cancellation.

8.5.1 Intuition from Anomaly Cancellation.

A quick inspection of some of the smallest $SU(5)$ representations immediately suggests an obvious solution. The smallest non-trivial representation of $SU(5)$ is the **5**, the vector representation. Its anomaly is normalized to 1. The symmetric tensor has dimension 15, and its anomaly is 9. The anti-symmetric tensor has dimension 10 and anomaly 1. All these representations are complex, and their conjugates have the opposite anomaly. Another interesting representation is **24**, the adjoint, which has anomaly zero because it is real. The next smallest representation has dimension 35, and that is a bit too large to be of interest. Clearly the only reasonable solution to the conditions listed above is to take $\mathbf{5}^* + \mathbf{10}$, of course all in the left-handed representation.

8.5.2 Matter in the Five-Dimensional Representation.

The representations contained in the **5** do not match any Standard Model particle, but the complex conjugates do. Hence we choose the anomaly-free representation $\mathbf{5}^* + \mathbf{10}$ (we could just as easily have conjugated the embedding of $SU(3) \times SU(2) \times U(1)$ in the **5**, but that is not the standard convention). Now the $\mathbf{5}^*$ precisely contains particles with the quantum numbers of d^c, e^- and ν, *i.e.* the representation

$$(3^*, 1, \frac{1}{3}) + (1, 2, -\frac{1}{2}) \tag{8.27}$$

Here and in the following all fermions are **left-handed** unless explicit subscripts R are shown.

8.5.3 Particle Content of the Ten-dimensional Representation

The decomposition of the **10** is

$$\mathbf{10} \to (3^*, 1, -\frac{2}{3}) + (1, 1, 1) + (3, 2, \frac{1}{6}) \ , \tag{8.28}$$

and we recognize the representations of the particles u^c, e^+ and the doublet u, d. Thus the $SU(5)$ representation $\mathbf{5}^* + \mathbf{10}$ contains precisely one family of the Standard Model.

8.5.4 Detailed Particle Decompositions

The precise decomposition of this representation into $SU(3)\times U(1)$ particle representations is as follows. By convention, the **5** contains $(3, 1, -\frac{1}{3}) + (1, 2, \frac{1}{2})$. In the five dimensional space, the first three components are reserved for $SU(3)$, and the last two for $SU(2)$. Within $SU(2)$ the ordering of the doublet is important, because $SU(2)$ is eventually broken, and two members of the same doublet will become particles with different charges. We will choose the 4^{th} component to coincide with the upper component of the $SU(2)$ doublet. Now we are able to write down the decomposition of the **5**: $(d_1, d_2, d_3, e^+, \nu)$, including the color index for the d's. Here d, e^+ and ν are nothing but short-hand notations for certain $SU(3) \times U(1)$ representations. Then the $\mathbf{5}^*$ decomposes to

$$\Psi = (d_1^c, d_2^c, d_3^c, e^-, \nu) \ . \tag{8.29}$$

[There is one subtlety here. In a normal $SU(2)$ doublet the upper component has an electric charge that is higher (by one unit) than that of the lower, because $Q_{\text{em}} = T_3 + Y$. This is true for the doublet (e^+, ν) in the **5** but not for the doublet (e^-, ν) in the $\mathbf{5}^*$. The reason is simple: the doublet in the $\mathbf{5}^*$ transforms in the complex conjugate representation 2^*, and not in the 2. These representations are equivalent, but the equivalence relation involves the invariant tensor ϵ_{ij}, which turns the doublet upside down.]

Now we construct the **10** by taking the anti-symmetric product of two **5**'s. This field is most easily represented by a 5×5 matrix, whose elements i, j have the quantum numbers of the tensor product of the i^{th} components times the j^{th} component of the **5**. Here $e^+ d_i$

yields the $SU(3)\times U(1)$ representation of u_i and $\epsilon_{ijk}d_i d_j$ that of u^c_k. The result is

$$\Delta = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & u^c_3 & -u^c_2 & -u_1 & -d_1 \\ -u^c_3 & 0 & u^c_1 & -u_2 & -d_2 \\ u^c_2 & -u^c_1 & 0 & -u_3 & -d_3 \\ u_1 & u_2 & u_3 & 0 & -e^+ \\ d_1 & d_2 & d_3 & e^+ & 0 \end{pmatrix} \tag{8.30}$$

The factor $\frac{1}{\sqrt{2}}$ is added to ensure that the kinetic terms have the proper normalization (note that every field appears twice in the **10**).

8.5.5 Distributing Family Members.

This describes one family, but one word of caution is needed. In the Standard Model it is customary to say that u and d belong to the first family, c and s to the second and t and b to the third, but that is not really true. The group $SU(2)$ does not relate the mass eigenstates u and d, but u and a linear combination of d, s and b (which is dominated by d). With the leptons the situation is even less clear. If, by definition, u belongs to the first family, then how do we know that its family member is e, μ or τ? We don't, and it makes no difference because there are no transitions between leptons and quarks in the Standard Model, or, equivalently, they never occur within one irreducible representation. However within $SU(5)$ they do, and it becomes meaningful to ask who belongs to the first family, e, μ or τ, or some linear combination. This question can only be answered once we know the mass matrices, and we will return to it later.

8.6 The Standard Model Higgs Field.

The Standard Model Higgs field denoted ϕ in section 4.2 belongs to the representation $(1,2,\frac{1}{2})$. In $SU(5)$ there is no two-dimensional complex representation that precisely contains it. We are forced to choose a bigger one, and hence get additional particles. It is clear from the previous paragraphs that the representation **5** contains $(1,2,\frac{1}{2})$, and hence is a natural candidate, and indeed the smallest one. It contains three extra components in the representation $(3,1,-\frac{1}{3})$, corresponding to scalars which have not been observed (yet). This $SU(5)$ extension of the Standard Model Higgs field will be denoted as H. There are other possibilities, see section 8.9.

8.7 Choosing the GUT-breaking Higgs Field

Up to now we have only embedded $SU(3)\times SU(2)\times U(1)$ in $SU(5)$. Clearly the full $SU(5)$ is not an exact symmetry of nature, and thus we have to find a mechanism to break $SU(5)$ to $SU(3)\times SU(2)\times U(1)$. In this process the 12 Standard Model gauge bosons should remain massless, and the other 12 should become massive. We will try to do this by the only method we know, the Higgs mechanism. Note that the name "Higgs" is used here as a generic name for a mechanism that breaks gauge symmetries.

The Higgs boson found at LHC is a remnant of one particular Higgs mechanism, the one that breaks $SU(3) \times SU(2) \times U(1)$ to $SU(3) \times U(1)$. There may be several more such mechanisms operative in nature. They must work at a higher energy scale, since otherwise we would presumably have detected them already. This implies in particular that the vacuum expectation value of these new Higgs mechanism must be larger than the one of the Standard Model, about 246 GeV.

The most important property of the new Higgs scalar field we are looking for is its coupling to the gauge bosons. As for all fields, this is completely determined by its gauge group representation. Which representation should we use? This is almost a science in itself. Many papers have been written about the question which representation of a group G and which potential breaks G to a certain subgroup H. These papers usually assume the Higgs potential to be quartic, so that the theory is renormalizable. Since we do not trust the renormalizability of the Higgs system that much anyway, this requirement should perhaps not be taken too seriously. Indeed, it is quite reasonable to expect couplings of the form $\Lambda^{-2}\Phi^6$, where Λ is the scale where the coupling constant blows up. This would not be allowed in renormalizable theories because it means that we cannot make sense of the theory for momenta larger than Λ, but this we cannot do for the scalar theory anyway. If Λ is close to the Higgs mass the theory is strongly coupled, and such higher order terms in Φ may be relevant for the determination of the minimum.

There is however one criterion that is important: the vacuum expectation value of Φ is invariant under the broken gauge group, by definition of the latter. Hence the decomposition of the representation of Φ with respect to the subgroup H must contain a singlet.

Searching again through the representations of $SU(5)$ we find that the smallest representation containing a singlet is the adjoint, **24**. Its full decomposition is

$$\mathbf{24} \rightarrow (8,1,0) + (1,3,0) + (1,1,0) + (3,2,-\tfrac{5}{6}) + (3^*,2,\tfrac{5}{6}) \ . \tag{8.31}$$

This can be derived easily by computing the tensor product of a $\mathbf{5}^*$ and a $\mathbf{5}$ and subtracting a singlet. This is also the decomposition for the gauge bosons, and we recognize the first three representations as those of the $SU(3)$, $SU(2)$ and $U(1)$ gauge bosons.

The rest is straightforward. We couple this Higgs scalar to the gauge bosons in the usual way. We cannot couple them to the fermions, because one cannot build a singlet out of $\mathbf{5}^*$, $\mathbf{10}$ and $\mathbf{24}$. The scalar gets a vacuum expectation value that breaks $SU(5)$ to $SU(3) \times SU(2) \times U(1)$ and that gives a mass to the 12 unwanted gauge bosons. They eat 12 of the Higgs, and the other 12 become massive. These massive Higgs bosons are of little interest since they do not couple to the fermions.

8.8 Baryon Number Violation

The 12 massive vector bosons are more important. They couple to the fermions via the $SU(5)$ generators that transform quarks into leptons. We see from Eq. (8.31) that they belong to the representation $(3,2,-\frac{5}{6})+(3^*,2,\frac{5}{6})$. They belong thus to an $SU(2)$ doublet, which we can decompose into two $SU(3) \times U(1)$ components. These components are massive vector bosons usually called X and Y. They are color triplets and have charges $\pm\frac{4}{3}$ and $\pm\frac{1}{3}$ respectively. Their coupling to fermions follows straightforwardly from the minimal couplings in the $SU(5)$ Lagrangian. They appear in these couplings as

$$X^1_{\mu,i}T^1(i,4) + X^2_{\mu,i}T^2(i,4) + Y^1_{\mu,i}T^1(i,5) + Y^2_{\mu,i}T^2(i,5) \ , \tag{8.32}$$

where i is the color index and

$$\begin{aligned} T^1(i,j)_{kl} &= \tfrac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \\ T^2(i,j)_{kl} &= \tfrac{1}{2}i(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}) \end{aligned}$$

These matrices are thus like $\frac{1}{2}\sigma_1$ and $\frac{1}{2}\sigma_2$. Just like one does for the W-bosons, we now go to the charge eigenstates $X^\pm = \frac{1}{\sqrt{2}}(X^1 \mp iX^2)$ and analogous for Y (the upper index $\pm$ refers only to the sign of the charge).

The full set of $SU(5)$ gauge bosons can in fact be represented as a matrix $G = A^aT^a$, where T^a is a matrix in the representation **5** and only the group structure is indicated; all space-time indices are suppressed. The group structure of the minimal coupling to the field Ψ is then $(\overline{\Psi})(-G^T)\Psi$, because $-G^T = -G^*$ is the matrix representing G in the $\mathbf{5}^*$. The representation **10** is the anti-symmetric tensor product of two **5**'s. If we label the field Δ_{mn} ($m,n = 1,\ldots 5$), then the group structure of the couplings to Δ is

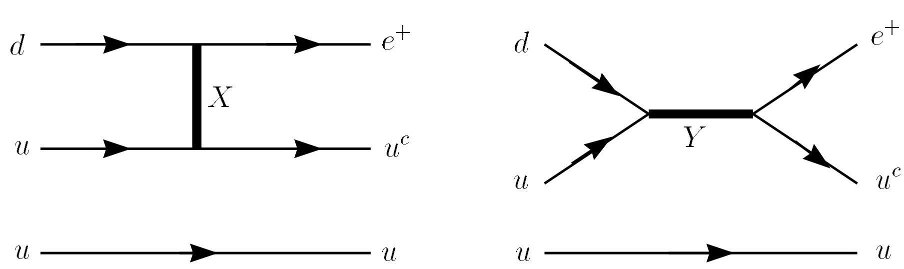
$$\bar{\Delta}_{mn}\left[G_{mk}\delta_{nl} + G_{nl}\delta_{mk}\right]\Delta_{kl} \ , \tag{8.33}$$

which can be written as $-2Tr\bar{\Delta}G\Delta$.

For one family the complete result is (the first terms in each line are derived from the coupling to the $\mathbf{5}^*$, the last from the coupling to the **10**).

$$\begin{aligned} \mathcal{L}_X &= \tfrac{g_5}{\sqrt{2}}X^-_\mu[\bar{e}^-\gamma_\mu d^c + \bar{d}\gamma_\mu e^+ - \bar{u}^c\gamma_\mu u] + \text{c.c} \\ \mathcal{L}_Y &= \tfrac{g_5}{\sqrt{2}}Y^-_\mu[\ \bar{\nu}\gamma_\mu d^c - \bar{u}\gamma_\mu e^+ - \bar{u}^c\gamma_\mu d] + \text{c.c} \end{aligned} \tag{8.34}$$

For simplicity we have suppressed color indices. They are contracted as follows for the X-boson couplings: $X_i d^c_i$, $X^i\bar{d}_i$, $\epsilon_{ijk}X_i\bar{u}^c_j u_k$ and analogously for the Y boson couplings. As expected these couplings violate both baryon number and lepton number. Diagrams for processes leading to proton decay are easy to construct, for example



These diagrams contribute to the process $p \to e^+m$, where m is a meson, which could be for example a π or a ρ. Note that we are not sure whether in the first diagram d

really is transformed to e^+ or to another charged lepton. This we can only determine after diagonalizing the mass matrices, and we will do that in a moment. If in fact the lepton is a τ then the process is forbidden by energy conservation. But there are other processes in which the lepton is a neutrino, which are allowed irrespective of the neutrino species.

The correct way to compute the couplings between the X and Y bosons is to take into account the matrices U and V that were introduced in Eqs. (4.27) and (5.11 to diagonalize the mass matrices. Then the couplings in Eq. (8.34) are replaced by matrices in flavor space, and instead of Eq. (8.34) we get

$$\begin{aligned}\mathcal{L}_X &= \tfrac{g_5}{\sqrt{2}}X^-_\mu\left[\bar{\mathcal{E}}^-_\alpha[U^\dagger_\mathcal{E}V_\mathcal{D}]_{\alpha\beta}\gamma_\mu\mathcal{D}^c_\beta+\bar{\mathcal{D}}_\alpha[U^\dagger_\mathcal{D}V_\mathcal{E}]_{\alpha\beta}\gamma_\mu\mathcal{E}^+_\beta+\bar{\mathcal{U}}^c_\alpha[V^\dagger_\mathcal{U}U_\mathcal{U}]_{\alpha\beta}\gamma_\mu\mathcal{U}_\beta\right]\\ \mathcal{L}_Y &= \tfrac{g_5}{\sqrt{2}}Y^-_\mu\left[\bar{\mathcal{N}}_\alpha[U^\dagger_\mathcal{N}V_\mathcal{E}]_{\alpha\beta}\gamma_\mu\mathcal{E}^+_\beta-\bar{\mathcal{U}}_\alpha[U^\dagger_\mathcal{U}V_\mathcal{E}]_{\alpha\beta}\gamma_\mu\mathcal{E}^+_\beta-\bar{\mathcal{U}}^c_\alpha[V^\dagger_\mathcal{U}U_\mathcal{D}]_{\alpha\beta}\gamma_\mu\mathcal{D}_\beta\right]\end{aligned}\qquad(8.35)$$

This makes many degrees of freedom of the previously unobservable rotation matrices U and V observable. Note that the matrix U rotates left-handed quarks or leptons, whereas V rotates anti-quarks and anti-leptons. The couplings of the Z and W-bosons involved rotation matrices of the form $U^\dagger(x)U(y)$, where x and y are identical for the Z bosons. Here even for couplings involving only one quarks species (the third term), no GIM-like cancellation is possible. Note also the appearance of the neutrino mixing matrix $U_\mathcal{N}$. Previously it appeared in the "CKM" matrix for the $e-\nu$ coupling, $U^\dagger(\mathcal{N})U(\mathcal{E})$. In the absence of neutrino masses it can be set equal to $U_\mathcal{E}$, which defines ν_e as the neutrino to which e^- decays. For more about the implications of these interactions for the stability of the proton see section 8.10.

8.9 Fermion Masses

In the foregoing paragraphs the fermion masses were discussed without regard to unification. This is not correct, of course. We know that in the Standard Model the fermions get their masses from a coupling to a scalar in the representation $(1,2,\frac{1}{2})$. If $SU(5)$ is a gauge symmetry above the breaking scale, it must be an exact symmetry, and hence the Higgs scalar must be part of an $SU(5)$ multiplet. In addition the Higgs Yukawa coupling must originate from an $SU(5)$ invariant coupling, and hence the $SU(5)$ representation of the Higgs scalar must couple to those of the quarks and leptons.

Since we work in the left-handed representation, the only couplings we can write down are of the form $\psi_1^TC\psi_2\phi$, where ψ_i represents one fermion, ψ_2 another and ϕ a scalar. Our fermions are in the representations $\mathbf{5}$ and $\mathbf{10}$, and hence the product of ψ_1 and ψ_2 can be in one of the following representations

$$\begin{aligned}\mathbf{5}^*\times\mathbf{5}^* &= \mathbf{10}+\mathbf{15}\\ \mathbf{5}^*\times\mathbf{10} &= \mathbf{5}+\mathbf{45}\\ \mathbf{10}\times\mathbf{10} &= \mathbf{5}^*+\mathbf{45}^*+\mathbf{50}\end{aligned}$$

Hence the candidates are $\mathbf{5},\mathbf{10},\mathbf{15},\mathbf{45}$ and $\mathbf{50}$ (note that we can use scalars and their conjugates to build Yukawa couplings). The decompositions of these fields with respect to $SU(3) \times SU(2) \times U(1)$ are

$$\begin{array}{rcl}
\mathbf{5} & \rightarrow & (1,2,\frac{1}{2}) + (3,1,-\frac{1}{3}) \\
\mathbf{10} & \rightarrow & (1,1,1) + (3^*,1,-\frac{2}{3}) + (3,2,\frac{1}{6}) \\
\mathbf{15} & \rightarrow & (1,3,1) + (3,2,\frac{1}{6}) + (6,1,-\frac{4}{3}) \\
\mathbf{45} & \rightarrow & (1,2,\frac{1}{2}) + (3,1,-\frac{1}{3}) + (3,3,-\frac{1}{3}) + (3^*,1,\frac{4}{3}) + (3^*,2-\frac{7}{6}) \\
 & & +(6^*,1,-\frac{1}{3}) + (8,2,\frac{1}{2}) \\
\mathbf{50} & \rightarrow & (1,1,-2) + (3,1,-\frac{1}{3}) + (3^*,2,-\frac{7}{3}) + (6^*,3,-\frac{1}{3}) + (6,1,\frac{4}{3}) + (8,2,\frac{1}{2})\ .
\end{array}$$

To break $SU(2) \times U(1)$ without breaking $SU(3)$ we need a representation $(1, R, q)$ where R and q are both non-trivial. The **5**, the **45** and the **15** meet that requirement, but in the latter case the candidate Higgs scalar is a triplet of $SU(2)$, and not a doublet. In addition the **15** can only couple $\mathbf{5}^*$ to itself. The fields in the $\mathbf{5}^*$ are, in the usual Standard Model notation, d_R and $(\nu, e^-)_L$, and mass terms between any pair of these fields are undesirable except for a possible Majorana mass for the neutrino. If that's the only mass we can get, it means that the **15** is not a useful representation (not by itself, at least).

We will only discuss scalars H in the **5** in some detail. The couplings to the combination $\mathbf{5}^* + \mathbf{10}$ are (here i, j, k, l, m, n are $SU(5)$ indices, and α and β are family indices)

$$g_1^{\alpha\beta} \Psi_i^\alpha(\mathbf{5}^*) C \Delta_{kl}^\beta(\mathbf{10}) H_m^* \delta_{ik} \delta_{lm} + \text{c.c} \tag{8.36}$$

Note that the **10** is an anti-symmetric tensor product of two **5**'s, so we can represent it as a field with two vector indices, satisfying $\Delta_{ij} = -\Delta_{ji}$. Since the indices i, k (and l, m) belong to conjugate representations, they can be contracted by a Kronecker δ. For the other coupling we need the invariant tensor ϵ_{ijklm} of $SU(5)$:

$$g_2^{\alpha\beta} \Delta_{ij}^\alpha(\mathbf{10}) C \Delta_{kl}^\beta(\mathbf{10}) H_m \epsilon_{ijklm} + \text{c.c}\ . \tag{8.37}$$

The fermion bi-linear is symmetric under the exchange of the two **10**'s (the sign change coming from interchanging the two fermions is canceled since $C = -C^T$), and hence $g_2^{\alpha\beta}$ must be symmetric in α and β. To right-handed neutrinos we have to add three singlet representations of $SU(5)$. These can get a Majorana mass, and in combination with the fermions in the representation **5** and the Higgs they can get a Dirac mass. The coupling to the Higgs boson is

$$g_{\text{neutrino}}^{\alpha\beta} \Psi_i^\alpha(\mathbf{5}^*) C \psi^\beta(\mathbf{1}) H_m \delta_{im} + \text{c.c}\ , \tag{8.38}$$

where $\psi^\beta(\mathbf{1})$ is the $SU(5)$ singlet field.

Let us assume that the field H acquires a vacuum expectation value just like it does in the Standard Model. This issue will require further discussion, since in principle the field H could choose an arbitrary direction within $SU(5)$. A completely random direction would break color, but there are Higgs potentials for which that does not happen. If color is not broken, the Higgs v.e.v will choose a direction within $SU(2) \times U(1)$. This direction is in principle arbitrary, but we have already fixed it by assigning particles to the elements of the $\mathbf{5}^*$ and the **10**. This is standard practice, but conceptually not very elegant (in the discussion of the Standard Model we did not follow this practice). Hence we choose $\langle H_i^* \rangle = \frac{1}{\sqrt{2}} v \delta_i^5$. Then the two Yukawa interactions yield the following fermion bi-linears

$$\frac{1}{\sqrt{2}} g_1^{\alpha\beta} \Psi_i^\alpha(\mathbf{5}^*) C \Delta_{i5}^\beta(\mathbf{10}) v + \text{c.c} \tag{8.39}$$

and

$$\frac{1}{\sqrt{2}} g_2^{\alpha\beta} \sum_{i,j,k,l=1}^{4} \Delta_{ij}^\alpha(\mathbf{10}) C \Delta_{kl}^\beta(\mathbf{10}) \epsilon_{ijkl} v + \text{c.c} \ . \tag{8.40}$$

The former leads to the following mass terms,

$$(\mathcal{D}^c)^T C M \mathcal{D} + (\mathcal{E}^-)^T C M \mathcal{E}^+ \text{c.c} \tag{8.41}$$

After switching back to $L-R$ notation we get.

$$-\bar{\mathcal{D}}_R M_{\mathcal{D}} \mathcal{D}_L - \bar{\mathcal{E}}_R M_{\mathcal{E}} \mathcal{E}_L + \text{c.c} \ , \tag{8.42}$$

where

$$M = M_{\mathcal{D}} = M_{\mathcal{E}}^\dagger = \frac{v}{\sqrt{2}} g_1 \ . \tag{8.43}$$

The Hermitean conjugate on $M_{\mathcal{E}}$ is due to the fact that the second term in Eq. (8.41 is not of the form $\psi^c M \psi$, where ψ is a particle and ψ^c the antiparticle spinor. Hence the lepton mass terms in Eq. (8.42) are obtained from the terms labeled "c.c" in Eq. (8.41) (*c.f.* Eq. (5.10)). This finds its origin in the fact that the $\mathbf{5}^*$ contains the anti down quark and the electron.

The second Yukawa coupling is just slightly more difficult to analyze. Note that because of the ϵ tensor only the first four components of the $\mathbf{10}$, the u-quarks, contribute. There are $4! = 24$ terms, divided over three colors, so that for each color the multiplicity is 8. Together with the normalization factor of the $\mathbf{10}$ and an overall $-$ sign we get then

$$-\bar{\psi}_{\mathcal{U}} M_{\mathcal{U}} \psi_{\mathcal{U}} \ , \tag{8.44}$$

where

$$M_{\mathcal{U}} = -2\sqrt{2} g_2 \ v \ , \tag{8.45}$$

which is a symmetric matrix.

We find thus a relation among the mass matrices for the leptons and the down quarks, whereas the up quarks have their own independent mass matrix. The mass relation implies in particular that the eigenvalues are the same, and that the diagonalization matrices are the same, so that the aforementioned problem of deciding which lepton belongs to which family does not occur: we have to order them according to increasing mass. Then $SU(5)$ with the set of Higgs bosons chosen here implies the following mass relations

$$\begin{aligned} m_d &= m_e \\ m_s &= m_\mu \end{aligned}$$

$$m_b = m_\tau$$

At first sight that does not look like a great success, but we have to remember that these relations hold at $M_{\rm GUT}$. Just like the coupling constants we have to extrapolate them to lower energies. Comparison with experimental data is not straightforward, since we do not measure the quark masses directly, and since in addition the required extrapolation for the d and s quarks is to mass scales that are much too low. With the τ-mass as input, the predicted value for m_b is somewhere between 5 and 7 GeV (depending on various assumptions), to be compared with the mass of the lowest $\bar{b}b$ bound state, the Y, 9.46 GeV. For the other relations it safer to compare the ratios m_d/m_s and m_e/m_μ, under the assumption that at least some of the unknown QCD effects cancel. The agreement is nevertheless not good, the discrepancy being almost a factor 10. It is noteworthy that in GUTs originating from string theory (in particular from heterotic strings) the relations between the gauge couplings are preserved, but that the bad predictions for the fermion masses do not hold.

Now let us consider briefly the effects of the other candidate Higgs boson, the **45**. Now the mass relations are $M_{\mathcal{E}}^\dagger = -3M_{\mathcal{D}}$, leading in particular to $m_\tau = 3m_b$, which is certainly not an improvement. For the u quarks the result is much worse. The coupling to the **45** yields an anti-symmetric mass matrix. This is a bad feature, because the eigenvalues of such a matrix come in pairs with opposite sign. The signs do not matter for fermion mass terms $m\bar{\psi}_R\psi_L$, because we may flip the sign of ψ_R without altering the kinetic terms. But then we get two masses that are equal, and a third one that is necessarily zero. Clearly $m_u = 0, m_c = m_t$ does not fit the quark masses very well.

If we choose a combination of a one **5** and one **45** (or more of each), the mass matrix M_U becomes an arbitrary complex matrix, and the mass matrices for the leptons and the down quarks read $M_{\mathcal{E}} = M(\mathbf{5})^\dagger + M(\mathbf{45})^\dagger$, $M_{\mathcal{D}} = M(\mathbf{5}) - 3M(\mathbf{45})$, which implies that they are two independent matrices. This means that we have lost all predictive power. In this case all mixing angles between the quarks and leptons are in principle non-trivial.

If only the **5** is used, the lepton and down quark masses can be diagonalized by the same matrices. This is most easily seen by considering Eq. (8.41)

$$\begin{aligned} M_{\mathcal{E}}(\text{diag}) &= U_{\mathcal{E}}^T M V_{\mathcal{E}} \\ M_{\mathcal{D}}(\text{diag}) &= V_{\mathcal{D}}^T M U_{\mathcal{D}} \; . \end{aligned}$$

We are clearly allowed to choose

$$U_{\mathcal{E}} = U_{\mathcal{D}} \quad \text{and} \quad V_{\mathcal{E}} = V_{\mathcal{D}} \; . \tag{8.46}$$

As was explained before when we discussed the CKM matrix, the matrices U and V are not determined uniquely by the requirement that the mass matrices be diagonal. We can replace V by VP and U by UP^*, where P is diagonal and unitary. This matrix P can be chosen differently for $\mathcal{D}$ and $\mathcal{E}$. Such phase rotations of the matrices $U_{\mathcal{U}}$ and $U_{\mathcal{D}}$ were used in the weak interactions to bring the CKM matrix to a definite form. This implies that the phases for the $\mathcal{U}$ sector and the $\mathcal{D}$ sector are already fixed, but the ones in the

$\mathcal{E}$ sector are free. It is furthermore not hard to show that the symmetry of $M_{\mathcal{U}}$ implies $V_{\mathcal{U}}^{\dagger}U_{\mathcal{U}}$ is diagonal (though in general not $\mathbf{1}$). The remaining two couplings, those of Y boson to ue and to ud have non-trivial mixing.

8.10 Proton Decay

8.10.1 B-L

In the $SU(5)$ theory baryon number is violated explicitly by combining quarks and leptons in a single multiplet; the same is obviously true for lepton number. There still is an unbroken global symmetry though, namely $B-L$. This symmetry is not very manifest in $SU(5)$, and one sees it in the following way. The gauge group $SU(5)$ couples to fermions in the representation $3(\mathbf{5}^*+\mathbf{10})$. There are six representations, and hence one can define six independent global symmetries. One linear combination is broken by anomalies with respect to $SU(5)$. Four linear combinations can be taken as differences between families. When the quarks get masses and mix, the only symmetry that remains of these is the relative lepton numbers. The last anomaly free symmetry is taken to be the same for all families, with charge $Q=-3$ for the $\mathbf{5}^*$, and $Q=1$ for the $\mathbf{10}$. This is anomaly-free with respect to $SU(5)$ since $I_2(\mathbf{5})=I_2(\mathbf{5}^*)=1$ and $I_2(\mathbf{10})=3$. To make its only cubic anomaly cancel, we need to add one Standard Model singlet per family, with Q charge 5. The cubic anomaly is then $5\times(-3)^3+10\times 1+125=0$, and the trace is $5\times(-3)+10\times 1+5$, which also vanishes. This is not $B-L$, however, since it assigns the same charge to all the members of the $\mathbf{5}$, d^c as well as e^- and ν. To get $B-L$ one has to combine it with Y, the weak hypercharge, which is an $SU(5)$ generator. Then one finds that $B-L=\frac{4}{5}Y+\frac{1}{5}Q$. Note that this is only an anomaly free symmetry if there are three Standard Model singlets with $B-L$ equal to 1, *i.e.* three particles with the quantum numbers of left-handed anti-neutrinos (or equivalently, right-handed neutrinos). Note also that a Majorana mass term is forbidden by $B-L$, and hence if there exists such a mass term in nature, this is a source of $B-L$ violation (Dirac mass terms respect $B-L$). However, if we simply drop the right-handed neutrino altogether, the only $B-L$ violation is via the cubic anomaly, and hence it is at best a non-perturbative effect, that cannot be seen in Feynman diagrams.

This shows that $B-L$ is a symmetry of the fermion minimal couplings. How about the Yukawa couplings? Under Q the combination $\mathbf{10}\times\mathbf{10}$ has charge 2, and $\mathbf{5}^*\times\mathbf{10}$ has charge -2. Hence Q and $B-L$ are preserved if we assign charge $Q=2$ to the Higgs. In $SU(3)\times SU(2)\times U(1)$ components this implies that the Higgs representation $(1,2,-\frac{1}{2})$, the Standard Model Higgs, has $B-L=0$, as it should. Note that the $SU(5)$ scalar that contains the Standard Model Higgs also contains an $SU(3)$ triplet. This triplet has $B-L=\frac{2}{3}$. For the gauge bosons $Q=0$, but since $B-L=\frac{4}{5}Y+\frac{1}{5}Q$ the X and Y bosons have a non-vanishing $B-L$ charge, equal to $-\frac{2}{3}$ for both.

The decay channels of the proton are restricted by $B-L$ because decays to for example 3 leptons are forbidden. The final state must thus necessarily contain a positron, a μ^+ or an anti-neutrino, and in addition there can be any number of lepton anti-lepton pairs.The main decay modes, if one disregards family mixing, are $e^+\pi, e^+\rho, e^+\eta, e^+\omega$, $\nu^c\pi^+, \nu^c\rho^+, \mu^+K^0$, etc.

8.10.2 The Proton Lifetime

The main technical complication in the computation of the proton decay width is that the initial and final states are not quarks, but hadrons and mesons. Therefore the result depends for example on the model used to describe the proton. Naively the estimate for the proton decay width would be

$$\Gamma = \left(\frac{g_5}{M_{\rm X}}\right)^4 C|\psi(0)|^2(E_{qq})^2 \ , \tag{8.47}$$

where E_{qq} is the energy of the quark-quark pair involved in the interaction, $\psi(\vec{r})$ is the wave function of the quark-quark pair (in an approximation where the third quark is ignored), $M_{\rm X}$ the mass of the X and Y bosons and C a dimensionless numerical constant, containing for example phase space, color and spin factors.

The wave function of the qq system at $\vec{r} = 0$ (where $\vec{r}$ is the spatial separation) is the amplitude for finding both quarks at the same place. Note that the unification scale is extremely high, so that to a very good approximation the interaction is point-like. In fact the first step in the computation is usually to "integrate out" the heavy X and Y bosons, so that their effect is described by an effective four fermion interaction, consisting of terms of the form $Ga\bar{\psi}_1\gamma_\mu\psi_2\bar{\psi}_3\gamma_\mu\psi_4$, where $G = \frac{1}{8}\sqrt{2}(g_5/M_{\rm X})^2$, a a numerical constant and ψ_i one of the quarks or leptons. This effective interaction is completely analogous to the usual fermi model for the weak interactions, with G playing the role of the Fermi constant. It can be obtained directly from Eqs. (8.34) or (8.35), depending on whether or not one wishes to ignore mixing effects.

It is not hard to get a rough estimate of the proton decay width. Clearly $\psi(0)$ which has dimension $\frac{3}{2}$ and is entirely determined by QCD, must be of order $(m_p)^{\frac{3}{2}}$, where m_p is the proton mass. The quark-quark center-of-mass energy E_{qq} must be about twice the effective quark mass (the constituent mass), which is roughly $\frac{1}{3}m_p$. If we assume that all numerical factors are 1, we get $\Gamma \approx (g_5/M_{\rm X})^4(m_p)^5$. Both g_5 and $M_{\rm X}$ can be computed directly from the evolution of the Standard Model coupling constants (for g_5 one uses $g_5(M_{\rm X})$). The result is about 10^{38} years, using the data on coupling constants of 1978, and was first computed in [4]. At that time the limit on the lifetime of the proton was about 10^{30} years, eight order of magnitude below this very naive theoretical prediction.

The main source of error in this prediction turns out to be the value of $M_{\rm X}$. This mass is predicted on the basis of a plot that is logarithmic in M. Small changes in $\log M$ can lead to very big changes in M.

8.10.3 Historical Remarks

In [4] a value for $M_{\rm X}$ of 3.7×10^{16} was computed. Several effects were ignored, such as the running of α up to $M_{\rm w}$, thresholds for heavy quarks, two-loop effects, and the contribution of the Higgs boson. It turned out that all these effects go in the same direction, and reduce $M_{\rm X}$ by a factor 100, and hence the proton lifetime by a factor 10^8! This still only uses a rather primitive treatment of the thresholds, namely a discontinuous change of the slope as in fig (6).

A second important class of corrections are the $SU(3) \times SU(2) \times U(1)$ loop effects on the effective four-fermi interaction, due to gauge boson exchanges between the four external legs. These can enhance the decay by factors of about 5 for gluon exchange and 2 for W, Z and photon exchange.

Another technical difficulty is the correct treatment of the proton structure. Various models for hadrons have been used, such as the bag model.

Not surprisingly, the final answer is subject to a large amount of uncertainty, and is about $10^{31\pm2}$ years. This range of values is however by now ruled out by experiment.

All of this is based on the "minimal" $SU(5)$ model. The simplest way of making the predictions for proton decay in agreement again with the experimental lower bound is to increase $M_{\rm X}$. Within the $SU(5)$ model that can be done by adding extra matter to the desert. We have seen the beginning of the chapter that simply adding a full $SU(5)$ multiplet is not going to change $M_{\rm X}$. It will only increase g_5 at $M_{\rm X}$, thus making the decay width larger instead of smaller. The only way to increase $M_{\rm X}$ without giving up $SU(5)$ altogether is to add broken $SU(5)$ multiplets. The arguments in section 8.1 assume that all particles in a multiplet contribute to the coupling constant evolution. If some are heavier than others, they will decouple, and then it is possible to change $M_{\rm X}$. The chiral fermions forming a family form an unbroken $SU(5)$ multiplet, and hence to first approximation their presence does not influence $M_{\rm X}$ (if one looks more carefully the mass splittings introduced by the weak interactions do have some effect on the evolution below $M_{\rm W}$). The standard model Higgs does have an effect, since one must assume that its triplet component is heavy (see below). Hence convergence and the value of $M_{\rm X}$ are sensitive to the number of Higgs scalars. Another set of fields that have an important influence turn out to be the gauginos in supersymmetric theories.

8.11 The Higgs System

In every broken gauge theory, the Higgs system is usually the most problematic part. GUTs are no exception. An additional complication is that there are two Higgses that enter the discussion, the Higgs field Φ in the **24** of $SU(5)$, and the Standard Model Higgs field H in the representation **5** of $SU(5)$.

To first approximation one may ignore H and consider only Φ. Just as was the case for the Standard Model Higgs, we can rotate the field in a convenient direction using the gauge transformation. The Higgs field Φ is a complex, Hermitean, 5×5 traceless matrix field (note that this does indeed have 24 degrees of freedom) that transforms under an $SU(5)$ transformation U (a unitary 5×5 matrix with determinant 1) as

$$\Phi \rightarrow U^{\dagger} \Phi U \qquad (8.48)$$

We can use the gauge freedom U to diagonalize Φ. Since it traceless, this leaves 4 parameters. For arbitrary choices of these parameters, it would break $SU(5)$ to $U(1)^4$. However, it turns out that this cannot happen. A single adjoint Higgs can break the group $G = SU(N)$ only to a so-called "maximal subgroup", which is a subgroup $H \subset G$ such that there is no intermediate group H' with $H \subset H' \subset G$, $H' \neq H$ and $H' \neq G$. With multiple adjoint Higgs field on can realize chains of symmetry breaking to smaller groups. To understand why this is true requires a more detailed study of Higgs potentials. We will just take it here as a fact. This implies that the vacuum expectation value of Φ can either be

$$\langle\Phi\rangle = \text{diag }(v, v, v, v, -4v) \tag{8.49}$$

or

$$\langle\Phi\rangle = \text{diag }(v, v, v, -\tfrac{3}{2}v, -\tfrac{3}{2}v)\ , \tag{8.50}$$

where v depends on the parameters in the potential. The first v.e.v. breaks $SU(5)$ to $SU(4) \times U(1)$, the second to $SU(3) \times SU(2) \times U(1)$. If we allow Φ to have more distinct eigenvalues we would get a subgroup of these groups. Note that there is an important difference with the Standard Model Higgs mechanism: not all possible Higgs vevs are gauge equivalent.

The obvious problem with the combined **24** and **5** Higgs is the hierarchy problem: Why does one of them get a vacuum expectation value so much smaller than the other?

But there is a second problem. The Higgs is a **5** of $SU(5)$, and in addition to the Standard Model Higgs boson this representation contains a color triplet scalar. This particle couples to quarks and leptons via the Yukawa couplings, and it is not hard to see that it can mediate proton decay. Therefore its mass must be of the order of $M_{\rm X}$. On the other hand its partner, the $SU(2)$ doublet, must get a mass of the order of the weak scale.

This is all possible, but in a very unsatisfactory way. To examine it more closely we consider the complete Higgs potential for H and Φ. If we go to quartic order and impose (for simplicity) the discrete symmetry $\Phi \to -\Phi$, the most general Higgs potential is

$$\begin{aligned} V(\Phi, H) \ &= \ -(\mu_5)^2 H^\dagger H + \tfrac{1}{4}\lambda(H^\dagger H)^2 - \tfrac{1}{2}\mu^2 \ \text{Tr}\Phi^2 + \tfrac{1}{4}a(\ \text{Tr}\Phi^2)^2 + \tfrac{1}{2}b\ \text{Tr}\Phi^4 \\ &+ \ \alpha H^\dagger H\ \text{Tr}\Phi^2 + \beta H^\dagger \Phi^2 H\ . \end{aligned}$$

For suitable parameter choices, a minimum of this potential is

$$\langle\Phi\rangle = \text{diag }(v, v, v, (-\tfrac{3}{2} - \tfrac{1}{2}\epsilon)v, (-\tfrac{3}{2} + \tfrac{1}{2}\epsilon)v); \qquad \langle H \rangle = \frac{1}{\sqrt{2}}(0, 0, 0, 0, v_0)^T\ . \tag{8.51}$$

The ϵ terms are induced by $SU(2)_{\rm w}$ breaking, and are slightly worrisome. The **24** decomposes as in Eq. (8.31). To break $SU(5)$ to $SU(3) \times SU(2) \times U(1)$ only the singlet component should get a v.e.v, as is the case in Eq. (8.49). The ϵ terms indicate that also the $SU(2)$ triplet component gets a v.e.v (all other components have non-trivial color, and the vacuum we consider here does not break color). This is undesirable, since they would contribute to the ρ-parameter. However, this problem at least takes care of itself, since it turns out that $\epsilon \propto \frac{v_0^2}{v^2}$. Since we clearly want $v_0 \ll v$ we see that $\epsilon v \ll v_0$, and hence the **24** gives a negligible contribution to $SU(2)_{\rm w}$ breaking in comparison to the **5**.

For any sensible choice of the parameters in the potential, the mixed $\Phi - H$ terms will induce a mass-term for H, or rather for the $SU(3)\times SU(2)\times U(1)$-components of H. The mass of the doublet component of H is, ignoring ϵ, equal to $-\mu_5^2 + \frac{15}{2}\alpha v^2 + \frac{9}{2}\beta v^2$. Here v is of order M_{X} and α and β can be expected to be of order 1, while the sum must be of order M_{W}. This is only possible if the two last terms almost cancel each other. There is no symmetry that can achieve this, and thus it requires a fine-tuning of α/β with a precision of about 25 digits. Once this has been achieved we do not have to worry much about the color triplet Higgs. Its mass is given by another combination of α and β, and the natural value of its mass is roughly M_{X}. So the problem with the $SU(5)$ Higgs sector is in the end just the naturalness problem of the Standard Model. The only difference is that it shows up in a more concrete way.

8.12 Magnetic Monopoles

The Maxwell equation $\vec{\nabla}\cdot\vec{B} = 0$ implies that there cannot be sources for the magnetic field $\vec{B}$. In other words, in classical electrodynamics there are no magnetic monopoles.

The reason magnetic sources cannot be added can be traced back to the formulation in terms of vector potentials. If one writes the Maxwell equations covariantly one gets

$$\begin{aligned} \partial_\mu F^{\mu\nu} &= J^\nu \\ \partial_\mu \tilde{F}^{\mu\nu} &= 0 \ . \end{aligned}$$

where $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$. The absence of magnetic monopoles follows from the zero on the left-hand side of the second equation. If we would put there a "magnetic current" the Maxwell equations have a magnetic-electric duality: they would be symmetric under the exchange of electric and magnetic quantities (fields and currents). As long as we just work with the field strength tensor $F_{\mu\nu}$ (or equivalently $\vec{E}$ and $\vec{B}$), we could simply add a magnetic current. However, the second equation is a trivial consequence of the expression for $F_{\mu\nu}$ in terms of the vector potential A_μ, and we need the description in terms of A_μ in order to quantize the theory.

One can make field configurations that *resemble* magnetic monopoles. Dirac observed that one can consider an infinitely long, infinitesimally thin solenoid with one pole at a point in space ($x = 0$) and the other at infinity. In such a configuration $\vec{\nabla}\cdot\vec{B} = 0$ is not violated, since the apparent outgoing magnetic flux at $x = 0$ is compensated by magnetic flux lines through the narrow solenoid, in such a way that all magnetic flux lines are closed loops through $x = 0$ and a point at infinity.

Classically the solenoid is infinitesimal and cannot be observed, but quantum mechanically one see flux going through a loop via the Aharonov-Bohm effect. The wave function of a charged particle changes by a phase when it moves in a non-trivial vector potential

$$\psi' = \psi\; e^{i\frac{e}{\hbar}\int d\vec{s}\cdot\vec{A}} \tag{8.52}$$

Although the magnetic field is present only inside the solenoid, the field A is non-zero outside, and can be detected by means of a charged particle. In particular we may carry the charge once around the solenoid and bring it back to the same point, so that it forms a closed loop. Note that the vector potential points in the tangential direction along circles around the solenoid, and falls of as $1/r$, where r is the distance to the solenoid. Using Stokes' theorem we can then convert the loop integral of $\vec{A}$ to a surface integral of $\vec{B}$:

$$\int d\vec{s} \cdot \vec{A} = \int dS \; \vec{n} \cdot B = \Phi \; , \tag{8.53}$$

where $\vec{n}$ is the normal vector of the surface and Φ the flux through the surface. Note that the left-hand side has the same value for any circle around the solenoid: the circumference of the circle increases with r, but the vector potential decreases with r.

This seems to lead to the conclusion that an infinitesimal solenoid can always be detected by means of finite size loops of charged particles. To express it in a more physical way, one may do interference experiments with particles going from point 1 to point 2 via different paths. If the loop formed by two such paths encloses the solenoid the interference pattern will change.

Note however that the change of the wave function is only by a phase. Hence the solenoid would still be unobservable if the phase equals 1, or

$$\Phi = n \frac{\hbar}{e} 2\pi \; , \tag{8.54}$$

where n is an integer. In the case under consideration Φ is the flux through the solenoid. Since there is no net magnetic flux escaping from any infinitesimal sphere around $x = 0$, Φ must be equal to minus the magnetic "monopole" flux that appears to emerge at $x = 0$. This in its turn is proportional to the magnetic monopole charge that one would define if indeed the end of the solenoid were a monopole. By analogy with the equation $\int_{S_\infty} dS \; \vec{n} \cdot \vec{E} = 4\pi e$ for the total electric flux from an electric point charge e, we define

$$\int_{S'_\infty} dS \; \vec{n} \cdot \vec{B} = 4\pi g \; . \tag{8.55}$$

Here S_∞ denotes a sphere at infinity, and the prime indicates that we omit the contribution of the infinitesimal solenoid, which precisely cancels the monopole flux. Substituting this into Eq. (8.54) we get the famous Dirac quantization condition for magnetic charges

$$eg = \tfrac{1}{2} n \hbar \; . \tag{8.56}$$

This result is a necessary condition for the existence of magnetic monopoles in a theory. It implies that if the theory contains particles with electric charge e, then if any monopoles exist they must have magnetic charges that are a multiple of $\hbar/2e$.

This result can be made more precise by showing that any field configuration with an asymptotic behavior such that the magnetic field has a monopole component can only be obtained from a vector potential that is not regular everywhere on the sphere. The singularities form a string as a function of the radius of the sphere, but if (and only if) the Dirac quantization condition is satisfied this singularity has no observable consequences.

Note that the Dirac quantization condition was derived by using a charge e particle, where e is the electron or proton charge. If (unconfined) particles of charge e/m exist in nature (where m is some integer), one could use them instead of an electron to detect Dirac strings, and hence the minimum magnetic charge would be a factor m larger. Conversely, if a magnetic monopole were found whose magnetic charge is precisely $1/2e$ (from now on $\hbar = 1$), we would know that all electric charges have to be a multiple of the electron charge.

In pure electrodynamics one cannot reasonably expect magnetic monopoles to exist, for two (not unrelated) reasons. First of all the Dirac field configuration just described not only has a string singularity, but also a singularity in the field strength at $r = 0$. Hence the magnetic field energy, which is part of the mass of the object, is infinite. One cannot really resolve this singularity without discovering that one is looking at the end of a solenoid and not at a monopole. With electric charges there is no such problem. Secondly, nothing in the theory forbids us *a priori* to add particles of arbitrary charge: there is no charge quantization mechanism. Hence one would expect the minimal value of g to be infinite.

It was realized by 't Hooft and Polyakov that these problems could be overcome if the electromagnetic gauge group was embedded in a non-abelian group. The canonical example is $U(1) \subset SO(3)$, with $U(1)_{\rm em} = T_3$. In this case there is a fundamental reason for charge quantization, since the representations of $SO(3)$ only allow integer eigenvalues for T_3.

To see how the singularity problem is solved we have to consider how $SO(3)$ is broken. One uses a Higgs ϕ in the triplet representation (the adjoint representation), which develops a v.e.v. which can be rotated to the form $\langle\phi\rangle = (0, 0, v)$. The surviving gauge group is $SO(2)$. However, the direction of the Higgs in group space is not relevant. We could choose any direction we want, and even choose different directions in different space-time points. The trick is now to make $\langle\phi\rangle$ point in the radial direction $\hat{r}$ for large r. Locally each asymptotic observer measures the same physical phenomena as with a fixed vacuum direction, but globally the configuration is different. In fact, no continuous transformation will bring it back to a fixed direction, and one says that configuration is "topologically non-trivial". Note that we are making an identification between two *a priori* unrelated groups, the $SO(3)$ gauge group and the $SO(3)$ rotation group.

This can be done asymptotically, but one cannot continue the Higgs field to $r = 0$ without encountering a singularity. This can be solved by choosing the Higgs field as

$$\langle\phi_a(\vec{r})\rangle = f(r) v \hat{r}_a \ , \tag{8.57}$$

where $f(0) = 0$ and $f(\infty) = 1$. In other words, the Higgs v.e.v. goes to zero at $r = 0$. We are not allowed to set $\phi = 0$ over all of space-time, since that would cost an infinite amount of energy, but we can do it in a finite space-time region at finite energy cost.

If one substitutes the above ansatz for the Higgs field into the equations of motion, one finds that for large r,

$$\partial_i \langle\phi_a\rangle = \frac{v}{r}(\delta_{ai} - \hat{r}_i \hat{r}_a) \tag{8.58}$$

hence $\int d^3x(\partial_i\phi_a)^2$ diverges. Here i is a space index, and a an $SO(3)$ vector index. These are thus indices of isomorphic representations. To get a configuration of finite energy, one can must make use of the coupling to the electromagnetic field, which modifies ∂_i to D_i. To get $D_i\phi$ to fall off sufficiently rapidly, we need (upper and lower indices are used here merely as a notational convenience, and have no special significance)

$$A_i^b(\vec{r}) = \epsilon_{ibj}\frac{\hat{r}_j}{er}[1 - K(r)] \ . \tag{8.59}$$

To get the proper asymptotic behavior, we need $K_\infty = 0$; to avoid a singularity at $r = 0$ we need $K(0) = 1$. Consider first the large r behavior. We would like to demonstrate that $\int d^3x(D_i\phi_a)^2$ falls off sufficiently fast at large r, unlike $\int d^3x(\partial_i\phi_a)^2$. Consider

$$D_i\phi_a = \left(\partial_i - ieA_i^bT^b\right)\phi_a \tag{8.60}$$

The gauge contribution to the covariant derivative is:

$$\left(-ieA_i^bT^b\right)\phi_a = -ie\epsilon_{ibj}\left(\frac{\hat{r}_j}{er}\right)(-i\epsilon_{bac})\phi_c \tag{8.61}$$

Here we used the fact that T^b are representation matrices in the adjoint representation, hence $T^b_{ac} = -i\epsilon_{bad}$. Substituting the asymptotic vacuum expectation value for ϕ_c, $v\hat{r}_c$, and performing the implicit sum over b we find in the large r limit

$$\left(-ieA_i^bT^b\right)\langle\phi_a\rangle = -\frac{v}{r}(\delta_{ai} - \hat{r}_i\hat{r}_a) \tag{8.62}$$

This precisely cancels the derivative contribution, Eq. (8.58). If we take into account the r dependence in $f(r)$ and $K(r)$ one gets small corrections that fall off sufficiently rapidly to keep the space integral finite. By making $f(r)$ go to zero at the origin, and $K(r)$ go to 1, one can ensure that the integral near the origin is finite as well. One may substitute Eqs. (8.57) and 8.59) into the equations of motion, and obtain a set of coupled differential equations for the functions f and K. These can only be solved numerically, and in some cases even analytically.

We have now obtained a non-trivial solution to the equations of motion with finite energy: $\int d^3x\mathcal{H}$ is finite, where $\mathcal{H}$ is the Hamiltonian density, defined in the usual way

$$\mathcal{H} = \pi_a(x)\dot{\phi}_a(x) - \mathcal{L} \tag{8.63}$$

The ϕ-dependent part of $\mathcal{H}$ is, for time-independent fields

$$\mathcal{H}_\phi = (D_i\phi_a)^2 + V(\phi) \tag{8.64}$$

In addition to this term there is an electromagnetic field energy, which is proportional to $\vec{E}^2 + \vec{B}^2$ for large r, and which has additional non-abelian contributions for small r. We will see in a moment that this contribution is also finite.

If we set the energy of the vacuum (corresponding to vanishing ϕ_a and A_i^b) to zero, then the "field configuration" we have obtained has a non-vanishing energy density localized around a point in space (namely $\vec{r} = 0$) with a finite total energy. Note that this breaks translation invariance. There is nothing wrong with that, it can simply be interpreted as an "object" localized at $\vec{r} = 0$. One may of course find a completely analogous solution to the equations of motion localized at other points, and one can find time-dependent solutions where these "objects" are moving as free particles. By studying their kinematics, one observes that they behave as particles, with a mass given by $\int d^3x\mathcal{H}$. Such solutions can be found in many classical field theories, and are generically called "solitons". So what we have found here is that the $SO(3)$ gauge-Higgs system has a soliton solution. In the quantum theory these give rise to new particles, in addition to the usual ones created from the vacuum by the quantum fields.

It is easy to check that the field strength $F_{\mu\nu}^a$ derived from the non-abelian field configuration Eq. (8.59) has asymptotically non-vanishing components only in the direction of the unbroken $U(1)_{\text{em}}$. Furthermore this field configuration looks asymptotically like the one of a magnetic monopole with magnetic charge $1/e$,

$$B_\infty^i = \frac{\hat{r}^i}{er^2} \ . \tag{8.65}$$

This implies in particular that for large r the integral $\int d^3x\vec{B}^2$ is finite.

To any distant observer this object would look like a magnetic monopole. Note that the vector potential Eq. (8.59) has no string singularity. It was avoided by making use of the embedding in $SO(3)$, which allowed us to make the vacuum point radially. By making a (singular) gauge transformation we can make the vacuum point in one direction only, but then inevitably a string singularity is introduced for the gauge field.

The monopole strength is twice the minimal Dirac value $1/2e$. The reason is that we may add fields in spinor representations of $SO(3)$ (so that the global gauge group becomes $SU(2)$ instead of $SO(3)$) whose charges are half-integer in comparison to the $SO(3)$ charges. These half-integral charges would be in conflict with a monopole of charge $1/2e$. Since the spinor representations are not involved in the classical field configuration, it is clear that the classical solutions in the $SU(2)$ theory are the same as in $SO(3)$. Hence the $SO(3)$ solutions already "anticipate" the possibility of half-integer charges.

The energy density of the object is localized around $r = 0$, and falls of exponentially for $r \to \infty$. This exponential fall-off can be used to define the size R of the object: $\mathcal{H} \propto e^{-r/R}$. The size is set by the only scale in the problem, the Higgs vev v. By computing the three-dimensional space integral of the energy density one obtains the energy of the field configuration, or the mass of the object. The result can be written as

$$M_{\text{mon.}} = \frac{4\pi}{e^2} M_{\text{W}} \xi(\lambda, e^2) \ , \tag{8.66}$$

where M_{W} is the mass of the massive vector bosons that are the result of the spontaneous symmetry breaking $SU(2) \to U(1)$. Hence $M_W \propto ev$. The function ξ depends on the gauge coupling e (in this case the $SU(2)$ gauge coupling and the canonically normalized

$U(1)$ gauge coupling are identical) and the Higgs quartic self-coupling λ. For simplicity we assume here a simple Higgs potential $V(\phi)$ with a quartic term $\lambda(\phi_a\phi_a)^2$. For $\lambda = 0$ we get $\xi = 1$, and in this limit (the Prasad-Sommerfield or Bogomol'nyi limit) the equations of motion can be solved analytically. The function ξ increases monotonically with λ, but reaches a finite value (≈ 1.7867) for $\lambda \to \infty$.

This kind of monopole solution also exists in $SU(5)$ grand unified theories, since they also have the properties that electric charges are automatically quantized. One can construct spherically symmetric solutions within suitable $SU(2)$ subgroups of $SU(5)$, which must include $U(1)_{\rm em}$. There are three spherically symmetric solutions with magnetic charges $1/2e$, $1/e$ and $3/2e$, *i.e.* once, twice and three times the Dirac charge. Their classical masses are

$$M_q = \frac{3}{8}\frac{M_{\rm X}}{\alpha} q\xi(\lambda_i, g_5, q) \ , \tag{8.67}$$

where q is the magnetic charge in units of $1/2e$, and α the fine structure constant $\frac{e^2}{4\pi}$. The factor $3/8$ is due to the conversion from g_5 to e. The function ξ is equal to 1 in the limit $\lambda_i = 0$, where λ_i is the (set of) quartic Higgs couplings. The mass increases monotonically with all λ_i's, and reaches a finite limit when all λ_i's got to infinity. For any value of the coupling constants the decay of the higher charge monopoles into minimal charge ones is energetically allowed.

It may seem incorrect that we obtain a minimal Dirac charge monopole even though the theory contains quarks with charges that are multiples of $\frac{1}{3}$. The fact that the quarks are confined should not matter, since QCD never entered the discussion so far. The resolution of this paradox is that the minimal and double charge monopole have long-range color fields. These color fields produce an Aharonov-Bohm phase for a particle with color charge, and when added to the electromagnetic phase this is indeed not observable, as required by Dirac.

The triple charge monopole satisfies the Dirac quantization condition without any need for long-range color fields even with respect to quarks, and indeed it does not have such color fields. Interesting questions suggest themselves regarding the fate of long range color fields in view of confinement, but we will not pursue this discussion any further here.

It goes without saying that the experimental observation of a magnetic monopole would be an extremely important and exciting event. The minimal charge magnetic monopole is a stable particle. If it is light one could pair-produce it in accelerators, but GUT-monopoles, which have a mass quit close to $M_{\rm Planck}$, will never be produced that way. Our only hope is then that some were formed during the early stages of our universe. The first estimate of monopole abundances led to results that were far above any reasonable limit. For example, a good limit (the "Parker bound") is obtained from the observed presence of galactic magnetic fields in space. Magnetic monopoles would "short-circuit" such fields, and since that does not happen one may deduce a limit of about 10^{-15} monopoles per $\mathcal{M}^2$ per second on the magnetic monopole flux, if one assumes that the monopoles are distributed homogeneously. Early cosmological models produced monopole abundances far above this (and other) limits.

In inflationary cosmological models the abundance is drastically reduced. In fact inflation washes out any topological structure, thus reducing the number of monopoles to about 1 per universe. If this is indeed true monopoles will never be seen.

Independent of cosmological limits, it is still interesting to look for monopoles on earth. Inflation might be wrong, and many cosmological bounds might not apply if there were a local enhancement of monopoles. The sensitivity of present experiments is still less than the Parker bound.

Monopoles have been searched for using superconducting current loops ("squids"). The passage of a monopole through such a loop increases the current by a definite, quantized amount, which should be an easily recognizable signal. A second, though less direct signal for monopoles might be "catalysis of proton decay". GUT-monopoles are expected to have a very large cross section for turning protons into leptons and mesons, violating baryon number. This is possible because monopoles carry inside their core classical X and Y vector bosons, and because the lowest quark and lepton partial waves can penetrate all the way to the core without encountering any barrier. The precise magnitude of the cross section is hard to calculate and somewhat controversial, however. In any case, no evidence for the existence of magnetic monopoles has been found so far.

8.13 Other GUTs

There are other models in which the Standard Model is embedded in larger gauge groups. The most interesting possibilities are $SO(10)$, E_6 and various subgroups of these groups that can be regarded either as intermediate stages of symmetry breaking, or as unified theories in their own right.

All these models have more complicated Higgs systems and in principle several steps of symmetry breaking, and hence several intermediate scales. This removes much of the predictive power of minimal $SU(5)$. It is precisely that predictive power that has already eliminated *minimal* $SU(5)$ as a candidate GUT.

8.13.1 $SO(10)$

The most attractive possibility is $SU(5) \subset SO(10)$. The main advantage of this embedding is that one Standard Model family can be fit within a single irreducible representation, the spinor, which has dimension 16. This decomposes into $SU(5)$ in the following way

$$\mathbf{16} \rightarrow \mathbf{5}^* + \mathbf{10} + \mathbf{1} \tag{8.68}$$

We see that in addition to a Standard Model family we get a singlet. This has the quantum numbers of a right-handed neutrino, so that in these models it would be natural for the neutrinos to have a Dirac mass.

Another advantage is that $SO(10)$ does not have a rank three invariant tensor, so that all its representations are automatically anomaly-free. In $SU(5)$ there is still a cancellation between the $\mathbf{5}^*$ and the $\mathbf{10}$ which is not understood in a fundamental way. Furthermore the $\mathbf{16}$ is a complex representation, so that no mass terms are allowed before the $SO(10)$ symmetry is broken.

In addition to $SU(5)$, $SO(10)$ contains a $U(1)$ which turns out to be $B-L$. This was already an exact symmetry in the Standard Model and its $SU(5)$ extension, and it can thus be gauged, even without $SO(10)$ unification. The gauge boson of $B-L$ must acquire a mass well above the weak scale, since no light vector boson has been observed. Note that the coupling of this extra gauge boson is related by unification to the Standard Model couplings, so it can not be extremely small.

In $SO(10)$ there are additional heavy gauge bosons, connecting the $\mathbf{5}^*$, $\mathbf{10}$ and $\mathbf{1}$ to each other. The proton decay width and the branching ratios will thus be different.

The breaking of $SO(10)$ to the Standard Model can proceed in many ways. Simply checking the maximal sub-algebras of $SO(10)$ leads to the following two main breaking chains

$$\begin{array}{rcl} SO(10) & \to & SU(5)\times U(1) \\ & \to & SU(5) \\ & \to & SU(3)\times SU(2)\times U(1) \end{array} \tag{8.69}$$

$$\begin{array}{rcl} SO(10) & \to & SU(4)\times SU(2)\times SU(2) \\ & \to & SU(3)\times SU(2)\times SU(2)\times U(1)_1 \\ & \to & SU(3)\times SU(2)\times U(1)_2\times U(1)_1 \\ & \to & SU(3)\times SU(2)\times U(1)\ . \end{array} \tag{8.70}$$

The first step in these two chains is a breaking to a maximal subgroup. The groups $SU(5)\times U(1)$ and $SU(4)\times SU(2)\times SU(2)$ are the only two acceptable maximal subgroups of $SO(10)$. All others either do not contain the Standard Model, or break the $\mathbf{16}$ to a real representation, or both. In principle every step requires its own Higgs mechanism, although it is sometimes possible to perform two steps at once with a single Higgs. This leads in general to a rather complicated Higgs Lagrangian, and one or more additional intermediate scales, which one can consider as independent input variables in addition to $M_{\rm x}$ and $M_{\rm w}$ in $SU(5)$. Needless to say, the discussion of the possible minima of the potential becomes extremely complicated in these models. We will not discuss that issue here.

The second breaking of $SO(10)$ leads to a unification model considered first by Pati and Salam, before the $SU(5)$ model was found. They already predicted the possibility of proton decay in these models. In the various breaking steps, a Standard Model family emerges in the following way

$$\begin{array}{rcl} \mathbf{16} & \to & (4,2,1)+(4^*,1,2) \\ & \to & (3,2,1,\frac{1}{6})+(1,2,1,-\frac{1}{2})+(3^*,1,2,-\frac{1}{6})+(1,1,2,\frac{1}{2}) \\ & \to & (3,2,0,\frac{1}{6})+(1,2,0,-\frac{1}{2})+(3^*,1,\frac{1}{2},-\frac{1}{6})+(3^*,1,-\frac{1}{2},-\frac{1}{6}) \\ & & +(1,1,\frac{1}{2},\frac{1}{2})+(1,1,-\frac{1}{2},\frac{1}{2}) \\ & \to & (3,2,\frac{1}{6})+(1,2,-\frac{1}{2})+(3^*,1,\frac{1}{3})+(3^*,1,-\frac{2}{3})+(1,1,1)+(1,1,0) \end{array} \tag{8.71}$$

Here $Q_Y = Q_1 + Q_2$ (see Eq. (8.70) for the definition of these two charges). The first $SU(4) \times SU(2) \times SU(2)$ representation yields thus the left-handed quarks and leptons (left-handed particles), while the second one yields the right-handed ones (left-handed anti-particles). In the first two stages the model has a left-right symmetry. It is *not* invariant under parity or charge conjugation: both would map $(4, 2, 1)_L$ to $(4^*, 2, 1)_L$ (after transforming back to left-handed fields), which is a representation that does not occur. However one can define a new exact symmetry by combining P or C with an interchange of the gauge bosons of the two $SU(2)$ groups (provided they have the same coupling constant).

This is true for the kinetic terms and minimal couplings; Yukawa couplings and the Higgs potential might not respect such a symmetry. If indeed there is such left-right symmetry it must be spontaneously broken. We end up with the usual $W^\pm$ bosons coupling to left-handed fermions, plus two similar but more massive bosons coupling to right-handed ones. At still higher mass scales there are bosons transforming quarks into leptons, due to the embedding of $SU(3) \times U(1)$ in $SU(4)$, and at still higher energies one encounters bosons coupling particles to anti-particles.

8.13.2 E_6

One can go one step further and embed $SO(10)$ in E_6. This group is also anomaly free while having complex representations. The simplest one is the **27**. It decomposes to

$$\mathbf{27} \to \mathbf{16} + \mathbf{10} + \mathbf{1} \ , \tag{8.72}$$

The first term represents one family, while the second and the third are real, and thus have a chance to become massive well above the weak scale. This does not look especially attractive, and nothing is gained by extending $SO(10)$ to E_6, but nature need not follow that kind of logic. The group E_6 contains $SO(10) \times U(1)$, and the extra $U(1)$ bosons must acquire a mass.

Just as above, the breaking of E_6 does not have to go via $SO(10) \times U(1)$, but one could also consider other maximal subgroups. The most popular one is $SU(3)^3$ (the other viable candidate is $SU(2) \times SU(6)$, but this does not seem to have been studied much). One of those $SU(3)$'s becomes the color group, whereas the other two contain $SU(2)_L$ and $SU(2)_R$, which are respectively the $SU(2)$ group of the Standard Model, and its counterpart for the right-handed fermions discussed above.

The main attraction of $SU(3)^3$ is the global S_3 permutation symmetry which one can impose. If one does, the coupling constants of the three groups are equal without a need for full unification into E_6. They will remain equal to all orders in perturbation theory. However, this symmetry must be broken spontaneously to get the Standard Model.

8.13.3 Flipped $SU(5)$

A possibility discussed recently is a different embedding of $SU(3) \times SU(2)$ in $SU(5)$. Instead of the decomposition $\mathbf{5}^* = (e^-, \nu, d^c)_L$ and $\mathbf{10} = (e^+, u^c, u, d)_L$ one chooses $\mathbf{5}^* =$

$(e^-, \nu, u^c)_L$ and $\mathbf{10} = (\nu^c, d^c, u, d)_L$, and one adds an $SU(5)$ singlet e_L^+. Note that there is one extra particle per family, a right-handed neutrino. As far as the $SU(3) \times SU(2)$ representation is concerned this is possible, since the difference between a flipped particles is just the electric charge. However, there is now only one way to get the correct electric charge, and that is to add an extra $U(1)$ factor. The Standard Model $U(1)_Y$ is then a linear combination of the $U(1)$ subgroup of $SU(5)$ and the extra factor. Let us denote the charges respectively as Q_F, Y and Q_5, where Q_5 is the $U(1)$ embedded in $SU(5)$. Thus Q_5 is exactly Y in standard $SU(5)$, and we will normalize it in the same way. It is then easy to check that the combination $Y = -\frac{1}{5}Q_5 + Q_F$ gives the correct answer, if we assign Q_F charges $-\frac{3}{5}$, $\frac{1}{5}$ and 1 to the five, ten and the singlet respectively. Note that Q_F is traceless. In fact, it turns out that this $SU(5) \times U(1)$ model is a subgroup of $SO(10)$, and that the charge Q_F is $B - L - \frac{4}{5}Y$.

This model seems to have few advantages and many disadvantages. Even the nice property of automatic charge quantization is lost, since there is an extra $U(1)$ factor. The main reason why this model was considered is that one can break it to $SU(3) \times SU(2) \times U(1)$ with a Higgs in the $\mathbf{10}$ of $SU(5)$. This was seen as an advantage for such a model in the context of superstring theories, since in most string theories one cannot get a Higgs in the $\mathbf{24}$ of $SU(5)$, but only in the $\mathbf{1}$, $\mathbf{5}$ or $\mathbf{10}$.

8.13.4 Still Larger Groups

One may continue along the same line, and embed E_6 in E_7 or E_8, or $SO(10)$ in $SO(10+n)$. The former two possibilities are rather unattractive sine E_7 and E_8 have only real representations. The group $SO(10+n)$ has complex representations for $n = 0 \bmod 4$, namely the spinor representations, but upon breaking to $SO(10)$ one gets an equal number of $\mathbf{16}$'s and $\mathbf{16}^*$'s, *i.e.* an equal number of families and mirror families.

Many other possibilities have been considered, but we will not enumerate them here.

8.14 Conclusions

The idea of Grand unification is *a priori* very attractive. This idea can in a natural way explain the following features of the Standard Model

+ Coupling constant convergence.

+ Charge quantization.

+ Structure of a Standard Model family.

+ Some quark/lepton mass relations.

As we have seen, not all models score equally well on all these points. On the other hand the following features are not explained – at least not in typical, non-contrived models:

– The smallness of the weak scale.

– Family repetition.

– Inter-family mass hierarchies.

– The strong CP problem.

The first of these problems is actually made more serious due to the explicit introduction of a large scale into the problem, which does not decouple naturally from the weak scale.

The minimal GUT model, based on $SU(5)$ does not agree with experiment: precise LEP measurements have shown that the three running coupling constants do not go exactly meet in one point, and the expected proton decay has not been found. All other models have extra parameters, and are much harder to rule out.

However, the idea of grand unification is far from dead. Two remarkable facts remain: that the coupling constants converge *approximately*, and that one family fits exactly in two representations of $SU(5)$, and – with an extra right-handed neutrino – in a single representation of $SO(10)$. These two observations will undoubtedly continue to play an important rôle in the future.

8.15 References

Most of this section was based on the extensive review by P. Langacker [20]. A useful review of group theoretical results for unification is [29].

9 Supersymmetry

Supersymmetry is a symmetry relating bosons to fermions. There is no doubt that it plays an important rôle in theoretical particle physics already. It has been used for proving index theorems, deriving positive energy theorems and lower bounds on soliton masses, to construct consistent fermionic strings and many other purposes. All of these are technical applications, however. The question is: could it be a symmetry of nature?

At first sight it seems that the answer must be negative. Among the particles in the Standard Model, there is at most one boson-fermion pair with the same mass (the photon and one of the neutrinos), and only one pair that belongs to the same $SU(3) \times SU(2) \times U(1)$ representation (a lepton doublet and the complex conjugate of the Higgs). So if supersymmetry is a symmetry of nature it must be badly broken. This is not a problem in itself, since we know from the Standard Model that badly broken symmetries can nevertheless play a crucial rôle in our understanding, but the difference is that at least we have always known several complete $SU(2)_{\mathrm{w}}$ multiplets. It is this difference that makes phenomenological supersymmetry a much more speculative subject. There simply is not the slightest piece of direct evidence in its favor. Many people hoped that after the first LHC run that finished in march 2013, some of the missing superpartners would finally have emerged, after several decades of expectations. But this has not happened.

There are several *a priori* motivations for attempting to supersymmetrize the Standard Model. The first and most primitive one falls under the category "why not". It is argued that supersymmetry is a very beautiful idea, and that it would be a pity if nature chose to ignore it. No further comments are needed here.

The second motivation is that supersymmetry is known to improve the divergent behavior of perturbation theory. For example, $N = 4, D = 4$ Yang-Mills theory (four supersymmetries in four dimensions) was shown to be finite to all orders in perturbation theory! This is a very remarkable result, but not a motivation to make $SU(3) \times SU(2) \times U(1)$ supersymmetric. In a finite theory without a scale the β-functions vanish and the coupling constants do not run, whereas we observe that they do run. After $N = 4, D = 4$ Yang-Mills was shown to be finite, it was hoped that one could also find a finite supersymmetric theory of gravity. So far the maximally supersymmetric theory, $N = 8$ supergravity, has not been demonstrated to be finite, nor has the contrary been shown convincingly. The hope of a finite theory of gravity has however probably been realized by superstring theory. The spectrum of this theory is supersymmetric,* and if indeed superstring theory is the only way to make sense of perturbative gravity, one could view this as an argument in favor of a supersymmetric spectrum. One should add immediately that finiteness is useful in practice only if it survives supersymmetry breaking. If it does, then it should not make any difference if supersymmetry is broken far below the Planck scale, or just slightly below. So even if this is an argument in favor of supersymmetry, it does not imply low-energy supersymmetry.

The primary motivation for believing that supersymmetry might have a rôle to play in particle physics is the hierarchy problem. One way to formulate that problem is that one cannot have scalars that are naturally massless without having supersymmetry. In comparison to large scales such as the GUT scale or the Planck scale all Standard Model particles are essentially massless. One may call that natural if there is an exact symmetry in the zero mass limit. Of course the particle of interest here is the scalar particle discovered in 2012, the first particle that might be a fundamental scalar: the Higgs boson. This particle is not *exactly* massless: before symmetry breaking the Higgs scalar ϕ has a mass^2 $\mu^2 < 0$ and after symmetry breaking a physical scalar η appears in the spectrum with mass $\sqrt{-2\mu^2}$. But, as explained in sec. 7, the value of μ^2 is extremely small in comparison to the GUT or Planck scale, so that to first approximation the Higgs scalar is a massless scalar.

The only massless (or nearly massless) particles that one can have in a sensible field theory have spin $0, \frac{1}{2}, 1, \frac{3}{2}$ or 2. There are good arguments for that in field theory, and string theory respects that rule as well. For each of these particles except spin 0, there is a natural symmetry that can protect them against large mass corrections. Particles of spin 1 are protected by gauge invariance. In order to make them massive, one has to find an additional degree of freedom to go from two to three polarizations. A Higgs scalar can provide that degree of freedom. But without such a scalar, there is no possibility for a

* This is true by definition: superstrings are supersymmetric string theories. There also exist non-supersymmetric string theories. They are finite at one-loop order, but beyond that it is difficult to make sense of them.

spin-1 particle to acquire a mass. The same argument holds for spin-2: a graviton has two polarizations, but a massive spin-2 particle has five. For fields of spin 2, the "protection symmetry" is general relativity, and the spin 2 field must be the graviton. For spin-$\frac{3}{2}$ (if such particles are ever observed) supergravity acts as the protection mechanism.

Spin-$\frac{1}{2}$ particles are protected by chiral symmetries. If a fermion mass is set to zero, a new symmetry emerges: one can now rotate the left- and right components of the fermion independently. Such symmetries are respected in perturbation theory, and hence no mass term will be generated if it was not already there. In this sense massless spin-$\frac{1}{2}$ particles are "protected". Unlike all previous symmetries the chiral symmetry does not have to be local, and the spin-$\frac{1}{2}$ particles are not in any sense gauge particles.

There are two known protection mechanisms for massless scalars: they could either be Goldstone bosons of some broken global symmetry,* or they could be protected by (global or local) supersymmetry. It is difficult to regard the Standard Model Higgs boson as a Goldstone boson, although ideas in that direction have been explored. One problem is that a Goldstone bosons would have derivative couplings with all fields, a property not shared by the Higgs field of the standard model.†

The supersymmetric protection mechanism is easy to understand: supersymmetry pairs the scalar with a fermion, whose mass is protected by chiral symmetry. Since supersymmetry requires the boson and fermion mass to be equal, the boson mass is now protected as well. It is thus natural to wonder if perhaps supersymmetry can be used to solve the hierarchy problem.

Note that supersymmetry has nothing to say about the weak interaction scale itself. If supersymmetry is unbroken the Higgs mass, which in the Standard Model is related to the weak scale, is an arbitrary parameter, as we will see; if supersymmetry is broken the weak scale is determined by the supersymmetry breaking scale, and then one can start arguing about the origin of that scale. Here technicolor appears to have the advantage. In that case the scale is determined by a gauge coupling constant becoming strong, and we know from the example of QCD that it is quite natural for this to happen at scales much below the Planck or unification scale. The most popular mechanism for supersymmetry breaking, gaugino condensation, also involves dynamical symmetry breaking, so that in such models the scale would be determined as well.

9.1 The Supersymmetry Algebra

The generators of supersymmetry must transform fermions and bosons. They must thus be anti-commuting, and be spinors. The simplest ($N = 1$) supersymmetric algebra has the form (see appendix A for index conventions and appendix D for a derivation of this

---

* There is also a second possibility for massless fermions, namely that they are Goldstinos of broken global supersymmetry. Attempts have been made to regard standard model fermions as Goldstinos, but without much success. † By "Higgs field" we mean here the complex scalar field ϕ in the unbroken Standard Model. One should not confuse this with the fact that three of the four real Higgs scalars become Goldstone bosons *after* $SU(2) \times U(1)$ breaking, which are eaten by the W and the Z. There is in any case also a fourth, physical Higgs boson, which cannot be a (pseudo) Goldstone boson.

algebra in a theory with a free boson and a free fermion).

$$
\begin{array}{rcl}
[Q_\alpha, P_\mu] & = & 0 \\
\{Q_\alpha, Q_\beta\} & = & 0 \\
\{\bar{Q}_{\dot{\alpha}}, \bar{Q}_{\dot{\beta}}\} & = & 0 \\
\{Q_\alpha, \bar{Q}_{\dot{\beta}}\} & = & 2\sigma^\mu_{\alpha\dot{\beta}} P_\mu \ .
\end{array}
$$

The trace of the last equation implies

$$H = P^0 = \frac{1}{4}(\bar{Q}_1 Q_1 + Q_1 \bar{Q}_1 + \bar{Q}_2 Q_2 + Q_2 \bar{Q}_2) \ . \tag{9.1}$$

Therefore, for some state $|\Psi\rangle$ we get

$$\langle\Psi| H |\Psi\rangle = \tfrac{1}{4}(|Q_1 |\Psi\rangle|^2 + |\bar{Q}_1 |\Psi\rangle|^2 + |Q_2 |\Psi\rangle|^2 + |\bar{Q}_2 |\Psi\rangle|^2) \geq 0 \tag{9.2}$$

Therefore the expectation value of H is precisely 0 if and only if $|\Psi\rangle$ is annihilated by all supersymmetry generators. Obviously such a state $|\Psi\rangle$ is the state with lowest energy, *i.e.* the vacuum. As with other symmetries, supersymmetry is a symmetry under two conditions: it must commute with the Hamiltonian, and it must be a symmetry of the ground state. The latter condition is

$$Q_\alpha |0\rangle = \bar{Q}_{\dot{\alpha}} |0\rangle = 0 \tag{9.3}$$

Then trivially $\langle 0| H |0\rangle = 0$, *i.e.* the ground state has zero energy. The contrary is also true. If $Q_\alpha |0\rangle \neq 0$ then $\langle 0| H |0\rangle$ can be written as $\frac{1}{2}\sum_\alpha |Q_\alpha |0\rangle|^2 > 0$.

The fact that the energy of the vacuum is zero is a first indication of cancellation between fermions and bosons. In a non-supersymmetric bosonic field theory the zero-point energy of the bosonic oscillators is positive and add up to infinity (which is then set to zero), whereas fermions give a negative contribution.

9.2 Multiplets

Since the supercharge transforms bosons into fermions and vice-versa, it is clear that it organizes the field content of the theory into super-multiplets, which are representations of the supersymmetry algebra. In the simplest case, $N = 1$, there are only two relevant multiplets, called the *chiral multiplet* and the *vector multiplet*. The former consists of a complex scalar and a complex Weyl fermion, the latter contains a real vector boson and a Majorana fermion. The fields in each multiplet must transform according to the same representation of any gauge symmetry. The members of a vector multiplet must thus both belong to the adjoint representation of a gauge group, of which the vector boson is the gauge boson.

Chiral supermultiplets can be left-handed or right-handed, because a scalar in a representation R may be paired either with a left-handed Weyl fermion in the representation R^*, or a right-handed Weyl fermion in the representation R. The Hermitean conjugate of a left-handed chiral multiplet contains a scalar and a right-handed Weyl fermion, both in the representation R^*. This is thus a right-handed chiral multiplet. Gravity requires an additional multiplet, containing a spin 2 and a spin-$\frac{3}{2}$ particle.

There also exist extended supersymmetries with more than one supercharge. Their representations are larger and can contain more different spins. If one requires that the highest spin that occurs is 2, the maximal number of supersymmetries is 8.

In extended supersymmetry theories every multiplet contains only real fermion representations. This does not look like a very promising starting point if one wants to obtain the Standard Model, which has complex matter representations. The simplest example of an extended supersymmetry is $N = 2$ supersymmetry. In this theory, matter belongs to "hyper-multiplets" which can be decomposed into a chiral and an anti-chiral multiplet of $N = 1$ supersymmetry. Hence a hypermultiplet consists of a right-handed Weyl fermion, a left-handed Weyl fermion and two complex scalars, all in the same representation. This means that for every right-handed fermion there is automatically a left-handed one: the theory is not chiral. This is not a good starting point for phenomenology (although this has not stopped all attempts in that direction).

On-shell (*i.e.* when the equations of motion are imposed) a Weyl-fermion and a Majorana fermion both have two degrees of freedom. A complex scalar and a real, on-shell vector boson also have two degrees of freedom. Hence each multiplet does indeed contain an equal number of bosonic and fermionic degrees of freedom.

To write down an action we need off-shell fields. The equations of motion follow from the action, but are not imposed on it. Off-shell a complex scalar still has two degrees of freedom, but a Weyl and a Majorana fermion have four, just as a vector boson. To realize supersymmetry off-shell additional fields have to be introduced, which can be removed from the action by their equations of motion, since they do not have kinetic terms. These are called auxiliary fields. For the scalar multiplet we need one complex bosonic auxiliary field to get the correct counting. For a vector multiplet one might expect to need none, but there is a complication since the reduction of the number of degrees of freedom for a vector boson involves not only the field equations, but also gauge invariance. In fact, the full set of auxiliary fields for the vector multiplet contains several bosons and fermions.

9.3 Constructing supersymmetric Lagrangians

For $N = 1$ supersymmetry an elegant formalism is available to construct invariant Lagrangians. This is the superfield formalism, which is reviewed in appendix D. Here we just summarize the main results.

Fields in non-supersymmetric field theory are combined into superfields. A superfield depends on the space-time point x in the usual way, and in addition on an parameter θ^α. This parameter is anti-commuting and the index α can take two values, 1 and 2. The anti-commutativity implies that $(\theta^1)^2 = (\theta^2)^2 = 0$. A superfield is a polynomial in θ, and it follows immediately that the highest order term in such a polynomial has order 2. This is usually written as θ^2, and is equal to $\theta_1\theta_2$. In addition we also introduce a parameter $\bar{\theta}^\alpha$, with analogous properties. It anti-commutes with θ.

There are two kinds of superfields.

- Chiral superfields Φ: By definition these depend only on x and θ , but not on $\bar{\theta}$.
- Vector superfields V : These depend on x, θ and $\bar{\theta}$, but are Hermitean: $V^\dagger = V$

Here the Hermitean conjugate of θ is $\bar{\theta}$. The indices of θ and $\bar{\theta}$ and the Lorentz transformations of these objects require more discussion, which can be found in the appendix.

A chiral superfield can be expanded in θ. Then one gets the following expression

$$\Phi(x,\theta) = \varphi(x) + \sqrt{2}\theta\psi(x) + \theta^2 F(x) \ , \tag{9.4}$$

The component $\varphi(x)$ is a complex scalar field, and $\psi(x)$ is a Weyl spinor. We choose it left-handed by convention. The strange factor $\sqrt{2}$ is also a convention. The component $F(x)$ is unphysical: it does not lead to propagating degrees of freedom. It is called an auxilliary field. Its rôle is make sure that the superfield contains the same number of fermionic components both on-shell and off-shell.

Here "off-shell" refers to the count of the field components. A Weyl spinor is complex two-component field. It must be complex, because $SO(3,1)$ Lorentz rotations acting on spinors are complex. Hence a Weyl spinor has four off-shell components. The complex scalar fields $\varphi(x)$ and $F(x)$ have two components each, so that the boson/fermion counting works out: $2 + 2 = 4$.

On the other hand "on-shell" refers to the counting of physical, propagating degrees of freedom. A Weyl fermion has two propagating degrees of freedom. This is because the word "on-shell" means that the Dirac equation is imposed as a constraint on the field. This can be seen explicitly in the Dirac propagator which contains a factor $\not{k} + m$ (see *e.g* 5.40). On-shell, if $k^2 = m^2$, this has two eigenvalues zero, reducing the number of propagating componets by a factor 2. This reduction works in the same way for a Weyl spinor. In both cases, one can start with a complex, four component Dirac spinor, which has eight degrees of freedom. The Dirac equation reduces this to four physical ones, and in addition Weyl spinors satisfy the constraint $\gamma_5\psi = 0$, which gives another reduction by a factor 2. A complex scalar has two physical degrees of freedom. As already stated, $F(x)$ is entirely unphysical, and hence the on-shell boson/fermion count also works: $2 + 0 = 2$.

The story for vector fields is similar, but more complicated. A full expansion gives many terms, but using gauge invariances several can be put to zero. The result is

$$V(x,\theta,\bar{\theta}) = -\theta\rho_\mu\bar{\theta}V^\mu + i\theta^2\bar{\theta}\bar{\lambda} - i\bar{\theta}^2\theta\lambda + \tfrac{1}{2}\theta^2\bar{\theta}^2 D + \dots \ , \tag{9.5}$$

Here V^μ represents a real vector field, λ a Majorana fermion, and D is an auxilliary field. The on-shell count is as follows: two d.o.f. for V^μ, two d.o.f. for λ and zero for D. The off-shell count involves gauge invariance and to check it one must include the omitted terms.

The rule for writing down supersymmetric Lagrangians using superfields are as follows. There are two kinds of terms

- F-terms: products of superfields that depend only on x and θ

- D-terms: products of superfields that depend on x, θ and $\bar{\theta}$, and are Hermitean

The rule is now to expand these products of fields in terms of θ and $\bar{\theta}$, and keep only the terms of higher order in the anti-commuting variables, *i.e.* proportional to $\theta_1\theta_2$ for F-terms and proportional to $\theta_1\theta_2\bar{\theta}_1\bar{\theta}_2$ for D-terms. The coefficient functions of these highest powers of θ can be shown to be supersymmetric. The resulting Lagrangian will in general depend on the auxilliary fields F and D. However, these fields always appear without derivatives. Hence there equations of motion are non-dynamical. They simply state that the variation of the Lagrangian with respect to F and D must vanish. This yields a simple algebraic constraint, that can be solved to eliminate F and D.

The F-terms are the easiest ones to deal with. They should not have any dependence on $\bar{\theta}$, and hence they can only be polynomials in terms of the chiral superfields. Just as in non-supersymmetric QFT, any term in the polynomial that respects all the symmetries is allowed. However, to get a renormalizable theory one may allow only terms of at most order three in the chiral superfields. This polynomial is called the *superpotential*. To derive it for the supersymmetrized Standard Model all we have to do is take all the left-handed Weyl spinors and assign a superfield to each, and then build the most general superpotential terms that are invariant under $SU(3) \times SU(2) \times U(1)$.

The D-terms are a bit more difficult to discuss. To construct these out of chiral superfields, one would like to consider the Hermitean conjugate of a chiral superfield, and build something like $\Phi^\dagger\Phi$. But $\Phi^\dagger$ does not transform correctly under supersymmetry. It is neither a chiral superfield nor a vector superfield. To solve this one first has to apply a transformation to $\Phi^\dagger$. Fortunately, D-terms play a rather simple rôle in the construction. They give rise to the kinetic terms of the fermions and the bosons, which we could easily have written down anyway.

$$\mathcal{L} = \Phi^\dagger\Phi\big|_{\mathrm{D}} = -\partial_\mu\varphi\partial^\mu\varphi + i\psi\rho^\mu\partial_\mu\bar{\psi} + FF^* \tag{9.6}$$

The only term of interest here is the dependence on the auxilliary field.

One can couple these kinetic term to gauge fields by considering

$$\mathcal{L} = \Phi^\dagger e^{2gV}\Phi\big|_{\mathrm{D}} \tag{9.7}$$

Under non-abelian symmetries, V transforms as $V \to UVU^\dagger$, and ϕ as $\Phi \to U\Phi$, so that gauge invariance is guaranteed. To be precise, if Φ is in some representation R with generators T^a, one writes $V = V^aT^a$; obviously one needs a vector superfield with as many components as there are gauge bosons. The effect of the coupling to gauge fields is that instead of (9.6) we get

$$\Phi^\dagger e^{2gV}\Phi\big|_{\mathrm{D}} = -|D_\mu\varphi|^2 - i\psi\rho^\mu D_\mu\bar{\psi} + 2ig[\varphi^*\lambda\psi - \varphi\bar{\lambda}\bar{\psi}] + FF^* + g\varphi^* D\varphi\ . \tag{9.8}$$

The first two terms are not unexpected: derivatives become covariant derivatives. There are two additional terms: a scalar-fermion-gaugino coupling and a coupling of two scalars to the auxilliary field.

There is on other kind of D-term that is clearly possible, and that is a term linear in V. This is invariant under supersymmetry, but only invariant under gauge transformations if one takes a trace. This can only be non-zero for abelian gauge theories. This kind of term does not play a major rôle in the following. Terms of quadratic and higher order in V have two many $\theta's$ to yield anything.

So far the formalism was fairly elegant, but this cannot be said about the gauge kinetic terms. These are actually F-terms, but these are constructed in rather baroque way. We refer to the appendix for details, but gauge invariance fixes most of the structure anyway:

$$\mathcal{L}_{\text{gauge}} = -\tfrac{1}{4}(F^a_{\mu\nu})^2 - \tfrac{1}{2}\bar{\Lambda}\gamma^\mu D_\mu \Lambda + \tfrac{1}{2}(D^a)^2 \ , \tag{9.9}$$

Not surprisingly, there are kinetic terms for the gaugino, and also not surprisingly they involve a covariant derivative. The only noteworthy term is the quadratic one involving the auxilliary fields.

This is all we need to write down a supersymmetric extension of the Standard Model.

9.4 The Supersymmetrized Standard Model

To write a supersymmetric version of the Standard Model we have to pair all fields into supermultiplets, and introduce the supersymmetric partners of all remaining fields. For the 12 gauge bosons of $SU(3)\times SU(2)\times U(1)$ no partners are available among the Standard Model fermions, and so we have to introduce 12 real fermions, namely 8 "gluinos", a "photino", two "winos" and a "zino". For the chiral fermions we need scalar partners. Here the only possible pairing would be between the Higgs doublet $C\phi^*$ (see the discussion following Eq. (4.24) for the notation), transforming in the representation $(1, 2, -\frac{1}{2})$ and one of the lepton doublets. However, one would like to assign lepton number 1 to the lepton doublet and 0 to the Higgs doublet in order to preserve lepton number. These values are in conflict with each other, making this identification impossible (the resulting violations of lepton number would be far too large)

For the other quarks and leptons we have no choice, and so we introduce left-handed chiral multiplets $\mathcal{Q}$, $\bar{\mathcal{U}}$, $\bar{\mathcal{D}}$, $\mathcal{L}$ and $\bar{\mathcal{E}}$ with Standard Model representations

$$(3, 2, \tfrac{1}{6}), (3^*, 1, -\tfrac{2}{3}), (3^*, 1, \tfrac{1}{3}), (1, 2, -\tfrac{1}{2}), (1, 1, 1)$$

respectively, plus the Standard Model singlet $\bar{\mathcal{N}}$ for the left-handed anti-neutrino (or equivalently the right-handed neutrino), if desired. The bars are added to remind ourselves of the fact that these fields represent anti-particles; the right-handed chiral multiplets that are the conjugates of these fields are denoted as $\mathcal{Q}^\dagger$,$\bar{\mathcal{U}}^\dagger$ etc. All these fields carry, in addition to their $SU(3)\times SU(2)\times U(1)$ indices, a flavor index with three distinct values. The resulting particles are called "squarks" (scalar quarks) and "sleptons". Note that there will be a squark for every left-handed and another one for every right-handed field (in the particle representation). Hence for example the up quark has two scalar partners, often denoted $\tilde{u}_L$ and $\tilde{u}_R$; of course since they are scalars the chirality index only refers to the fermion they belong to.

The kinetic terms of these fields require no further discussion, but the Yukawa couplings are more interesting. In the non-supersymmetric Standard Model we needed the Higgs scalar ϕ as well as the conjugate $C\phi^*$ to give mass to all quarks. Suppose we introduce a left-handed chiral superfield H_d in the representation $(1, 2, -\frac{1}{2})$. Thus H_d transforms exactly like $\mathcal{L}$, and the scalar component of H_d transforms exactly like the complex conjugate Higgs field $C\phi^*$. Using this field we can write down the following Yukawa couplings:

$$g_{\mathcal{D}} \mathcal{Q} H_d \bar{\mathcal{D}} + g_{\mathcal{E}} \mathcal{L} H_d \bar{\mathcal{E}} \tag{9.10}$$

Here all indices have been suppressed, but they are exactly as in Eq. (4.24). The two terms given in Eq. (9.10) yield the complex conjugates of the last two terms in Eq. (4.24), when one considers only the terms involving Standard Model particles. Of course both fermions in the resulting Yukawa coupling will be left-handed, and one has to convert one of them to right-handed notation to get Eq. (4.24) (up to an irrelevant overall phase). The structure of Eq. (9.10) is dictated by gauge invariance, and in particular the $SU(2)$ indices must be contracted as $\mathcal{Q}^a H_d^b \epsilon_{ab}$.

Now we would like to write down the equivalent of the first term in Eq. (4.24), and we would also like to introduce neutrino Yukawa couplings. The obvious guess is $\mathcal{Q} H_d^\dagger \mathcal{U}$, but $H_d^\dagger$ is a right-handed superfield, and there exists no supersymmetric coupling to the two left-handed superfields $\mathcal{Q}$ and $\mathcal{U}$. This forces us to introduce a new field H_u which transforms like $(1, 2, \frac{1}{2})$. Then the missing Yukawa couplings are

$$g_{\mathcal{U}} \mathcal{Q} H_u \bar{\mathcal{U}} + g_{\mathcal{N}} \mathcal{L} H_u \bar{\mathcal{N}} + \text{c.c} \tag{9.11}$$

Here again all indices are contracted in the obvious way, flavor indices as in Eq. (4.24), and all others as dictated by gauge invariance.*

There is another reason why we are anyway forced to introduce an additional Higgs doublet. The supermultiplet H_d contains a left-handed fermion in the representation $(1, 2, -\frac{1}{2})$. This field contributes to the $SU(2)\times U(1)$ and $U(1)^3$ anomalies of the Standard Model, and hence we have to introduce additional matter to cancel these anomalies. The simplest solution is to add a left-handed chiral superfield in the representation $(1, 2, \frac{1}{2})$.

This gives also another reason why it is not a good idea to identify the fields H_d with one of the flavors of the lepton doublets $\mathcal{L}_i$: to get masses for the up quarks we would in any case need a chiral superfield in the representation $(1, 2, \frac{1}{2})$, and to cancel the anomalies introduced by this field we need to add a $(1, 2, -\frac{1}{2})$ superfield. So the fields H_d and H_u are needed in any case.

9.5 Additional Interactions

The Yukawa couplings are not the only interactions one can write down. First of all one may have a term

$$\mu H_d H_u \tag{9.12}$$

* We have chosen H_d and H_u to transform in the same $SU(2)$ representation and not in complex conjugate representation, as some others do. The difference is merely an ϵ tensor.

in the superpotential. If one works out the scalar potential one finds that this simply gives rise to a mass term

$$|\mu|^2(|h_d|^2 + |h_u|^2) \tag{9.13}$$

for the higgs scalars H_d and h_u in the superfield, as well as a contribution to higgsino mass matrix. Note that these Higgs scalar masses are free parameters, just as in the Standard Model, but that – unlike the Standard Model Higgs mass term – $|\mu|^2$ can only be positive. Hence there is no possibility for spontaneous $SU(2) \times U(1)$ breaking with the present form of the potential. This gives us no reason for concern: all this is true only as long as supersymmetry is not broken, but we know that it has to be broken. There are no other renormalizable superpotential contributions involving only H_d and H_u. Note that we do not get any quartic scalar potential terms from the superpotential.

The extra term, Eq. (9.12), is not good news. Unlike the Yukawa coupling constants, μ has the dimension of a mass. If our ambition is only to build a theory with naturally protected hierarchies, μ poses no problem: as we will see the coefficients of the superpotential are not renormalized, and hence we can give μ any value we like in a "natural" way. But unless μ is of order M_{Planck} its existence introduces a μ/M_{Planck} hierarchy problem (here instead of M_{Planck} one can substitute any other large scale that occurs in the theory). Of course μ cannot be of order M_{Planck}, because it will contribute to the Higgs mass parameter after supersymmetry breaking, and hence its natural value is of order the weak scale (or smaller).

In addition one can add the following four terms:

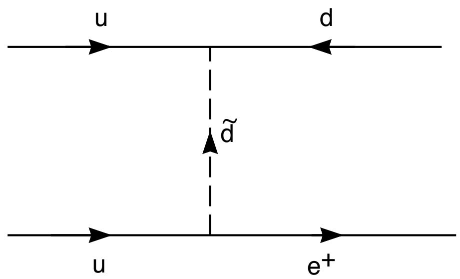
$$\mathcal{Q}\mathcal{L}\bar{\mathcal{D}}; \quad \mathcal{L}\mathcal{L}\bar{\mathcal{E}}; \quad \mathcal{L}\bar{H}_u; \quad \bar{\mathcal{U}}\bar{\mathcal{D}}\bar{\mathcal{D}}; \tag{9.14}$$

where again the index structure is dictated by gauge invariance. Each term would appear with a coupling tensor with as many flavor indices as there are fields.

These terms are undesirable, since they manifestly violate either baryon or lepton number. They do not appear in the standard model although they would be allowed by $SU(3) \times SU(2) \times U(1)$ group theory. The reason is that in the Standard Model Lorentz-invariance forbids them: one cannot couple three fermions to a singlet, or a fermion to a scalar. This is a clear disadvantage of the supersymmetric extension of the Standard Model.

Note that the first three terms are simply the Yukawa couplings of the field H_d, with H_d replaced by $\mathcal{L}$. This gives us yet another reason why one should not identify H with $\mathcal{L}$, because in that case such undesirable couplings are certainly inevitable.

A contribution to proton decay due to these terms is shown below. Here the dashed line indicates a scalar component of a superfield and the solid line a fermion component. The diagram corresponds to the decay $p \to e^+ + \ldots$ (the terms denoted by $\ldots$ are hadrons that are needed for energy-momentum conservation, and that would be created when the proton breaks up).



The arrow convention is such that it shows the flow of color charges. In any vertex there must be the same number as in- and outgoing arrows, except for vertices that make use of the ϵ-tensor coupling; they have three incoming (or outgoing) arrows. By adding a free d-quark line, this particular process can be interpreted as $p = (uud) \to (e^+)(dd^c)$, and the dd^c becomes a neutral pion. One may suppress this decay either by making the coupling constant extremely small or the mass of the scalar component of the d-quark (usually called the d-squark) extremely large. From GUTs we know that with couplings of order 1 the mass of this squarks would have to be of order 10^{15} GeV. This would imply an extremely large supersymmetry breaking, and it is hard to see how supersymmetry could in that case still have something to do with the breaking of weak interaction symmetries.

The limits on the three terms that only violate lepton number are less severe (the best limit is about .01, for the first family $\mathcal{U}\mathcal{L}\bar{\mathcal{D}}$ coupling), but nevertheless these couplings are usually set to zero. They are zero by definition in the minimal supersymmetric Standard Model (MSSM).

In general one cannot simply set allowed couplings to zero. This is possible only if there is a symmetry protecting them. This symmetry will then be preserved by all quantum corrections, so that the undesirable terms will not be generated if we omit them from the Lagrangian. So we should try to find symmetries of the Standard Model Lagrangian that are not symmetries of the undesirable terms. There are in fact many global symmetries that are broken by these unwanted terms. The most obvious choice is B or L or $B - L$ (where B and L are assigned to the entire supermultiplet, *i.e.* a squark has $B = \frac{1}{3}$). But B and L are not good candidates for a fundamental symmetry of nature, because these symmetries have an anomaly with respect to $SU(2)_{\mathrm{Weak}}$. The combination $B - L$ does not have that problem, but this global symmetry would forbid a mass-term for right-handed neutrinos (just as L by itself) and hence would inhibit the see-saw-mechanism. But perhaps the see-saw mechanism is not realized in nature. Then one has to accept unnaturally small neutrino Dirac masses, which is a high price to pay for solving another naturalness problem, the hierarchy problem. Furthermore, the idea that $B - L$ could be an exact global symmetry is not likely to be correct, because it is believed that gravity does not allow exact global symmetries. But if $B - L$ is a local symmetry, we should see an extra abelian gauge boson.

There is another possibility. The Standard Model has five $SU(3) \times SU(2) \times U(1)$ matter representations and a single Higgs boson. The MSSM has one extra representation, namely H_u. Therefore one may expect an additional $U(1)$ global symmetry X. Indeed there is, if the terms (9.12) and (9.14) are absent. The corresponding $U(1)_X$ charges are 1 for H_d and H_u, and $-\frac{1}{2}$ for all quark and lepton left-handed chiral superfields. All extra terms discussed in this section break X, or in other words if we impose it all of these terms are forbidden. Note that in the Standard Model H_d and H_u correspond to $C\phi^*$ and ϕ, so that this charge assignment is not possible.

9.6 Continuous R-symmetries

In supersymmetric theories there is one extra global symmetry called R-symmetry, which unlike all others does not act equally on all members of a supermultiplet. In the superfield formalism this symmetry acts on the θ parameter by changing it by a phase:

$$\theta \to e^{i\alpha}\theta \ . \tag{9.15}$$

By definition, θ has R-charge 1. One may assign furthermore R-charges to all superfields. The components of the superfield transform then according to the number of factors of θ by which they are accompanied. For a left-handed chiral superfield ϕ with R-charge r, the decomposition

$$\phi(x,\theta) = \varphi(x) + \sqrt{2}\theta\psi(x) + \theta^2 F(x) \ , \tag{9.16}$$

implies that φ has charge r, ψ charge $r-1$ and F charge $r-2$.

For a vector superfield

$$V(x,\theta,\bar\theta) = -\theta\sigma_\mu\bar\theta V^\mu + i\theta^2\bar\theta\bar\lambda - i\bar\theta^2\theta\lambda + \tfrac{1}{2}\theta^2\bar\theta^2 D + \dots \ , \tag{9.17}$$

we find that V^μ and D have charge r, λ charge $r+1$, $\bar\lambda$ charge $r-1$. However, V^μ is a real field and hence it cannot transform with a phase. Then r must vanish, and λ and $\bar\lambda$ are conjugate spinors with charges 1 and -1 respectively. Without taking R-charges into consideration these spinors form together a Majorana spinor (two degrees of freedom). Now they can be described either as left-handed Weyl-spinor with R-charge 1 or a right-handed one with R-charge -1. The number of degrees of freedom of a Majorana and a Weyl spinor in four dimensions is the same.

Since F-terms in the action are built by means of an integral $\int d^2\theta$, which removes two θ's, all terms in the superpotential must have total R-charge $r=2$. Then the F-term that remains has total R-charge $r-2=0$, and the action is invariant. The total R-charge of any D terms in the action must be zero.

In the absence of the unwanted terms (9.14) as well as the supersymmetric Higgs mass term (9.12) an R-symmetry of the action is $R=1$ for all left-handed chiral quark and lepton superfields, and $R=0$ for H_d and H_u. This symmetry has the nice property that all Standard Model fields (including h_d and h_u) have R-charge 0, and all squarks, sleptons, gauginos and higgsinos have R-charge ±1. It is not anomaly-free, even though all quarks and leptons have zero charge: the left-handed gauginos have charge 1, and contribute $+6$ and $+4$ to the $SU(3)\times U(1)_R$ and $SU(2)\times U(1)_R$ anomaly respectively, and the Higgsinos have charge -1 and contribute together -2 to the $SU(2)\times U(1)_R$ anomaly.

Any other assignment of R-charges differs from the previous one by some global symmetry. For example, one may choose a different R-charge assignment R' which allows (9.12) but not (9.14) by choosing $R' = R + X$ (with X as defined in the previous subsection). Then all quark and lepton superfields have R' charge $\frac{1}{2}$ and H_d and H_u have $R' = 1$. It is also possible to find a linear combination of R, X, B and L that is completely free of anomalies with respect to non-abelian groups and even with respect to gravity, as well as a combination that allows all of the terms in (9.12) and (9.14), namely $R + X - \frac{1}{2}(B - L)$.

The problem with continuous R symmetries is that the gauginos are in a complex representation of any R-symmetry and hence cannot become massive as long as the R-symmetry is exact. If we break the R-symmetry spontaneously we get a Goldstone boson. This boson is massless if the symmetry is anomaly free, and has a very small mass (like an axion) if the broken symmetry has an $SU(3)$ or $SU(2)$ anomaly. This is phenomenologically unacceptable unless one can make the axion extremely weakly coupled, *i.e.* invisible. Since there is no way of breaking the R-symmetry in the supersymmetric Lagrangian (not even if we include the unwanted B and L violating terms (9.14), we will have to worry about this problem later.

9.7 R-Parity

In most work on the phenomenology of supersymmetry the unwanted B and L violating terms are removed by imposing a discrete symmetry called R-parity. This symmetry is

$$R_p = (-1)^R = (-1)^{3(B-L)+2S} , \tag{9.18}$$

where S denotes the spin. Since S is always conserved modulo integers, R-parity is conserved if $B - L$ is conserved. The only reason for introducing the spin-dependent sign is the following convenient characterization: all Standard Model particles have R-parity $+$, while all their superpartners have R-parity $-$. Consequently, in interactions involving ordinary matter (the only interactions we can cause to happen using accelerators), superpartners are produced in pairs. Furthermore, a superpartner can only decay in another superpartner, and hence there must exist a lightest superpartner (often called "LSP") that is absolutely stable. This is one of the most important handles we have experimentally on supersymmetry.

Note that the terms (9.14) are forbidden by R-parity (indeed, they break $B - L$), whereas (9.12) is not. A gaugino mass-term is also allowed by R-parity. Therefore most phenomenology assumes that the continuous R-symmetry is in some way broken to R-parity. This is in particular part of the MSSM definition.

9.8 Supersymmetry Breaking

Supersymmetry cannot be an exact symmetry at low energies, since none of the partners of the known quarks, leptons and gauge bosons has been seen. It may seem bizarre to postulate a new symmetry when of each multiplet at most one member has been observed.

In fact, the situation is not as bad as it seems since the particles we have seen so far are precisely those whose masses are forbidden by unbroken $SU(3) \times SU(2) \times U(1)$. Indeed, the gauginos are in real representations, and can thus have a Majorana mass, the scalars are allowed to have a mass no matter what their representation is, and the Higgsinos from H_d and H_u can combine with each other to form a massive Dirac fermion. On the other hand, the masses of the gauge bosons are protected by gauge invariance, and those of the quarks and leptons by chirality. The only Standard Model particle whose mass is not protected by $SU(3) \times SU(2) \times U(1)$ is the Higgs boson, the last Standard Model particle that has been discovered. The fact that its mass is not protected is precisely the hierarchy problem. An optimistic point of view about supersymmetry is that we may just be crossing the borderline between protected and unprotected particles.

In view of this the natural course to follow is to break supersymmetry first at some scale M_S, so that all non-Standard Model particles acquire masses, and so that the Higgs scalar mass is still protected by supersymmetry cancellations, and then break the weak interaction gauge symmetries at the lower scale $M_{\rm W}$.

To break supersymmetry we have three options

- Explicit breaking
- Spontaneous breaking of global supersymmetry
- Spontaneous breaking of local supersymmetry

Explicit breaking means that supersymmetry is not an exact symmetry of nature, but just a "coincidental" property of the low-energy spectrum (where "low" means a few TeV, *i.e.* low with respect to the next higher energy scale, for example the Planck scale). In other words, perhaps nature is not fundamentally supersymmetric, but for some reason the part of the spectrum that lies well below the Planck scale consists of equal numbers of bosons and fermions for any gauge group representation. Remarkably, it is possible to break supersymmetry explicitly without loosing the good properties it has with respect to scalar masses. However, for a fundamental theory of nature this does not look very attractive, since we would never understand why these coincidences are occurring.

Spontaneous breaking of global supersymmetry has many problems, most obviously the appearance of a massless Goldstone fermion related to the broken symmetry, the "Goldstino". In addition, it is well known that if we want to couple a supersymmetric theory to gravity (which undoubtedly we will have to do), the global supersymmetry must become local.

Thus if we reject option 1, and wish to see supersymmetry as a fundamental symmetry of nature, we are inevitably led to local supersymmetry, also known as supergravity. In supergravity the Goldstino is eaten by the gravitino, a massless spin-$\frac{3}{2}$ field. If supersymmetry breaks this particle must become massive, and its number of degrees of freedom must increase from 2 to 4. The extra two degrees of freedom are provided by the Goldstino, just as the Higgs scalar contributes the extra degree of freedom needed to make a vector boson massive. This solves already the most obvious problem of spontaneously broken global supersymmetry.

The MSSM makes no statement about the kind of symmetry breaking. In all three cases, one assumes that low-energy physics is described in terms of the supersymmetry Lagrangian plus so-called soft supersymmetry breaking terms, which do not affect the cancellations due to supersymmetry. In the last two cases theses soft breaking terms are generated by the spontaneous symmetry breaking, whereas in the first case they are put in by hand.

9.9 Non-renormalization Theorems

Let us now examine the cancellations of divergences that supersymmetry produces. The result is usually phrased in the form of a non-renormalization theorem. For $N = 1$ supersymmetry, one finds that F-terms in the Lagrangian are not renormalized at all. The only radiative corrections that can appear are D-terms. The radiative corrections to $\int d^4\theta\phi\phi^\dagger$ for example yield the usual wave function renormalization of the chiral superfields. Other terms yield the gauge coupling constant renormalization.

The non-renormalization of the F-terms does not mean that the parameters of the superpotential are not renormalized. Take for example a superpotential $g\phi^3$. It leads, among others, to a contribution $g\varphi\psi\psi$ in the Lagrangian. In a non-supersymmetric theory counter-terms have to be introduced to cancel the divergences, which can be combined to

$$Z_g\sqrt{Z_\varphi}Z_\psi g\varphi\psi\psi \ . \tag{9.19}$$

Here gZ_g is the renormalized coupling, $\sqrt{Z_\varphi}\varphi$ the renormalized scalar field and $\sqrt{Z_\psi}\psi$ the renormalized fermi field. In a supersymmetric theory there are no infinities in the $g\varphi\psi\psi$ vertex diagram (after subtraction of lower order sub-divergences), and hence no counter-term is needed. Therefore $Z_g\sqrt{Z_\varphi}Z_\psi = 1$, and the coupling constant renormalization can be expressed in terms of the wave function renormalization.

A similar remark holds for the mass terms, and in particular those of the scalar fields. Here we get after introducing counter-terms $Z_m Z_\varphi m\varphi^2$, with $Z_m Z_\varphi = 1$. This expresses Z_m in terms of Z_φ. The latter has only logarithmic divergences, and no quadratic ones. Hence Z_m is free of quadratic divergences as well. Technically this comes from a cancellation between the fermion loop against the scalar loop diagram. This is precisely why one introduces supersymmetry to solve the hierarchy problem.

Supersymmetry was not supposed to make the theory finite, but it was supposed to remove quadratic divergences. There is still one term that can be quadratically divergent though, namely

$$\int d^2\theta d^2\bar{\theta}\ V \tag{9.20}$$

This yields simply the D-term of the superfield V, and is gauge invariant only if V is a vector superfield of an abelian gauge symmetry. It receives quadratically divergent corrections proportional to $\mathrm{Tr}Q$ at one loop (and at one loop only), and hence there is no problem if $\mathrm{Tr}Q = 0$ (which was also the condition for absence of gravitational anomalies). Apart from this problem (which is easy to circumvent) all corrections are logarithmic.

9.10 Soft Supersymmetry Breaking

Remarkably, the absence of quadratic divergences can be maintained even if certain terms are added to the Lagrangian that break supersymmetry explicitly. The allowed terms are

$$m_{ij}\varphi_i\varphi_j^* \ ; \quad \alpha_{ij}\varphi_i\varphi_j + \text{c.c} \ ; \quad \beta_{ijk}\varphi_i\varphi_j\varphi_k + \text{c.c} \ ; \quad \mu(\lambda\lambda + \bar{\lambda}\bar{\lambda}) \ , \tag{9.21}$$

where λ is a gaugino and φ_i a scalar field from one of the chiral multiplets; m, α, β and μ are arbitrary parameters. The most interesting terms that are *not* allowed are mass terms for the fermions in chiral multiplets, Yukawa couplings of such fermions to Higgs bosons, and fourth order scalar interactions.

Note that the second and third terms have precisely the structure of term in the super-potential, when the scalar field φ is replaced by a superfield ϕ. The conditions for invariance under global and local symmetries that commute with supersymmetry are identical for these terms. However, they appear directly in the potential, whereas the similar-looking superpotential terms lead to totally different term in the potential. The last soft breaking term does not respect continuous R-symmetries, since the gaugino transforms non-trivially under such a symmetry. Most of the terms of the second and third type will generically also violate R-symmetries.

If other terms are added to the action this leads in general to quadratic divergences, so that everything one hoped to get from supersymmetry is lost. There are exceptions however. The analysis leading to an enumeration of soft breaking terms assumes arbitrary supersymmetric theories. In a specific theory the expected disasters may not occur, and indeed there are examples of that. However, it does not seem that such potentially dangerous soft breaking terms are actually generated in spontaneous symmetry breaking.

9.11 Spontaneous Supersymmetry Breaking

Supersymmetry is broken if and only if $\langle 0| H |0\rangle \neq 0$. The operator H can be derived from the Lagrangian of the theory. Since vector fields, fermions and derivatives cannot get non-trivial vacuum expectation values without breaking Lorentz invariance, the relevant part to consider is the full scalar potential. It turns out to be convenient to express it in terms of the auxiliary fields. From Appendix D we find that the result is

$$V = F_i^* F_i + \tfrac{1}{2} D^a D^a \tag{9.22}$$

Clearly we can arrange to have $\langle 0| H |0\rangle = \langle 0| V |0\rangle \neq 0$ by having either $\langle 0| F_i |0\rangle \neq 0$ or $\langle 0| D^a |0\rangle \neq 0$ (or both). The former is called O'Raifeartaigh breaking and the latter Fayet-Illiopoulos breaking.

The supersymmetry breaking scale is defined in terms of the value of the potential after supersymmetry breaking. Since the potential has the dimension of a mass to the fourth power we define

$$M^2_{\text{Susy}} = \sqrt{F_i^* F_i + \tfrac{1}{2} D^a D^a} \tag{9.23}$$

O'Raifeartaigh breaking occurs when the superpotential is chosen in such a way that the conditions $F_i = -\partial W/\partial\varphi_i = 0$ do not have a solution for all i simultaneously. A standard example of such a superpotential is $\lambda_1\phi_1(\phi_3^2 - M^2) + \lambda_2\phi_2\phi_3$, where ϕ_i are superfields, and λ_i and M parameters.

Fayet-Illiopoulos breaking occurs when there it is not possible to have a simultaneous solution to the equation $D^a = 0$ for all a. In the absence of a ξ-term (and in particular for non-abelian fields), the condition $D^a = 0$ becomes (see Appendix D)

$$D^a = -g\varphi_i^* T_{ij}^a \varphi_j = 0 \ , \tag{9.24}$$

and can always be satisfied by setting $\varphi_i = 0$. [We are assuming here that the conditions $F_i = 0$ are trivially satisfied. In principle all conditions on D and F have to be considered, and this could still force us to have a non-trivial v.e.v. for φ_i. This might lead to a breaking of supersymmetry if the right-hand side of Eq. (9.24) is non-zero, and in any case leads to a breaking of gauge symmetry, since φ manifestly transforms non-trivially under gauge transformations.] In the presence of the ξ-term one gets the condition

$$D = -g'\varphi_i^* Q_i \varphi_i - \xi = 0 \ . \tag{9.25}$$

If the product $g'Q_i\xi > 0$ for all i this has no solution, and supersymmetry is broken. The minimum of the potential is at $\phi_i = 0$ (unbroken gauge symmetries), $D = -\xi$. If on the other hand there is a possibility for cancellation among the terms on the right-hand side of Eq. (9.25) the minimum breaks gauge-symmetry, but not supersymmetry ($D = 0$).

Note that the condition $g'Q_i > 0$ implies that all superfields coupling to the $U(1)$ symmetry under consideration must have charges with the same sign, which makes it impossible to cancel the Q^3 anomalies. A possible way out is to build a superpotential in such a way that all field with a certain sign of the charge a forced to have vanishing v.e.v's by the $F_i = 0$ conditions, so that they cannot contribute to Eq. (9.25), but this is highly contrived. Fayet-Illiopoulos symmetry breaking is thus *a priori* not a very attractive option. [Fayet-Illiopoulos symmetry breaking has however found an interesting application in four-dimensional string theory, where there is a new mechanism to cancel $U(1)$ anomalies.]

9.12 The Goldstino

Another way of looking at spontaneous supersymmetry breaking is to consider the vacuum expectation value of the anti-commutator $\{Q, \Psi\}$ where Ψ is some fermion in the theory. This object is bosonic (unlike $[Q, \text{boson}]$) and hence can in principle have a v.e.v. without breaking Lorentz invariance. If it has a v.e.v, clearly $Q\,|0\rangle \neq 0$, so that supersymmetry is broken. This implies that Q acting on the vacuum creates a fermionic state, which can be written in terms of the fermions in the theory.

The fermion Ψ can be either a member of a chiral multiplet or a gaugino. In the former case

$$\delta\psi_i = \{\alpha Q + \bar{Q}\bar{\alpha}, \psi\} = \sqrt{2}(\alpha F_i + \sigma^\mu \bar{\alpha}\partial_\mu \varphi_i) \ , \tag{9.26}$$

whereas in the latter case

$$\delta\lambda^a = \alpha\sigma^{\mu\nu} F_{\mu\nu}^a + \alpha D^a \tag{9.27}$$

In both cases only the auxiliary field can get a v.e.v, as we already know. The fermionic state created by Q out of the vacuum is then

$$\eta \;\; \propto \;\; \sum_i \langle F_i \rangle \psi_i + \frac{1}{\sqrt{2}} \sum_a \langle D^a \rangle \lambda^a \; , \tag{9.28}$$

assuming all fermi fields are orthonormal, *i.e.* the fermion propagators have residue δ_{ij} in the space of all fermions.

Just as for Goldstone bosons the Goldstino is a fluctuation around the vacuum in the direction of an exact symmetry. Hence it is a massless particle.

9.13 Mass Sum Rules

When supersymmetry is unbroken, the masses of the members of each supermultiplet are degenerate. When supersymmetry is spontaneously broken by the O'Raifeartaigh mechanism alone, there still is a relation, namely

$$\sum_S (-1)^{2S}(2S+1)M_S^2 = 0 \tag{9.29}$$

This relation holds at tree level, and can easily derived from the action. This sum rule plays an essential rôle: it guarantees the absence of quadratic divergences in the one-loop effective potential. It can be shown that those divergences are proportional to the right-hand side of Eq. (9.30).

The sum rule Eq. (9.30) holds in fact for the gauge and matter supermultiplets separately. This is bad news, especially if one hopes to break supersymmetry first and then, at a lower scale, break the weak interaction symmetries. Then the quarks, leptons and gauge bosons should remain massless after supersymmetry breaking, but the sum-rule can then only be satisfied with massless squarks, sleptons and gauginos. Even if we evaluate the sum rule including the masses of the quarks, leptons and gauge bosons after weak interaction symmetry breaking the results are disastrous. For example, the sum of the square of all 12 gaugino masses is predicted to be equal to $\frac{3}{2}(M_{\rm W}^2 + M_{\rm Z}^2)$, so that the lightest of them cannot be heavier than about 20 GeV. However, the current lower limit on the gluino mass from the Tevatron is about 135 GeV, so that the gluinos by themselves already violate the sum rule. The application of the sum rule to the matter sector is somewhat more difficult, since both the Higgs mass and the Higgsino mass are unknown, but for any reasonable guess for these masses the results are equally bad.

There are several possibilities to escape from these sum rules.

1. *Non-standard matter*
   One may add extra gauge fields which acquire a mass when supersymmetry breaks, so that there are extra contributions to the vector terms in the sum rule. Something similar has to be done in the matter sector. This is arbitrary, difficult to arrange and unattractive. To appreciate the difficulty note that the scalars giving mass to the extra gauge bosons must at the same time give more mass to the gluinos than to the extra gauginos.

2. *Fayet-Illiopoulos breaking*
   In this case the mass sum rule is modified to

$$\sum_S (-1)^{2S}(2S+1)M_S^2 = -2g'\langle D\rangle \ \mathrm{Tr}Q \tag{9.30}$$

   This is useful only if one has a $U(1)$ gauge group with a generator that is not traceless. This is thus in any case not the $U(1)$ factor of the Standard Model, so that we need to add extra gauge fields. As we have already seen, it is hard to avoid $\mathrm{Tr}Q^3$ anomalies, and manifestly impossible to avoid gravitational anomalies proportional to $\mathrm{Tr}Q$. In addition, the Fayet-Illiopoulos mechanism by itself requires a ξ term; this in combination with $\mathrm{Tr}Q \neq 0$ leads to quadratic divergences, which is what we wanted to avoid in the first place by means of supersymmetry. Thus this does not look like an attractive option either.

3. *Breaking at one loop level*
   The mass sum rules are only valid at tree level, and are subject to radiative corrections. Since the tree level sum rules are badly violated, one cannot expect radiative corrections to help much. The only way out is then to leave supersymmetry unbroken at tree level, and to break it at the loop level. This has been tried, but not with much success.

4. *Supergravity*
   The sum rule was derived for global supersymmetry. If one considers instead local supersymmetry, there is a correction proportional to the gravitino mass. The result (in the absence of Fayet-Illiopoulos breaking) is

$$\sum_S (-1)^{2S}(2S+1)M_S^2 = 2(N-1)m_{3/2}^2 \tag{9.31}$$

   Here N is the number of chiral superfields. As before, this formula is valid only at tree level. This is usually considered the most attractive way out of the sum rule problem.

9.14 The Minimal Supersymmetric Standard Model

Most of the supersymmetry phenomenology ignores the origin of supersymmetry breaking, and starts with a "low-energy" theory in which the effects of supersymmetry breaking are parametrized by soft supersymmetry breaking terms. From the considerations of the previous sections one arrives then at the so-called "minimal supersymmetric Standard Model", or "MSSM".

The MSSM is defined as follows. One has a gauge theory $SU(3) \times SU(2) \times U(1)$, and a corresponding set of vector superfields, each containing a gauge boson and a gaugino. There are three families of quarks and leptons in the representations $\mathcal{Q}, \mathcal{U}, \mathcal{D}, \mathcal{L}$, $\mathcal{E}$ and $\mathcal{N}$. For each quark and lepton flavor there are then two squarks or sleptons, usually denoted $\tilde{q}_L$

and $\tilde{q}_R$ (or $\tilde{l}_L, \tilde{l}_R$). Here L and R indicate the chirality of the quark or lepton that are the supersymmetric partners of these scalars. Furthermore there are two Higgs superfields H_d and H_u, each containing a complex Higgs scalar h_d and h_u in the representations $(1, 2, -\frac{1}{2})$ and $(1, 2, \frac{1}{2})$ respectively, and a left-handed Higgsino in the same representation.

The Lagrangian consists of the supersymmetric kinetic terms with minimal gauge coupling, plus a superpotential for the chiral superfields. This superpotential contains the three standard Yukawa coupling terms plus the scalar mass term $\mu H_d H_u$. There is an exact R-parity forbidding the other possible terms (9.14).

The soft supersymmetry breaking terms are

$$\begin{aligned}\mathcal{L}_{\text{soft}} &= -\sum_{i,j}(m^2)_{ij}(\varphi_i)^\dagger\varphi_j - \tfrac{1}{2}\sum_a M_a\bar{\lambda}_a\lambda_a \\ &+ \left[m_{ud}^2 h_d h_u + g_{\mathcal{U}}A_{\mathcal{U}}\varphi_{\mathcal{Q}}\varphi_{\bar{\mathcal{U}}}h_u + g_{\mathcal{D}}A_{\mathcal{D}}\varphi_{\mathcal{Q}}\varphi_{\bar{\mathcal{D}}}h_d + g_{\mathcal{E}}A_{\mathcal{E}}\varphi_{\mathcal{L}}\varphi_{\bar{\mathcal{E}}}h_d + \text{c.c}\right] \quad (9.32)\end{aligned}$$

Here φ_i denotes the scalar component of the superfield $i = (H_d, H_u, \mathcal{Q}, \bar{\mathcal{U}}, \bar{\mathcal{D}}, \mathcal{L}, \bar{\mathcal{E}})$; instead of φ_{H_i} we usually write h_i, and the squark and slepton fields are often denoted as $\tilde{u}_L, \tilde{u}_R$, etc. (note that $\tilde{u}_L$ is the upper component of the $SU(2)$ doublet $\varphi_{\mathcal{Q}}$, and that $\tilde{u}_R = \varphi^*_{\bar{\mathcal{U}}}$). The parameters denoted here as "$(m^2)_{ij}$" are in fact matrices in all degeneracy spaces of Standard Model representations (the square is just intended to indicate that these parameters have the dimension of a mass-squared). This means that they are 3×3 Hermitean matrices in family space for each of the Standard Model multiplets $\mathcal{Q}, \mathcal{U}, \mathcal{D}, \mathcal{L}$ and $\mathcal{E}$. This would also allow a soft breaking term of the form $\varphi_L h_u$ between the slepton doublet and a higgs (having the same structure as the $\mathcal{L}H_u$ superpotential term), but we will assume that R-parity remains unbroken, so that such a term does not appear.

The parameter m_{ud}^2 is in principle a complex number, which can be chosen real and positive by absorbing a phase in h_d (or h_u).

The parameters $g_{\mathcal{U}}$, $g_{\mathcal{D}}$ and $g_{\mathcal{E}}$ are the Standard Model Yukawa coupling matrices, which are modified by matrices $A_{\mathcal{U}}$, $A_{\mathcal{D}}$ and $A_{\mathcal{E}}$, which have the dimension of a mass.

We have ignored all neutrino contributions in the soft breaking terms, because we do not know exactly how many singlet neutrinos $\mathcal{N}$ there are. If there are three, one can add an extra term $g_{\mathcal{N}}A_{\mathcal{N}}\varphi_{\mathcal{E}}\varphi_{\bar{\mathcal{N}}}h_u$ completely analogous to the up-quark couplings. In addition there could be a supersymmetric Majorana mass matrix for the superfields $\mathcal{N}$, plus some extra soft breaking terms for the scalars in $\mathcal{N}$.

The additional parameters are then counted as follows: five 3×3 Hermitean matrices for the soft scalar masses of the squarks and sleptons, plus two masses for the two Higgses, giving a total of 47; three Majorana masses for the gauginos, plus 3 unrestricted 3×3 matrices A_x with 54 parameters, plus a real parameter m_{ud}. The total number of soft parameters is then 105, ignoring any neutrino contributions.

In principle all (or most) of these parameters are determined by the supersymmetry breaking mechanism, and for example in supergravity models one usually finds that they are determined by a much smaller number of input parameters. Nevertheless, if one really wants to compare the MSSM as defined so far to the data in a supersymmetry-breaking-independent way, one should keep all these parameters.

This is a fairly hopeless task, and what one usually does is make some additional "unification" assumptions. One assumes relations among these parameters at some high scale Λ_U, and then one uses renormalization group evolution to derive the low-energy parameters. These assumptions may include gauge coupling unification *à la* $SU(5)$, universal gaugino masses ($M_a = m_{1/2}$, for all a), universal scalar masses ($(m^2)_{ij} = m_0^2\delta_{ij}$), and universal tri-linear couplings ($A_x = m_0 A\mathbf{1}$, for all x). If one makes all these assumptions, the set of parameters of the soft terms is reduced to $g_{\mathcal{U}}, g_{\mathcal{D}}, g_{\mathcal{E}}, m_{1/2}, m_0^2, m_{ud}^2$ and A. The latter four, plus the parameter μ, are then the parameters which are added to the Standard Model by supersymmetry. Note that the Standard Model parameters in the Higgs potential, μ^2 and λ, are not present in the MSSM; the parameter μ in the super-potential term $\mu H_u H_d$ should not be confused with one in the Standard Model potential term $\mu^2\phi^\dagger\phi$. The complete set of parameters of the MSSM consists of the unified gauge coupling g plus $g_{\mathcal{U}}, g_{\mathcal{D}}, g_{\mathcal{E}}, m_{1/2}, m_0^2, m_{ud}^2, A$ and μ. Note that neither neutrino masses nor strong CP violation have been taken into account.

For all the foregoing assumptions one can give more or less convincing arguments, of two types: either they hold in a certain class of models, or violating them would in general have undesirable phenomenological consequences (some of these will be discussed later).

The equality of the gaugino and scalar masses is not as unreasonable as it may seem at first sight, if we imagine that supersymmetry breaking is an effect involving (super)gravity interactions. With respect to gravity all matter is on equal footing, and hence it would not be a total surprise if all chiral multiplets and all vector multiplets, regardless of their gauge properties, experience the same supersymmetry breaking. Since gravity *is* sensitive to differences in spin, it is also not unreasonable that gaugino masses and scalar masses come out different. The relations among the tri-linear couplings are less easy to understand from this point of view.

In addition one sometimes assumes the $SU(5)$-inspired relation $g_{\mathcal{D}} = g_{\mathcal{E}}$ or one introduces a parameter B so that $m_{ud} = m_0\mu B$, which replaces m_{ud}. The dimensionless parameters A and B play a similar rôle in the sense that both are appearing as factors of terms that also appear in the super-potential, but that now appear in the potential as soft breaking terms. Note that the term $\mu H_d H_u$ in the super-potential leads to a term $\mu^2(h_d^2+h_u^2)$ in the potential. The corresponding soft breaking term is $\mu B h_d h_u$, and appears directly in the potential. Sometimes B is eliminated as a free parameter by imposing the relation $B = A - 1$, an assumption inspired by a simple supergravity model, which we will discuss later.

With these five parameters, the MSSM really has some predictive power, but unfortunately it can never be ruled out completely convincingly with these restrictions.

9.15 The Higgs Potential

In the MSSM supersymmetry is already broken, but $SU(2)\times U(1)$ is not. The supersymmetric contribution to the Higgs potential is derived as follows. One can express the scalar potential entirely in terms of auxiliary fields (see appendix D for further details):

$$V(h_d, h_u) = F_d^* F_d + F_u^* F_u + \tfrac{1}{2}D^i D^i|_{SU(2)} + \tfrac{1}{2}D^2|_{U(1)} \qquad (9.33)$$

Note that there are contributions from the superpotential, as one might expect, but also from the D-terms that yield the kinetic terms of the scalars. This potential can now be computed as follows. We will need the solution for the auxiliary fields D given in Eq. (D.73): $D^a = -g\varphi_i^* T^a_{ij}\varphi_j$. Furthermore we need the analogous solutions for F.*

$$F_d = -\mu h_u \; ; F_u = -\mu h_d \tag{9.34}$$

Since H_d and H_u are both in the doublet representation of $SU(2)$ one has, using the correct normalization for the generators $T^i = \frac{1}{2}\sigma^i$:

$$D^i = -\tfrac{1}{2}g_2 h_d^\dagger \sigma^i h_d - \tfrac{1}{2}g_2 h_u^\dagger \sigma^i h_u \ , \tag{9.35}$$

and since they have opposite charges $\pm\frac{1}{2}$ the Y-charge D-terms contribute:

$$D = \tfrac{1}{2}g_1 h_d^\dagger h_d - \tfrac{1}{2}g_1 h_u^\dagger h_u \ . \tag{9.36}$$

Hence the Higgs potential has the form

$$V(h_d,h_u) = |\mu|^2|h_d|^2 + |\mu|^2|h_u|^2 + \tfrac{1}{8}g_1^2\left(|h_d|^2 - |h_u|^2\right)^2 + \tfrac{1}{8}g_2^2(h_d^\dagger\vec{\sigma}h_d + h_u^\dagger\vec{\sigma}h_u)^2 \ , \tag{9.37}$$

where $|h_d|^2 \equiv h_d^\dagger h_d$ Using the identity

$$\sigma^i_{\alpha\beta}\sigma^i_{\gamma\delta} = 2\delta_{\alpha\delta}\delta_{\beta\gamma} - \delta_{\alpha\beta}\delta_{\gamma\delta} \ , \tag{9.38}$$

we can also write this as

$$\begin{aligned} V(h_d,h_u) \;\; = \;\; & |\mu|^2|h_d|^2 + |\mu|^2|h_u|^2 \\ & + \tfrac{1}{8}(g_1^2+g_2^2)(|h_d|^2 - |h_u|^2)^2 + \tfrac{1}{2}g_2^2|h_d^\dagger h_u|^2 \ . \end{aligned} \tag{9.39}$$

Note that the quartic terms are determined entirely in terms of gauge couplings, and that there is no free four-scalar coupling constant associated with them. This potential has manifestly positive quadratic terms, and hence there is no possibility to break $SU(2)\times U(1)$. But we still have to add the soft supersymmetry breaking terms. Including them, one gets

$$\begin{aligned} V(h_d,h_u) \;\; = \;\; & \mu_d^2|h_d|^2 + \mu_2^2|h_u|^2 - (m_{ud}^2 h_d h_u + \text{c.c}) \\ & + \tfrac{1}{8}(g_1^2+g_2^2)(|h_d|^2 - |h_u|^2)^2 + \tfrac{1}{2}g_2^2|h_d^\dagger h_u|^2 \ , \end{aligned} \tag{9.40}$$

where $\mu_d^2 = |\mu|^2 + m_{h_d}^2$ and $\mu_u^2 = |\mu|^2 + m_{h_u}^2$. The importance of supersymmetry breaking is that now these parameters can be negative. Note that $SU(2)$ indices are suppressed here. In order to get an $SU(2)$-invariant, the explicit form of $h_d h_u$ must be $h_d h_u = h_d^\alpha h_2^\beta \epsilon_{\alpha\beta}$.

It should be emphasized that the positivity of $|\mu|^2$ is independent of perturbative corrections. We will see later that the parameters μ_d^2 and μ_u^2 may be positive at some scale, and then evolve to negative values at some lower scale. This would not be possible

* Here we are ignoring all scalar fields except h_d and h_u.

if supersymmetry were unbroken. Then the superpotential, from which Eq. (9.39) is derived, is not renormalized, and hence the form of (9.39) cannot change.

However, this is not the most general Higgs potential one can write down with two scalar fields h_d and h_u. The most general one has the same set of quadratic terms, but has the following quartic terms

$$\begin{aligned}&\lambda_1(h_d^\dagger h_d)^2+\lambda_2(h_u^\dagger h_u)^2+\lambda_3 h_d^\dagger h_d h_u^\dagger h_u+\lambda_4|h_d^\dagger h_u|^2\\ +&\left[\lambda_5(h_d h_u)^2+\lambda_6 h_d^\dagger h_d(h_d h_u)+\lambda_7 h_u^\dagger h_u(h_d h_u)+\text{c.c}\right]\ ,\end{aligned}\tag{9.41}$$

As usual $SU(2)$ invariant contractions are not explicitly indicated. The Higgs potential in the MSSM satisfies the additional constraints $\lambda_5=\lambda_6=\lambda_7=0$, and $\lambda_1=\lambda_2=-\frac{1}{2}\lambda_3$. There is no symmetry one can impose to enforce such a relation. For example, the interchange $h_d\leftrightarrow Ch_u^*$ would explain one of these relations, but because of the Yukawa couplings this cannot be a symmetry of the MSSM. For the same reason one cannot impose a symmetry $h_i\rightarrow -h_i$ $(i=u,d)$ to get rid of the last two terms. These constraints are in fact due to supersymmetry. For example, the term with coefficient λ_5 does not appear because it does not come from the supersymmetric part of the action, nor is it a soft term.

This has several consequences. First of all the special form of the potential ensures that the Higgses h_d and h_u align correctly. A potential danger of a two-Higgs potential with two Higgses that have to get a non-trivial vacuum expectation value (as is the case here) is that the two Higgses choose an "arbitrary" direction with respect to each other. Then $SU(2)\times U(1)$ does not break to $U(1)_{\rm em}$ but to nothing at all, and the photon gets a mass. Even an extremely small misalignment would clearly be fatal. Let us assume that the mass parameters in the potential are such that the two Higgses do indeed get a non-trivial vacuum expectation value. Using $SU(2)\times U(1)$ rotations we may bring the $\langle h_d\rangle$ to the form

$$\frac{1}{\sqrt{2}}\begin{pmatrix}v_d\\0\end{pmatrix}\ .\tag{9.42}$$

Note that h_d has $U(1)_Y$ charge $Y=-\frac{1}{2}$, so that with this choice the vacuum has charge $Q_{\rm em}=T_3+Y=0$. The correct alignment of h_u $(Y=\frac{1}{2})$ is then

$$\langle h_u\rangle=\frac{1}{\sqrt{2}}\begin{pmatrix}0\\v_u\end{pmatrix}\ .\tag{9.43}$$

However, let us assume that h_u is misaligned by an arbitrary $U(2)$ rotation. This can be parametrized by choosing

$$\langle h_u\rangle=\frac{1}{\sqrt{2}}\begin{pmatrix}v_u\ e^{i\alpha}\sin\gamma\\ v_u\ e^{i\eta}\cos\gamma\end{pmatrix}\ .\tag{9.44}$$

The terms in the potential that depend on the orientation are

$$\begin{aligned}h_d^\dagger h_u &= v_d v_u e^{i\alpha}\sin\gamma\\ h_d h_u &= \epsilon_{ij}h_d^i h_u^j = v_d v_u e^{i\eta}\cos\gamma\end{aligned}$$

Substituting this into the Higgs potential (9.40), but with the more general quartic interactions shown in (9.41), we get

$$\begin{aligned} V(v_d, v_u, \alpha, \eta, \gamma) \;\; = \;\; & \mu_d^2 v_d^2 + \mu_u^2 v_u^2 - 2m_{ud}^2 v_d v_u \cos\eta \cos\gamma \\ & + \lambda_1 v_d^4 + \lambda_2 v_u^4 + \lambda_3 v_d^2 v_u^2 + v_d^2 v_u^2 \lambda_4 \sin^2\gamma \\ & + 2v_d^2 v_u^2 \lambda_5 \cos 2\eta \cos^2\gamma + 2(\lambda_6 v_d^3 v_u + \lambda_7 v_d v_u^3) \cos\eta \cos\gamma \end{aligned}$$

If $\lambda_5 = \lambda_6 = \lambda_7 = 0$ and $\lambda_4 > 0$, as is the case in Eq. (9.40), the quartic terms are minimized for $\sin\gamma = 0$, and the quadratic ones for $\cos\eta = \cos\gamma = \pm 1$ (if $m_{ud}^2 v_d v_u > 0$) or $\cos\eta = -\cos\gamma = \pm 1$ (if $m_{ud}^2 v_d v_u < 0$). No matter how we choose the signs, the solutions are always $\gamma = 0 \bmod \pi$ and $\eta = 0 \bmod \pi$, so that h_d and h_u are indeed aligned properly. For the general potential the minimization is more complicated. For example, if we change the sign of λ_4 and keep $\lambda_5 = \lambda_6 = \lambda_7 = 0$, the minimum occurs for a non-trivial value of γ due to competition between the quadratic and quartic terms. In general there are regions in parameter space where the true minimum respects $U(1)_{\rm em}$, and hence the alignment is not something unnatural even for the full potential. Fortunately the extra constraints due to supersymmetry put us precisely in a region of parameter space where the alignment is automatic.

9.15.1 A Weak Symmetry Breaking Minimum

The Higgs potential (9.40) has an interesting feature: the quartic terms vanish if one choose $h_u = e^{i\alpha} C h_d^*$ for any phase α. This means that if we vary the fields along this direction, the quartic terms cannot guarantee that the potential is bounded from below. Along this direction in parameter space, the quadratic terms are equal to

$$(\mu_d^2 + \mu_u^2 - 2m_{ud}^2 \cos\alpha)|h_d|^2 \ , \tag{9.45}$$

and hence we see that there is a positivity condition, to ensure that the potential is bounded from below in the limit $|h_d| \to \infty$:

$$\mu_d^2 + \mu_u^2 \geq 2|m_{ud}^2| \tag{9.46}$$

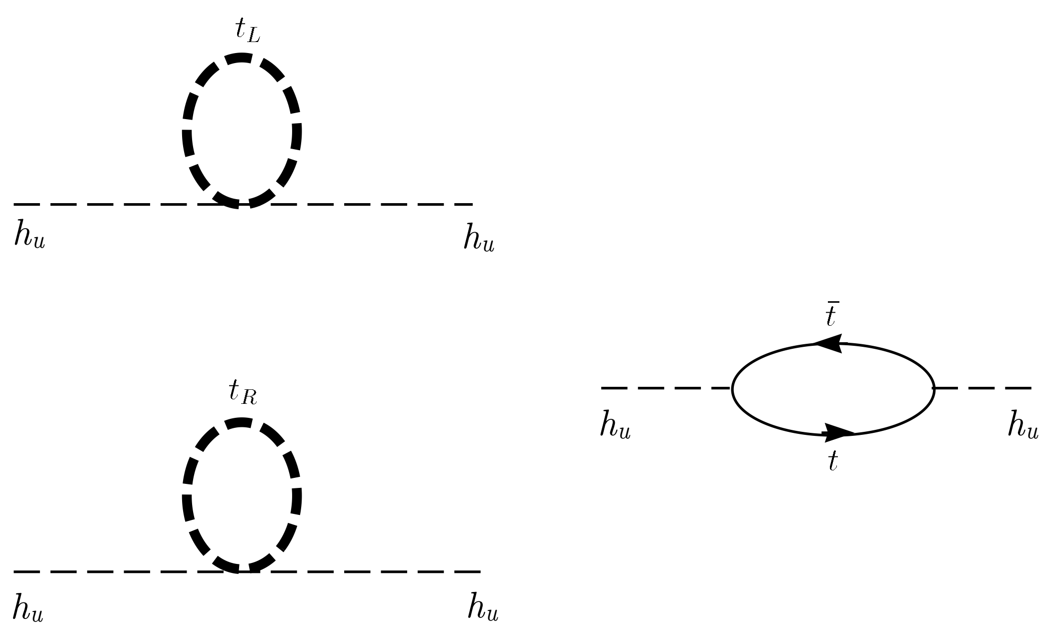
The condition for the occurrence of symmetry breaking is that the mass matrix of h_d and h_u has a negative eigenvalue. The existence of a single negative eigenvalue is equivalent to the requirement that the determinant be negative:

$$|m_{ud}^2|^2 > \mu_d^2 \mu_u^2 \tag{9.47}$$

Of course this is not relevant if both eigenvalues are negative, but usually one is interested in a situation where the determinant is positive at high energies, and changes sign when evolved to lower energies.

If we make the "unification" assumption $m_u^2 = m_d^2$ (universal soft scalar masses) , it follows that at the unification scale Λ_U $\mu_d^2 = \mu_u^2$. Then conditions (9.46) and (9.47) can just not be satisfied: choosing $m_{ud}^2 = \mu_d^2 = \mu_u^2$ saturates both inequalities. The potential is flat along the lines $h_u = e^{i\alpha} C h_d^*$, and there is no symmetry breaking, but all vacua along these lines are exactly degenerate.

Now what happens if we evolve these parameters to lower mass scales? Note that the Higgs potential looks symmetric in h_d and h_u, but the Yukawa couplings are not. Most importantly, h_u couples to the top quark, and h_d does not. Since the top quark is very heavy, it has a large Yukawa coupling, and this coupling turns out to dominate the evolution. Some of the contributing diagrams are



In the supersymmetric limit they would exactly cancel, so that m_u^2 and m_d^2 are renormalized only by wave function renormalizations. Both Higgs masses are equal to μ^2 in the supersymmetric limit, since they both come from the superpotential term $\mu H_d H_u$. Radiative corrections may change the value of μ, but not the form of the superpotential, nor the resulting equality $m_u = m_d$. But once supersymmetry is broken the scalar in the loop gets a mass, while the fermion remains massless. This suppresses the positive scalar contribution with respect to the negative fermion contribution. Hence the net effect of this contribution is to drive the scalar mass of the external lines to lower values. It is then possible that even if condition (9.47) is not satisfied at the higher scale, it is satisfied at a lower one.

There is a competing effect due to non-cancellation of the gauge boson and gaugino contributions. In this case a fermionic contribution, namely that of the gaugino, is suppressed, and hence in this case the effect is precisely opposite. If the Yukawa coupling is sufficiently large the first effect will be larger than the second, and the mass will indeed decrease with decreasing energy.

But there is a further complication. The lines in the diagrams shown above can be interchanged to get corrections to the masses of t_L due to t_R and h_u, and corrections to t_R due to t_L and h_u. Hence there is an effect not only for m_2, but also for $m_{\tilde{t}_L}$ and $m_{\tilde{t}_R}$. Now we would like the scalar m_2 to get a v.e.v. and certainly not the stop squarks, since this would break color. There is an intuitive way to see which effect will win, namely by considering the fermion loop (to which all diagrams are proportional in the supersymmetric limit). In the correction to m_2 there is a color triplet loop (formed by t_L and t_R), which gives a factor 3; in the correction to $m_{\tilde{t}_R}$ there is an $SU(2)$ doublet loop (the higgsino and t_L), which gives a factor 2, and finally the correction to $m_{\tilde{t}_L}$ involves a color and $SU(2)$ singlet loop (the higgsino and t_R), so that there is no enhancement. Hence h_u receives the largest contribution, and if there is any mass^2 that changes sign it will be that of h_u.

The net result of a quantitative calculation is the following set of renormalization group equations (for $t = \log Q$, and Q the energy scale)

$$\frac{dm_i^2}{dt} = \frac{1}{8\pi^2}\left[-\sum_{a=1,2,3} c_a(i) g_a^2 M_a^2 + c_i g_t^2 (m_{\tilde{t}_L}^2 + m_{\tilde{t}_R}^2 + m_2^2 + A_t^2)\right] \tag{9.48}$$

where only the top quark and its superpartners are taken into account. Here m_a and A_t are parameters appearing in Eq. (9.32), g_t the top Yukawa coupling, $c_a(i)$ is a set of numerical coefficients, and so is c_i. The most important point is that $c_i = 3$ for h_u, $c_i = 2$ for t_R and 1 for t_L. If the second term dominates, and the masses of all scalars are equal at some scale, then all masses will decrease with decreasing t. But due to the factor $c_i = 3$ the mass of h_u will go through zero before any of the others. When that happens (actually already earlier, namely when condition (9.47) is satisfied) $SU(2)\times U(1)$ breaks, and many particles acquire a mass and decouple from the renormalization group equations. Hence below this scale these equations show a different behavior, and in particular it is possible that none of the other masses goes through zero.

All of this is hand-waving, and a detailed study of the full set of coupled equations is required to show that indeed this mechanism works. This is quite complicated, even under the drastic simplifications of the unification conditions on the masses. A more detailed analysis does appear to show that indeed regions in parameter space exist where this mechanism could work.

9.16 Higgs Masses

It is straightforward to diagonalize the mass-matrix of the Higgses after $SU(2) \times U(1)$ breaking. Altogether the Higgs system h_d and h_u contains eight real degrees of freedom. Three of them are eaten by the $W^\pm$ and Z bosons, leaving five physical Higgs scalars.

Four modes of the Higgs system are in the direction of the vacua $\langle h_d\rangle$ and $\langle h_u\rangle$, corresponding to two two phases and two scale factors of these vacua. These four fields all have zero electromagnetic charge $T_3{+}Y$. The relative phase degree of freedom corresponds to the longitudinal component of the Z_0, because h_d and h_u have opposite $U(1)$-charges. This leaves one boson corresponding to the *common* phase degree of freedom, called A^0, and two corresponding to the scalings of the two vacua, called h^0 and H^0 (the lighter one being h^0). The field A_0 is precisely the axion discussed in chapter 5. The m_{ud} breaks the PQ-symmetry discussed there, and hence the "axion" mass will be proportional to it. In a one-Higgs system there is of course no particle like A^0, and only one rescaling, corresponding to the Standard Model Higgs boson.

The Higgs potential is almost symmetric under the CP-symmetry $h_i \leftrightarrow h_i^\dagger$. The only possible violation could be the term proportional to m_{ud}^2, if that parameter is not real (in general it could be any complex number, although the notation might suggest otherwise). However, we can always make m_{ud}^2 real and positive by a relative phase rotation of h_d and h_u.* We could just as well have called this symmetry C , and assign positive parity to both h_d and h_u, since parity is manifestly a symmetry of the Higgs action (provided both Higgses are assigned the same parity). However, both C and P are badly broken when we couple the Higgses to the fermions, and CP is a symmetry to a quite good approximation. It follows then that A^0 is CP-odd and h^0 and H^0 are CP even. Since the Higgs potential with real m_{ud}^2 is CP-invariant the mass matrix will not mix CP-odd and CP-even states, and furthermore any radiatively induced mixing is proportional to the CP-violating terms in the full action, and hence probably quite small. (Unless there are large CP-violating terms that do not manifest themselves in our present experiments).

From the condition that the higgs v.e.v's are a local minimum of the action one derives rather easily

$$\mu_d^2 + \mu_u^2 = m_{ud}^2 \frac{v_d^2 + v_u^2}{v_d v_u} \tag{9.49}$$

To see this, act with the differential operator $(v_d\partial_{v_u} + v_u\partial_{v_d})$ on the potential V shown in Eq. (9.40), with the vacuum expectation values substituted. In a local minimum, $(v_d\partial_{v_u} + v_u\partial_{v_d})V$ must be zero. The potential has the form (note that $h_d^\dagger h_u = 0$ because of the vacuum alignment):

$$V(v_d, v_u) = \mu_d^2 v_d^2 + \mu_2^2 v_u^2 - 2m_{ud}^2 v_d v_d + \tfrac{1}{8}(g_1^2 + g_2^2)(v_d^2 - v_u^2)^2$$

The differential operator annihilates the quartic terms, and requiring that it vanishes on the quadratic terms yields Eq. (9.49).

If m_{ud}^2 has been chosen positive one may assume without loss of generality that both v_d and v_u are positive. It is customary to define

$$\tan\beta \equiv \frac{v_u}{v_d} \; . \tag{9.50}$$

Furthermore one has

$$M_{\rm W}^2 = \tfrac{1}{4} g_2^2 (v_d^2 + v_u^2) \; , \tag{9.51}$$

which fixes the value of $(v_d^2 + v_u^2)$ to the usual value (246 GeV)^2.

The quartic terms in the potential are completely independent of the field A^0. Its mass is thus independent of g_1 and g_2, and one finds

$$m_{A^0}^2 = \frac{m_{ud}^2}{\cos\beta\sin\beta} = m_{ud}^2 \frac{v_d^2 + v_u^2}{v_d v_u} \; . \tag{9.52}$$

* Note that this is not true for the full Higgs potential (9.41), since one cannot simultaneously remove the phases in λ_3, λ_6 and λ_7. Indeed, the two-Higgs model has been proposed as a model for CP-violation.

This relation is often used to replace the parameter m_{ud}^2 by the directly measurable quantity $m_{A^0}^2$.

The two CP-even states mix with each other. Their mass matrix does depend on the quartic terms in the Higgs potential, but only on the terms proportional to $g_1^2 + g_2^2 = 4M_{\rm Z}^2/(v_d^2+v_u^2)$. The mass matrix can in fact be expressed completely in terms of $M_{\rm Z}, m_{A^0}$ and β. The eigenvalues are

$$m_{H^0,h^0}^2 = \tfrac{1}{2}\left(m_{A^0}^2 + M_{\rm Z}^2 \pm \sqrt{(m_{A^0}^2 + M_{\rm Z}^2)^2 - 4m_{A^0}^2 M_{\rm Z}^2 \cos^2 2\beta}\right) \tag{9.53}$$

From this relation we find immediately that the lightest particle, h^0 has a mass that is less than that of the Z-boson!

This prediction is a consequence of the fact that the quartic terms in the tree level potential are completely determined by the gauge couplings, whereas in the Standard Model there is a free parameter λ.

The remaining four degrees of freedom of the Higgs system are charged. Two of them are absorbed by $W^\pm$, whereas the remaining two form a charge conjugate pair $H^\pm$ whose masses are easily found to be equal to $M_{\rm W}^2 + m_{A^0}^2$.

9.17 Corrections to the Higgs Masses

All results obtained so far are based upon the tree level potential, and are subject to radiative corrections. These are computed either by first calculating the one-loop effective potential or directly by diagrammatic calculation. There are at least two important things to check: is the alignment of the Higgses respected, and does the tree-level result $m_{h^0} < M_{\rm Z}$ survive.

The statement that the alignment is respected is equivalent to the statement that the original vacuum remains stable in charged directions, in other words that $H^\pm$ do not develop a v.e.v. This in its turn implies that the mass M^2 of the charged Higgses must remain positive. At tree level these masses are larger than that of the W-boson, so that it appears implausible that radiative corrections would make them negative. In other words, we saw before that the potential forces alignment of H_d and H_u if $\lambda_4 > 0$. In the MSSM, this parameters is $\frac{1}{2}g_2^2$, and hence comparable in size to the other quartic terms. It is unlikely that radiative corrections would lead to a sign flip of a parameter of order 1, such as λ_4/λ_1. Explicit computations confirm this: there are corrections that increase the masses of the charged Higgses, and corrections that decrease them. Only in rather extreme limits of parameter space $m_{H^\pm}^2$ could in principle be negative, but for generic parameter values it receives fairly moderate corrections.

On the other hand, for most of the parameter space the result $m_{h^0} < M_{\rm Z}$ does not survive. There are rather large radiative corrections to $m_{h^0}^2$ proportional to $\alpha m_t^4/M_{\rm W}^2$:

$$\Delta M^2 = \frac{3}{8\pi^2}\frac{g_2^2 m_t^4}{M_{\rm W}^2 \sin^2\beta}\log(1+\frac{m_0^2}{m_t^2}) \ , \tag{9.54}$$

where m_0 is the universal squark mass.* Since the top quark mass is large, these corrections can be considerable, and for most of the parameter space they push m_{h^0} above $M_{\rm z}$. Just to give an example: for $m_t = 175$ GeV and $m_0 = 100$ GeV one finds that the absolute maximum for the light Higgs mass is about 95 GeV. If we choose m_0 equal to 1 TeV, as it might well be, the maximum increases to 130 GeV; for $m_0 = 5$ TeV (considered a very high value) one gets 150 GeV. Here the maximum value is obtained by maximizing with respect to all the other parameters on which m_{h^0} depends, in particular m_{A_0} and $\tan\beta$. These maxima were discussed for about two decades, until a Higgs scalar was finally discovered in 2012, with a mass of about 126 GeV, close to the upper limit for $m_0 = 1$ TeV. However, the complete story is far more complicated, and in addition one has to take into account that no evidence for supersymmetry has been found. The consensus is that the 126 GeV Higgs mass puts supersymmetry under stress, but does not rule it out. Indeed, the fact that the supersymmetry breaking scale is higher than expected reduces the tension.

9.18 Neutralino Masses

The photino, Zino, and the two Higgsinos all have the same $SU(3) \times U(1)$ quantum numbers, and hence they mix (if lepton number is violated they can of course also mix with the neutrinos). The mass matrix is

$$\begin{pmatrix} M_1 & 0 & -M_{\rm z}\cos\beta\sin\theta_{\rm w} & M_{\rm z}\sin\beta\sin\theta_{\rm w} \\ 0 & M_2 & M_{\rm z}\cos\beta\cos\theta_{\rm w} & M_{\rm z}\sin\beta\cos\theta_{\rm w} \\ -M_{\rm z}\cos\beta\sin\theta_{\rm w} & M_{\rm z}\cos\beta\cos\theta_{\rm w} & 0 & -\mu \\ M_{\rm z}\sin\beta\sin\theta_{\rm w} & M_{\rm z}\sin\beta\cos\theta_{\rm w} & -\mu & 0 \end{pmatrix}, \tag{9.55}$$

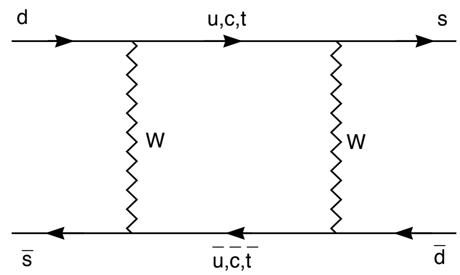
where M_1 is the gaugino mass in the $U(1)$ factor of $SU(3) \times SU(2) \times U(1)$, and M_2 the gaugino mass in the $SU(2)$ sector.

Perhaps the most noteworthy feature of this matrix is that its determinant is proportional to μ, so that there is a zero eigenvalue if $\mu = 0$.

9.19 Rare Processes

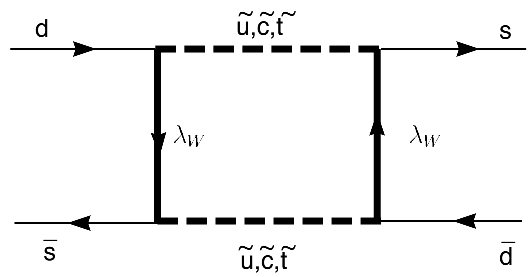
Several low energy processes put strong constraints on the parameter space of supersymmetric models. Consider for example the mass difference between the two neutral K-mesons. These particles are the two mass eigenstates in the $K^0 = \bar{s}d$ and $\bar{K}^0 = s\bar{d}$ system. These strong-interaction eigenstates are mixed by the weak interactions, the relevant diagrams being

* This is the correction to one of the diagonal entries of the 2×2 mass matrix, namely the h_u-h_u entry; since h_d does not couple to the top quark, the other entries receive negligible corrections.



These diagrams cancel if all intermediate quarks u, c and t are degenerate in mass. Since they are not, there is a non-vanishing mixing matrix element between the strong interaction eigenstates. In general this amplitude violates CP, but in the excellent approximation that CP is conserved the new mass eigenstates are the CP-eigenstates $K^0_L = \frac{1}{\sqrt{2}}(\bar{K}^0 + K^0)$ and $K^0_S = \frac{i}{\sqrt{2}}(\bar{K}^0 - K^0)$ (the subscripts stand for "Long" and "Short", referring to the rather different lifetimes of these particles. This is due to the fact that the final states allowed by CP are different).

In any supersymmetrized version of the Standard Model there are additional diagrams. The following ones are sensitive to the up squark masses

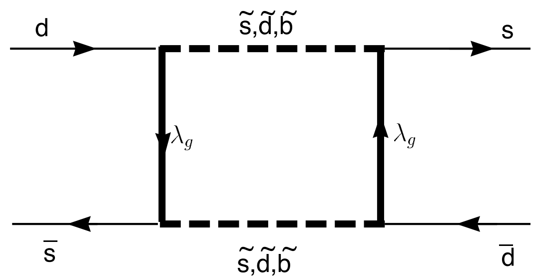


These diagrams yield the following bound on the mass differences

$$\frac{1}{M^2}\left(\frac{\Delta\tilde{m}^2_{\mathcal{U}}}{\tilde{m}^2_{\mathcal{U}}}\right) < 10^{-7}\ \mathrm{GeV}^{-2}\ , \tag{9.56}$$

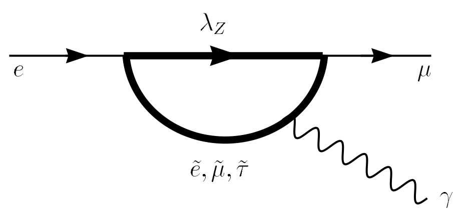
where M is the maximum of the Wino mass and the up squark masses, $\tilde{m}_{\mathcal{U}}$. In this result it was assumed that the relevant mixing angles are precisely the same as those of the quarks, *i.e.* the CKM matrix. The amplitude necessarily involves off-diagonal elements of that matrix, since we have to go from d to s quarks. The corresponding matrix for quarks does not have to be the same as those of quarks, but it it would be equal if the up squark and up quark mass matrices are diagonalized by the same matrices. But note that the off-diagonal CKM matrix elements are rather small, and it would be wishful thinking to assume that the corresponding elements for squarks are even smaller. So the assumption made here about the size of the matrix elements is a conservative one. For quarks, $\Delta m^2_{\mathcal{U}}$ is dominated by the top quark, so that this quantity is essentially equal to $m^2_{\mathcal{U}}$. If the ratio within parentheses is close to 1 also for the squarks, it means that the scale M must be 3 TeV or more, which seem rather large in comparison to the weak scale. This is an argument in favor of the unification assumption that all squarks masses are equal at the unification scale. Then renormalization group corrections will still generate differences, but it is possible that those differences are small enough.

A similar limit on the down squarks is obtained from the same diagrams with the Winos replaced by gauginos.



Here again the result depends on unknown mixing angles. If the down quarks and down squarks are diagonalized by the same matrix, the gluino-quark-squark coupling is flavor diagonal, and the diagram vanishes. Note that precisely this assumption was used above for arguing that the Wino-quark-squark coupling might be identical to the CKM matrix. But it is not plausible that an exact cancellation occurs for the Wino diagrams discussed above, since there is no reason why the down quark and up squark mass matrices could be related. The bound for the down squarks is more stringent since the relevant coupling is α_s instead of α_w, but unfortunately rather model-dependent due to the unknown angles.

Other constraints come from the flavor changing neutral current processes $\mu \to e\gamma$, $K_L \to \mu\mu$ and others. A diagram contributing to $\mu \to e\gamma$ is



This diagram vanishes if the Zino-lepton-slepton coupling is flavor diagonal. This gives a bound similar in order of magnitude to the previous ones, for some “reasonable” assumption regarding the mixing angles.

Although no definite conclusions can be reached, it seems that in any case the diagonalization of the fermion and sfermion mass matrices must not be vastly different, or that the sfermions should be nearly degenerate in mass. If the mass matrix of the fermions is equal to that of the fermions plus an arbitrary diagonal matrix this condition is automatically fulfilled.

Note that it is not true that strangeness changing neutral currents only annoy technicolor model builders. The main difference is that in supersymmetric models one has considerably more freedom, especially as long as one has not settled on a supersymmetry breaking mechanism.

9.20 Direct Searches

One of the predictions of most supersymmetric models, and in particular the MSSM, is the existence of a stable particle, whose decay is forbidden by an exact R-parity. This particle, the "LSP" (lightest supersymmetric particle), is an obvious signal to look for in present and future experiments. It also has a second important consequence: it contributes to the matter density of the universe without being easily visible, and hence is a "dark matter" candidate. Depending on the mass and the interactions of the LSP, and on the density one assumes the universe should have, one can obtain bounds on such particles. There is a lot to be said about cosmological bounds on stable particles, but we will not discuss this important topic here.

To give honest model-independent bounds on supersymmetric particle masses is extremely complicated. Most limits quoted in the literature involve many stated or unstated assumptions. One of these assumptions is the precise version of the MSSM one is using. To give bounds on supersymmetric particles independent of the assumptions of the MSSM is essentially impossible.

One usually assumes that the LSP is a neutral particle. Compare for example the gluino and the Wino, Zino masses. At the unification scale they are assumed to be equal to each other. Then the renormalization group effects are such that at any lower scale $M_i/M_j = g_i^2/g_j^2$. Since the strong coupling constant is much larger than the $SU(2)$ coupling constant, one expects the gluino to be heavier than the $SU(2) \times U(1)$ gauginos. Note that this uses the MSSM assumption that all gauginos are born with equal mass. To discriminate between Winos and other neutralinos is less straightforward, since the latter have a complicated mass matrix, mixing them with the Higgsinos which furthermore contribute an additional free parameter μ. Among the sfermions one expects the sneutrinos to be the lightest particles, for similar reasons.

The LSP can then either be the gravitino a sneutrino, or a linear combination of the two higgsinos, the photino and the Zino. If it is the gravitino there is a next-to-LSP which lives extremely long, but decays ultimately to the gravitino, to which it couples only very weakly. Hence for all practical purposes we may then ignore the gravitino, and regard the next-to-LSP as the LSP.

The celebrated LEP-result that limits the number of neutrinos species to 3 also has implications for the LSP. If it couples with reasonable strength (*i.e.* like a neutrino) to the Z-boson and has a mass less than about 40 GeV it contributes to the width of the Z. This is the present limit on most neutral superpartners (stable or not), but the LSP could have escaped observation if its coupling to the Z is much smaller than that of a neutrino. The limit is then much worse, about 15 GeV only.

Squarks and gluinos will be produced in pairs in hadron colliders. They will decay into quarks and gluons plus the LSP, possibly in several steps. The details of their decays are model dependent, but the fact that among the decay products there is an LSP follows from R-parity. If it is neutral, the LSP cannot be seen directly, but it carries transverse momentum. The signal to look for is thus a multi-jet event with a large amount of missing transverse momentum that cannot be attributed to neutrinos. Such signals have been looked for at FermiLab, and since nothing was seen one obtains a limit, which is about 100 GeV for squarks and 200 GeV for gluinos. These limits are based on several assumptions, but I will not elaborate on that.

Charged sleptons are much harder to see in hadron colliders due to the small production cross section. However, the limits from Z decay mentioned above apply, and indeed the present limits are about 45 GeV. Unfortunately these limits are also dependent on some assumptions. The same limits apply to charged Wino-Higgsino mixtures.

9.21 Supersymmetric Unification

The minimal supersymmetric grand unified theory is the straightforward supersymmetrization of the $SU(5)$ GUT. The particle content consists of superfields $\mathcal{Y}_i$ in the representation $\mathbf{5}^*$ and $\mathcal{X}_i$ in the $\mathbf{10}$, where i labels the three families. Furthermore there is a gauge multiplet in the $\mathbf{24}$, and for the usual reasons we need two Higgses that couple to the fermions, one in the $\mathbf{5}$ (H_d) and one in the $\mathbf{5}^*$ (H_u). Finally we need a Higgs Φ in the $\mathbf{24}$ to break $SU(5)$ to $SU(3) \times SU(2) \times U(1)$. Note that Φ is a chiral superfield, and therefore complex. Hence it contains *two* real scalars $\varphi(\mathbf{24})$. One cannot put a real scalar in a supermultiplet, since in four dimensions every fermion has at least two degrees of freedom, and a real scalar has just one.

The most general superpotential one can write down consists of the following terms. First of all there are Yukawa couplings:

$$g_{ij}\mathcal{X}_i\mathcal{X}_j H_d + g'_{ij}\mathcal{X}_i\mathcal{Y}_j H_u \ , \tag{9.57}$$

which shows that H_i play the same rôle as before. Then there are interactions among Higgs bosons

$$\lambda_1 H_d \Phi H_u + \lambda_2 \Phi^3 \ , \tag{9.58}$$

as well as mass terms for the Higgs bosons

$$M\Phi^2 + \mu H_d H_u \ , \tag{9.59}$$

and finally there are undesirable terms that lead to direct B-L violation

$$\mathcal{Y}_i H_d, \mathcal{X}_i\mathcal{Y}_i\mathcal{Y}_k, \mathcal{Y}_1\Phi H_d \tag{9.60}$$

These are omitted exactly as before. Even though $SU(5)$ unification also leads to proton decay, the extra terms due to supersymmetry would give a proton decay rate that is certainly much too large.

The unification of coupling constants works differently because there is additional matter: squarks, sleptons, gauginos and higgsinos and an additional Higgs. In the discussion of $SU(5)$ unification we have seen that matter in complete $SU(5)$ multiplets does not affect the fact that coupling constants unify, nor the mass scale at which they unify. Only the unified coupling constant is affected, and becomes larger. The squarks and sleptons are in complete $SU(5)$ multiplets, but the gaugino and the higgs are not. For the gaugino this is a direct consequence of $SU(5)$ breaking. It should be noted that $SU(5)$ breaks at a scale that it supposed to be above the supersymmetry breaking scale. Hence we may assume that supersymmetry breaking may be ignored at the $SU(5)$ breaking scale. Then the components in the **24** that are supersymmetric partners of the X and Y vector bosons get a mass of order the unification scale, whereas the other gauginos (the partners of the $SU(3) \times SU(2) \times U(1)$ gauge bosons) remain massless. The fact that the fields H_i and $\bar{H}_i$ contribute as incomplete $SU(5)$ multiplets – i.e. that their triplet components are getting a mass of order the unification scale – is on the other hand not natural. Nevertheless, we will have to assume that this happens, because if these triplets are light they would generate proton decay at much too large rates.

9.21.1 MSSM β-functions

The leading terms of the β-functions of the MSSM are obtained by just adding the contributions of the squarks, gluinos, the fermionic partner of the Higgs and the additional Higgs boson. In formula Eq. (6.27) one has to take into account one complex scalar for every Weyl fermion. Together they contribute $2I_2(R_f) + \frac{1}{2} \times 2I_2(R_f) = 3I_2(R)$. Similarly, for every complex scalar there is a Weyl fermion, so that the term $2 \times \frac{1}{2}I_2(R_s)$ (note the factor of 2 for a complex scalar) becomes $3I_2(R_s)$. This is then multiplied with another factor 2 to take into account the two Higgs bosons H_u and H_d, so that the total Higgs supermultiplet contribution is six times that of the Standard Model. For the strong interaction sector these add up as follows

$$\begin{aligned} b_0(SU(3)) = -\frac{1}{96\pi^2}[\quad & 11 \times 2 \times 3 && \text{(gauge bosons + ghosts)} \\ & -3 \times 12 && \text{(12 quark supermultiplets)} \\ & -2 \times 6 && \text{(gauginos: Majorana fermions with } I_2 = 6)] \end{aligned}$$

so that we get

$$b_0(SU(3)) = -\frac{18}{96\pi^2} \tag{9.61}$$

For the weak interactions the computation is

$$\begin{aligned} b_0(SU(2)) = -\frac{1}{96\pi^2}[\quad & 11 \times 2 \times 2 && \text{(gauge bosons + ghosts)} \\ & -3 \times 12 && \text{(12 weak doublet supermultiplets)} \\ & -2 \times 4 && \text{(gauginos: Majorana fermions with } I_2 = 4) \\ & -3 \times 2 && \text{(Higgses: 2 supermultiplets)}] \end{aligned}$$

leading to a positive (not asymptotically free) result

$$b_0(SU(2)) = \frac{6}{96\pi^2} \tag{9.62}$$

Finally, for the abelian factor of the Standard Model the non-supersymmetric result was $40+1$ (fermions+Higgs), and this now becomes $\frac{3}{2}\times 40 + 3\times 2\times 1$=66.

$$b_0(U(1)) = \frac{66}{96\pi^2} \tag{9.63}$$

9.21.2 MSSM versus SM Unification

In the eighties of last century, when supersymmetric evolution was first studied seriously, the situation did not look very favorable for supersymmetry. One way to check the convergence of the couplings is to use the QED and QCD coupling constants as input, and compute $\sin\theta_{\rm w}$, using the $SU(5)$ unification condition $g_2 = g_3 = \sqrt{\frac{5}{3}g_1}$. Then the intersection of the curves for g_2 and g_3 determines the unification scale. This yielded $M_U \approx 10^{14}$ GeV, $\alpha_5 \approx \frac{1}{40}$ and $\sin\theta_{\rm w} \approx .215$. Here M_U is the unification scale and α_5 is the unified coupling constant (to be precise, $\alpha_5 = \frac{g_5^2}{4\pi}$). The only known quantity, $\sin\theta_{\rm w}$, was in remarkable agreement with the best measurements at that time. It was found that in the minimal Susy-GUT these values were respectively $\approx 10^{16}$ GeV, $\approx \frac{1}{25}$ and $.23\pm .01$.

Since then all coupling constants have been measured with much greater precision, and it has become clear that they simply do not go through a single point anymore, as they seemed to do earlier. In particular the value of $\sin\theta_{\rm w}$ has increased to .232. Furthermore the proton was found not to decay at the rate expected by minimal non-supersymmetric $SU(5)$, so either one would have to resort to something more complicated, or reject the idea of unification.

With the every accurate LEP-data from the early nineties the picture looks much better for Susy-GUTS. One finds that within the error the coupling constants do merge, provided one takes into account in the evolution the extra particles predicted by supersymmetry, from a scale M_S to M_U. Here M_S is a common mass scale for all the superpartners, which was determined from the data. Amazingly this scale was found to be about 1 TeV (with a large error, though). Of course this scale is an extra parameter in the fit, so that it is not really all that surprising that the three coupling constants could be made to merge: success is essentially guaranteed. Furthermore one can argue that there is an infinite number of solutions to this problem, if one allows arbitrary $SU(3)\times SU(2)\times U(1)$ representations to populate the desert. Nevertheless, it must be said that the fact that M_S comes out with a "reasonable" value is remarkable. Of course this is no proof of supersymmetry, but a hint to be taken quite seriously.

The fact that the unification scale comes out about two orders of magnitude higher is in principle also good, since one can escape the bounds from the proton decay experiments. However, in Susy-GUTs there are additional diagrams leading to proton decay.

9.21.3 Proton Decay

In non-supersymmetric $SU(5)$ all diagrams leading to proton decay must involve the bosons X and Y or the triplet component of the $\mathbf{5}^*$ Higgs. In any case, a boson, whose propagator contributes $\frac{1}{M_U^2}$ to the amplitude. In terms of an effective Lagrangian, any term contributing to proton decay must thus have M_U^2 in front of it, *i.e.* it must be an operator of dimension six (indeed, the relevant operators are four-fermi interactions). In supersymmetry one can also exchange fermions with mass M_U to get B-violation, namely the partners of the aforementioned bosons. The amplitude for these processes is only suppressed by one power of M_U, and the corresponding operators are of dimension 5. Examples can easily be constructed by taking a diagram of non-supersymmetric $SU(5)$ and replacing two external fermions, as well as the interchanged vector boson, by their superpartners. Such a dimension 5 operator is built out of two fermions and two scalars. In such a process, two ordinary quarks and/or leptons are transformed into two squarks and or sleptons. These are much too heavy to form a valid decay product for the proton. Hence a second step is required to get rid of the supersymmetric particles, this time involving the exchange of a gaugino or a higgsino. Then the complete decay process is of higher order in the coupling constant and suppressed by powers of masses of Susy-partners.

One can analyze systematically which dimension 5 operators are possible. If we require $B-L$ or R-Parity to forbid the undesirable dimension 4 operators discussed earlier, only two combinations of superfields are possible, namely the F-terms $\mathcal{QQQL}$ and $\bar{\mathcal{U}}\bar{\mathcal{U}}\bar{\mathcal{D}}\bar{\mathcal{E}}$, where we use $SU(3)\times SU(2)\times U(1)$ superfield notation. In the second expression the color anti-symmetry enforces a flavor anti-symmetry for the two $\bar{u}$ fields. Hence if one of them is a u-quark, the other is necessarily charm or top, into which the proton cannot decay. Hence this operator can be ignored. The first operator involves the doublet fields $\mathcal{Q}$. One cannot take these all within the first family, since the combinations uud or ddu cannot be made anti-symmetric in the color labels (here u and d denote the upper and lower components of the superfield $\mathcal{Q}$, *i.e.* they are superfields each containing a quark and a squark). However, the combination uds is allowed. The conclusion is in general that proton decay through dimension 5 operators must involve particles from at least two families. The most important decay mode would then be $p\to K^+\bar{\nu}_\mu$ (which conserves B-L), instead of the processes $p\to\pi^+\bar{\nu}_e$ or $p\to\pi^0 e^+$ expected to be important in the non-supersymmetric case.

9.22 Conclusions

In comparison to GUTs and technicolor the case for supersymmetry is *a priori* very weak. The only problem it promises to solve is the stabilization of large mass hierarchies. Unlike technicolor, it does not explain why there is such a hierarchy of scales. It requires a lot of courage to conclude so much on the basis of so little information. It requires even more courage to state that the minimal form of this idea should be the correct model to test, although doing anything else is essentially impossible.

Nevertheless, whether one likes it or not, nothing has been found so far that rules out the MSSM, unlike minimal $SU(5)$ or technicolor. The observed coupling constant convergence – skeptical as one may and should be about it – may even be viewed as a first positive hint.

Upon closer examination, the model has some interesting features that were not put in, but come out anyway: the fact that all unobserved superpartners have $SU(3) \times SU(2) \times U(1)$-allowed masses, and the fact that the mass of one of the Higgs bosons runs to zero faster than all other scalar masses for example.

Apart from the "technical" hierarchy problem, supersymmetry at such solves none of the remaining Standard Model problems: family structure, family replication, quark and lepton mass hierarchies etc. all have to be put in by hand. One ends up with a theory with considerably more parameters than the Standard Model, although the situation improves if one combines supersymmetry with the idea of Grand Unification.

Is the MSSM falsifiable? Unfortunately the superpartner masses can be pushed to large values without any real harm. The scale that determines these masses also determines M_{W}, but the Higgs potential has enough freedom to get the correct value of M_{W} even with very large superpartner masses. Although this is "unnatural", unfortunately it is impossible to obtain an upper bound from such a principle. Hence it will not be possible to rule out supersymmetry by *not* finding, for example, sleptons, although it may be possible to diminish the number of believers.

It is clear that the idea of low-energy supersymmetry will still be with us for many years. If it turns out to be realized in nature this would be an incredible theoretical achievement, given the tiny amount of experimental evidence on which the case is presently based.

9.23 References

Most of the results presented here were based on the Physics Reports by H. Nilles [23] and lecture notes by H. Haber [16]. The superfield formalism is explained in the book by Bagger and Wess [2]. Other sources are [37], [38], [24], [9], [18] and [17], and references cited in these papers.

10 Supergravity

As we have seen before, spontaneously broken local supersymmetry does not easily yield an acceptable theory. The supertrace formula for the squared masses does not allow us to get reasonable multiplet splittings, except perhaps if one uses Fayet-Illiopoulos breaking, which however is unappealing for other reasons. In addition one gets in the spectrum a massless fermion, the Goldstino, which has not been seen. Furthermore it is clear that ultimately we would like to couple supersymmetry to gravity. One would expect that in order to keep exact supersymmetry the graviton has to belong to a supermultiplet itself. Indeed, supersymmetry requires the existence of N superpartners of spin $\frac{3}{2}$ called *gravitinos* if there are N supersymmetries. Particles of spin larger than $\frac{1}{2}$ can only exist in an interacting field theory if they are gauge particles of some symmetry. For spin 1 this symmetry is local invariance with respect to some gauge group, for spin 2 it is general coordinate invariance, and for spin-$\frac{3}{2}$ the gauge symmetry turns out to be supergravity, or, what is the same, local supersymmetry. Thus the combination of supersymmetry and gravity inevitably leads us to supergravity.

There is an alternative. We could simply view supersymmetry as a coincidence without deeper meaning. The world would then be described by a supersymmetric theory with soft explicit supersymmetry breaking terms. At low energies this theory would not look supersymmetric at all, like the world we observe, and at higher energies it would look more supersymmetric, but never exactly supersymmetric. Such a theory would still have all the miraculous cancellations of quadratic (and some logarithmic) divergences we expect from supersymmetry. Once we couple it to gravity those properties might be lost, but coupling a theory to gravity leads to problems anyway. However, if this would turn out to be the solution nature has chosen it would be extremely disappointing.

Most people believe that if supersymmetry has something to do with the Standard Model Higgs mechanism, it must be a local symmetry, *i.e.* supergravity. The phenomenology of supergravity is still in its infancy. For global supersymmetry there is at least a "minimal standard model", although perhaps the restrictions imposed on the parameter space may not convince everyone. Some of the problems of the MSSM are hoped or expected to be solved when supergravity is added, but at the moment there only exists a rather large collection of interesting ideas, each with obvious shortcomings.

10.1 Local Supersymmetry

The maximum number of supersymmetries one can have in a locally supersymmetric theory is $N = 8$. With more supersymmetries one inevitably gets a massless spin-$\frac{5}{2}$ particle, and it is not possible to write down a consistent field theory involving such particle. Although the convergence properties of the theory improve when there are more supersymmetries, even $N = 8$ supergravity is not expected to be finite or renormalizable, although that has not been proved yet: the expected infinities occur for the first time in seven loops diagrams! [without coupling to gravity the maximum number of supersymmetries is $N = 4$; this theory, $N = 4$ super Yang-Mills theory *is* finite.] It follows that from now on we will not be dealing with renormalizable theories anymore. Our best hope at present of solving these problems is string theory. Supersymmetric string theory is believed to yield a finite theory with $N = 1, \ldots, 8$ supergravity automatically contained in it. However, it is better to discuss first supergravity by itself, and worry about superstrings later. Since $N = 1$ supersymmetric theories look most promising, in particular since they allow chiral multiplets, we will restrict ourselves to $N = 1$ in the following.

Supersymmetry can be made local by assuming that the infinitesimal parameter in a supersymmetry transformation depends on the space-time point x^μ. Then the commutator of two such transformations yields

$$[\alpha(x)Q, \bar{Q}\bar{\alpha}(x)] = 2\alpha(x)\sigma_\mu\bar{\alpha}(x)P^\mu$$

The right-hand side is a space-time dependent translation $\epsilon(x)_\mu P^\mu$, in other words a general coordinate transformation. Once we have general coordinate transformations it is inevitable that we have to couple the theory to gravity.

The minimal particle content of such a theory is a graviton and its supersymmetric partner the gravitino. Since the graviton has spin 2, it is not a surprise that the gravitino has spin $\frac{3}{2}$. Since the graviton has two physical components (helicity ± 2), the gravitino must have two as well. This is consistent with its mass being zero; a massive spin-$\frac{3}{2}$ particle must necessarily have all four spin states $-\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}$ since we go to its rest frame and apply an $SO(3)$ rotation to it.

Since we do not observe massless gravitinos the gravitino must acquire a mass. It can do so by the super-analog of the Higgs mechanism: it absorbs two degrees of freedom by eating the Goldstino appearing when supersymmetry breaks. In this way we remove the Goldstino from the spectrum.

10.2 The Lagrangian

The Lagrangian for $N = 1$ supergravity without any other fields than the graviton and the gravitino is

$$\mathcal{L} = -\tfrac{1}{2\kappa^2}\sqrt{g}R - \tfrac{1}{2}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\mu\gamma_5\gamma_\nu\partial_\rho\psi_\sigma \ , \tag{10.1}$$

where ψ_μ is the gravitino field. Just as the graviton it must satisfy gauge constraints reducing its number of degrees of freedom to 2. The supersymmetry transformations are

$$\begin{aligned} \delta e_\mu^m &= \tfrac{1}{2}\kappa\bar{\alpha}\gamma^m\psi_\mu \\ \delta\psi_\mu &= \tfrac{1}{\kappa}(\partial_\mu + \tfrac{1}{2}\omega_\mu^{mn}\sigma_{mn})\alpha \equiv \tfrac{1}{\kappa}D_\mu\alpha \end{aligned}$$

Here e_μ^m is a vierbein (also called "tetrad"), which has one space-time index μ and a local Lorentz tangent space index m, ω_μ^{mn} the spin connection and κ is the gravitational coupling constant, related to the Planck mass by

$$\kappa = \sqrt{\frac{8\pi}{M_{\text{Planck}}}} \ . \tag{10.2}$$

The vierbein is related to the metric by

$$\gamma_{\mu\nu} = e_\mu^m e_\nu^n \eta_{mn} \ , \tag{10.3}$$

with $\eta_{mn} = \text{diag}\,(-1, 1, 1, 1)$. It can be used to replace space-time indices by local tangent space indices. For example, $\gamma^m = e_\mu^m\gamma^\mu$. Furthermore $\sigma^{mn} \equiv \frac{1}{4}[\gamma^m, \gamma^n]$ is an $SO(3,1)$ generator. The covariant derivative is thus very similar to a gauge covariant derivative, with ω_μ^{mn} interpreted as an $SO(3,1)$ gauge potential. However, the analogy is not perfect. One difference is that, unlike a gauge field A_μ, the spin connection is not an independent physical degree of freedom (indeed, that would violate supersymmetry), but can be eliminated in terms of the vierbein. Finally, g is the absolute value of the determinant of the metric.

Now we must couple the graviton and the gravitino to chiral multiplets and vector multiplets. Since there is no need anymore for the theory to be renormalizable, we begin by dropping that requirement. The most general supersymmetric action with at most two derivatives is

$$\mathcal{L} = \int d^4\theta \phi(\bar{S}e^{2gV}, S) + \text{ Re } \left[\int d^2\theta f_{ab}(S) W^a_\alpha W^b_\beta \epsilon^{\alpha\beta} + \int d^2\theta g(S) \right] \tag{10.4}$$

Here S denotes the full set of chiral superfields, V the set of vector superfields, $W_\alpha = (\bar{D}\bar{D})e^{-gV}D_\alpha e^{gV}$. The action depends on three functions of the chiral superfields which are usually called f, ϕ and g [This notation is somewhat unfortunate in view of previous definitions, but we will respect the traditional notation here. One should not confuse the function ϕ with a chiral superfield or a scalar field, nor confuse the function g with a gauge coupling.] These functions have the following properties

- The function $f(z)$ is holomorphic in z (*i.e.* $\partial_{\bar{z}} f(z) = 0$), and transforms under gauge transformations as the symmetric product of two adjoint representations. In a renormalizable theory the only possibility is $f_{ab} \propto \delta_{ab}$. The proportionality constant may in fact be complex. Then the real part multiplies the gauge kinetic terms $F^2_{\mu\nu}$, and the imaginary part appears in front of the topological term $F_{\mu\nu}\tilde{F}_{\mu\nu}$.

- The function $\phi(z, \bar{z})$ must be real. In a renormalizable theory it must be proportional to $\bar{z}z$.

- The function $g(z)$ is a holomorphic function of z, and is nothing but the superpotential. In a renormalizable theory is must be a polynomial in z of degree three (or less).

There is a deceptively simple expression way of coupling this globally supersymmetric action to supergravity. Instead of Eq. (10.5) one writes

$$\mathcal{L} = \int d^4\theta E \left(\phi(\bar{S}e^{2gV}, S) + \text{ Re } R^{-1} \left[f_{ab}(S) W^a_\alpha W^b_\beta \epsilon^{\alpha\beta} + g(S) \right] \right) \tag{10.5}$$

Here E is the "superspace determinant" and R the "chiral curvature scalar". All these fields are fields in "curved superspace". We will not explain this further here.

The Lagrangian Eq. (10.5) can be written out in components. It turns out that the result depends only on two independent functions instead of the three one has on the global case: the functions ϕ and g only appear in the combination

$$G(z, \bar{z}) = 3 \log(-\phi) - \log(|g|^2) \ . \tag{10.6}$$

This function is called the Kähler potential. It turns out that the scalars in supergravity can be viewed as coordinates of a complex manifold with some special properties, which is called a Kähler manifold. The metric on such a manifold can be expressed in terms of the Kähler potential

$$G^i_j \equiv \frac{\partial^2 G}{\partial z_i \partial \bar{z}^j} \tag{10.7}$$

Here and in the following derivatives with respect to z_i are denoted by an upper index i, and with respect to $\bar{z}^i$ by a lower index i. The indices i label all different scalars that may appear, as well as their indices in any of the gauge group representations.

The Lagrangian can be split into four kinds of terms: bosonic kinetic terms, fermionic kinetic terms, the scalar potential, and all remaining terms without derivatives. The bosonic kinetic terms are

$$\mathcal{L}_{B_{\rm kin}} = e\left[-\tfrac{1}{2}R + G^i_j D_\mu z_i D^\mu \bar{z}^j - \tfrac{1}{4}(\text{ Re } f_{ab})F^a_{\mu\nu}F^{\mu\nu,b} - \tfrac{1}{4}(\text{ Im } f_{ab})F^a_{\mu\nu}\tilde{F}^{\mu\nu,b}\right] \tag{10.8}$$

Here and in the following D_μ denotes a derivative that is covariant with respect to the gauge group as well as gravity, and e is the determinant of the vierbein (*i.e.* $e = \sqrt{|g|}$, where g is the space-time metric). This part of the Lagrangian is exactly as one could have expected.

The scalar potential has the form

$$\mathcal{L}_{B_{\rm pot}} \equiv -V = e^{-G}\left[3 + G_k(G^{-1})^k_l G^l\right] - \tfrac{1}{2}g^2 \text{ Re } \left[f^-_{ab}(G^i T^{aj}_i z_j)(G^k T^{bl}_k z_l)\right] . \tag{10.9}$$

Here g is the gauge coupling (which cannot be confused anymore with the superpotential, since the latter has been absorbed in G). If the gauge group is semi-simple the second term becomes a sum over all factors, each of which may have a different coupling constant, and a different function f_{ab}. The coupling constant is only normalized in the standard way if in the kinetic terms in the vacuum one considers are properly normalized: $f_{ab}(\langle z\rangle) = \delta_{ab}$. By $(G^{-1})^k_l$ we mean the inverse Kähler metric, and *not* the double derivative of the function G^{-1}. This metric must be regular and have an inverse for the theory to make sense. Finally, T^{aj}_i is a generator of the gauge group in the (in general reducible) representation of the scalars.

The terms involving fermions are much more complicated, and we will not present them here. These results are valid if $g \neq 0$ and in the absence of a Fayet-Illiopoulos term.

10.3 Spontaneous Symmetry Breaking

Just as for global supersymmetry, the condition for spontaneously broken local supersymmetry is that the auxiliary fields have a vacuum expectation value. Equivalently one may require that $\langle\{Q,\Psi\}\rangle \neq 0$. The auxiliary fields are precisely equal to the terms in this anti-commutator that do not contain space-time derivatives. Terms with derivatives cannot get a v.e.v. without breaking Lorentz invariance.

Explicitly these terms are

$$F_i = e^{-G/2}(G^{-1})^j_i G_j + \tfrac{1}{4}f_{ab,k}(G^{-1})^k_i \lambda^a\lambda^b - (G^{-1})^k_i G^{jl}_k \psi_j\psi_l - \tfrac{1}{2}\psi_i G_j \psi^j \tag{10.10}$$

and

$$D_a = i \text{ Re } f^{-1}_{ab}(-gG^i T^{bj}_i z_j + \tfrac{1}{2}if^i_{bc}\psi_i\lambda^c - \tfrac{1}{2}if^{bc}_i\psi^i\lambda_c) - \tfrac{1}{2}\lambda_a(G^i\psi_i) \tag{10.11}$$

Note that in addition to purely bosonic terms there are also fermionic ones. These terms disappear in the limit $M_{\rm Planck} \to \infty$ (the dependence on $M_{\rm Planck}$ is not been explicitly given here, but can be inferred from the dimensions. In this limit all non-renormalizable terms must vanish as well).

Supersymmetry can now be broken by a vacuum expectation value of a fermion bi-linear, or by a purely bosonic v.e.v. We first consider the purely bosonic terms. The fermionic terms in the action, which we did not write down, contain among many others the terms*

$$e^{-G/2}[\bar{\psi}_{\mu,L}\sigma^{\mu\nu}\psi_{\nu,R} - \bar{\psi}_{\mu,L}\gamma^{\mu}\eta_L - \tfrac{2}{3}\bar{\eta}_R\eta_L] \ , \tag{10.12}$$

where

$$\eta = e^{G/2}[e^{-G/2}G^i\psi_i + \frac{1}{\sqrt{2}}D^a\lambda^a] \tag{10.13}$$

If one of the coefficients on the right-hand side has a non-zero v.e.v, there are bi-linears $\psi_\mu\psi_i$ or $\psi_\mu\lambda^a$ in the action, and hence we see that the fields ψ_i and/or λ^a mix with the gravitino. It can be shown that then η can be removed from the action by a shift of the gravitino field

$$\psi_\mu \rightarrow \psi_\mu - \tfrac{1}{3}e^{-\langle G\rangle/2}\gamma_\mu\eta - \tfrac{2}{3}\partial_\mu\eta \ . \tag{10.14}$$

In the global limit η reduces to the Goldstino field, up to normalization.

Furthermore, if $\langle e^{-G/2}\rangle \neq 0$ we see from (10.12) that the gravitino gets a mass:

$$m_{3/2} = \kappa^{-1}e^{-G/2} \ , \tag{10.15}$$

where on the right-hand side a factor $\kappa^{-1} = M_{\text{Planck}}/\sqrt{8\pi}$ was inserted for dimensional reasons. Note that this mass vanishes if $\langle g\rangle = 0$. In that case G is not well-defined, but $e^{-G/2} \propto |g|/\phi^{\frac{1}{3}} = 0$.

It is in fact not quite correct to interpret the gravitino mass in this way, nor is it correct to conclude that $\langle e^{-G/2}\rangle \neq 0$ implies that supersymmetry is broken. Indeed, supersymmetry is broken if *and only if* the F or D term has a vacuum expectation value.

To see this more clearly, consider the scalar potential. Usually an additional assumption is made, namely that the Kähler metric and the function f_{ab} are proportional to the unit matrix:

$$G^i_J = -\delta_{ij}; \quad f_{ab} = \delta_{ab} \tag{10.16}$$

This is called "minimal coupling" of Yang-Mills and matter to supergravity. Under this assumption, the scalar potential can be written as

$$V = -3e^{-G} + |F_i|^2 + \tfrac{1}{2}D_a^2 \tag{10.17}$$

where in F_i and D_a we only take the bosonic terms into account. Just as in the global case we define the supersymmetry breaking scale in terms of the vacuum expectation value of the auxiliary fields:

$$M^2_{\text{Susy}} = \sqrt{|F_i|^2 + \tfrac{1}{2}D_a^2} \ . \tag{10.18}$$

* In some (most?) of the literature on supergravity $\bar{\psi}_L$ is apparently defined as $(\bar{\psi})_L = \psi^\dagger\gamma_4 P_L$. In that notation one has bi-linears $\psi_L\psi_L$, $\psi_L\gamma_\mu\psi_R$, etc. Here we adopt the standard convention of the non-supergravity literature, namely $\bar{\psi}_L = \overline{(\psi_L)} = \psi^\dagger P_L\gamma_4$. The bi-linears have then the form $\psi_L\psi_R$, $\psi_L\gamma^\mu\psi_R$, etc

This is the square root of the shift of the potential due to supersymmetry breaking.

The scalar potential in supergravity shows a new feature in comparison with the one of supersymmetry, namely the extra term $-3e^{-G}$. Its presence implies that the potential is not positive definite. Furthermore, even in a supersymmetric point ($\langle F_i \rangle = \langle D_a \rangle = 0$) the potential does not necessarily vanish.

In non-supersymmetric theories, the value of the potential can be changed by a constant. Hence the cosmological constant can be tuned to zero, although we have no insight in the reason for this fine-tuning. In supersymmetric theories one cannot add an arbitrary constant to the potential. In a globally supersymmetric theory the cosmological constant is zero before supersymmetry breaking, but is definitely non-zero and positive after supersymmetry breaking. In the absence of gravity we could however ignore this problem. Fortunately, local supersymmetric theories are better in this respect. One can again tune the cosmological constant to zero, not by adding an arbitrary constant, but by requiring a suitable value for the constant $-3e^{-\langle G \rangle}$. Also in this case there is no fundamental insight in the mechanism that might impose such a fine-tuning, nor is the value of Λ_c protected against corrections due to further shifts in the potential, for example in weak interaction symmetry breaking.

If we do not do such a fine-tuning, we end up in de-Sitter and anti-deSitter space, and we cannot even interpret the masses we get in the conventional way. For example, if $e^{-G} \neq 0$ but all auxiliary fields have zero v.e.v's supersymmetry is unbroken, but the gravitino has a mass. Since we are in anti-de Sitter space our usual notions about masslessness are no longer valid, and these two facts are not in contradiction. Clearly it would not make much sense to compute a mass spectrum from supergravity if the v.e.v. of the potential is not tuned to zero.

This then gives us immediately an expression relating the gravitino mass to $M_{\rm Susy}$ and κ. Requiring $V = 0$ in Eq. (10.17), and using Eqs. (10.18) and (10.15) we get

$$m_{3/2} = \frac{M_{\rm Susy}^2 \kappa}{\sqrt{3}} = \sqrt{\frac{8\pi}{3}} \frac{M_{\rm Susy}^2}{M_{\rm Planck}} \; . \tag{10.19}$$

One may also compute the tree level mass matrices for the remaining fermions, the scalars and the spin-1 fields. Minimal coupling implies that G has the form

$$G(z, \bar{z}) = -z\bar{z} - \log |g(z)^2| \; . \tag{10.20}$$

With these choices the scalar and Yang-Mills have their canonical form. Under this assumption one can derive the mass sum rule Eq. (9.31). This rule is valid at tree level for minimal coupling and if $\langle D \rangle = 0$. If one also includes D-type breaking, it is generalized to

$$\sum_S (-1)^{2S} (2S+1) M_S^2 = (N-1)[2m_{3/2}^2 - \kappa^2 \langle D_a^2 \rangle] - 2g_a \langle D_a \rangle \ {\rm Tr} T^a \tag{10.21}$$

This mass sum rule does not give all the information that is available. In fact one can express all tree-level mass matrices completely in terms of F_i and D_a. It is somewhat disturbing that in these expressions a non-zero gaugino mass requires the corresponding

D_a term to have a non-zero vacuum expectation value. However, D-type breaking remains undesirable. One possible way out of this is to use non-minimal couplings.

If we omit the D-terms, we see that the chiral multiplet splittings induced by supersymmetry breaking are of order $m_{3/2}$. In particular a contribution of this order of magnitude may be expected for the masses of the Higgs scalars, which in their turn determine the weak scale. The relation between these two scales is in fact a rather complicated function of all MSSM parameters, and there exist regions in parameter space where $m_{3/2}$ is much larger than $M_{\rm Z}$. However, this is considered "unnatural", and current prejudice says that we should have $m_{3/2}$ not too much higher than $M_{\rm Z}$. Substituting $m_{3/2} \approx 100$ GeV in Eq. (10.19) gives $M_{\rm Susy} \approx 10^{10}$ GeV.

Up to now we have only considered non-vanishing v.e.v's for bosonic fields. It is also imaginable that fermion bi-linears get a v.e.v that breaks supersymmetry. Consider for example Eq. (10.11). The second term involves a gaugino bi-linear. If $f_{ab,k} \neq 0$ (which means that the couplings are not minimal) this may yield a contribution to $\langle F_i \rangle$, which is proportional to $\langle \lambda\lambda \rangle$. This is called gaugino condensation. Since $\langle \lambda\lambda \rangle$ has dimension three, we conclude that, after tuning the resulting effective potential to zero we will get supersymmetry breaking with an associated scale

$$M_{\rm Susy}^2 \sim \frac{\langle \lambda\lambda \rangle}{M_{\rm Planck}} \tag{10.22}$$

Here we define $M_{\rm Susy}^4$ as the shift in the potential due to the symmetry breakdown. Of course the foregoing calculations are not valid in this case, but clearly $M_{\rm Susy}^2$ will be proportional again to $\langle F_i \rangle$, and for dimensional reasons there must then be a factor $1/M_{\rm Planck}$. For the gravitino mass one may then expect a formula like

$$m_{3/2} \sim \frac{\langle \lambda\lambda \rangle}{M_{\rm Planck}^2} \tag{10.23}$$

This yields $\langle \lambda\lambda \rangle \approx 10^{13}$ if $m_{3/2} \approx 100$ GeV. It goes without saying that the mass-squared sum rule is not valid in this case, although one may expect a similar formula to hold.

The attractive point about gaugino condensation is that it is now possible that $M_{\rm Susy}$ is generated dynamically, just as in Technicolor models. Suppose we add to the Standard Model an extra gauge group G, whose coupling merges with the Standard Model couplings at $M_{\rm GUT}$ (or perhaps $M_{\rm Planck}$). Then the value of the coupling at that scale is fixed, and we can use renormalization group evolution to compute at which scale it becomes large. Just as the fact that the $SU(3)_{\rm color}$ coupling becomes large triggers chiral symmetry breaking via quark condensates, it seems plausible that when the coupling constant of G becomes strong it forms condensates of the fermions it couples to. Since supersymmetry is still unbroken, those fermions include the gauginos of G, and possibly nothing else. Just as the there is no hierarchy problem for $\Lambda_{\rm QCD}/M_{\rm Planck}$, there would be no hierarchy problem for $M_{\rm Susy}/M_{\rm Planck}$ either. This large ratio would be explained as in terms of an exponential e^{-1/g^2}, where g is a coupling constant which is small, but of order 1. The unattractive point this mechanism is that very little is know about whether and how exactly it works.

10.4 Hidden Sector Models

It is clear that we cannot tolerate vacuum expectation values of order M_{Susy} for quantities that carry non-trivial $SU(3) \times SU(2) \times U(1)$ quantum numbers. One usually assumes that supersymmetry breaking takes place in a sector of the theory that has trivial Standard Model representations, and couples to the visible world only via (super)gravitational interactions. This is called the "hidden sector".

A frequently used toy model to describe hidden sector symmetry breaking is the Polonyi-model. This model has just one chiral superfield in the hidden sector, whose scalar component we will call, as before, z. The couplings are assumed to be minimal. Let us first derive some useful results for minimally coupled theories with an arbitrary number of scalars. We will ignore D-terms in the following, in other words we assume that they do not get v.e.v's.

Since $G(z_i, \bar{z}^j) = -\sum_i (z_i \bar{z}^i) - \log |g(z_i)|^2$ we find

$$G^i = \partial_{z_i} G = -\bar{z}^i - \frac{g^i}{g} \tag{10.24}$$

Hence

$$F^i = e^{\frac{1}{2} z_i \bar{z}^i} |g| (\bar{z}^i + \frac{g^i}{g}) \ , \tag{10.25}$$

and

$$V = e^{z_i \bar{z}^i} \left[|\bar{z}^i g + g^i|^2 - 3|g|^2 \right] \ . \tag{10.26}$$

In all these expressions κ is set to 1. It can be restored easily using the fact that z_i has dimension 1, g has dimension 3, and κ has dimension -1.

In the Polonyi model one chooses $g = m^2(z + \beta)$. The value of F is then $e^{z\bar{z}/2} m^2 (1 + \bar{z}(z + \beta))$. If $\beta < 2$ this quantity is positive for any value of z, and hence supergravity must be broken. The potential is equal to

$$V = e^{z\bar{z}} m^2 [|1 + \bar{z}(z + \beta)|^2 - 3|z + \beta|^2] \tag{10.27}$$

It is easy to check that the conditions $V = \partial_z V = \partial_{\bar{z}} V = 0$ have a solution $\beta = 2 - \sqrt{3}$, $z = \sqrt{3} - 1$ (there is another solution with opposite signs for z and β, and two more solutions with $\beta > 2$). This solution gives broken supergravity with $V = 0$ and

$$\langle F \rangle = e^{\frac{1}{2}(\sqrt{3}-1)^2} \sqrt{3} m^2 \ . \tag{10.28}$$

The gravitino mass is thus

$$m_{3/2} = \frac{8\pi m^2}{M_{\text{Planck}}} e^{\frac{1}{2}(\sqrt{3}-1)^2} \ . \tag{10.29}$$

Note that the vacuum expectation value of z is of order 1 with respect to M_{Planck}. To get a reasonable value of $m_{3/2}$ we have to fine-tune m to about 10^{10} GeV. In addition the value of β has been fine-tuned to get $V = 0$. Furthermore there is a hidden fine-tuning in the choice of g. In principle there could also be terms of second and third order in z, but this destroys in most cases the possibility of having a minimum at $V=0$. Clearly this is no more than a toy model.

More generally, the hidden sector can be coupled to the observable sector simply by writing the complete superpotential as a sum of two terms,

$$g(z_i, y_m) = h(z_i) + k(y_m) \ , \tag{10.30}$$

where z_i are hidden sector scalars and y_m observable sector scalars. The scalar potential is

$$V = e^{\kappa^2(|z_i|^2+|y_m|^2)} \left[|h_i + \kappa^2 \bar{z}^i g|^2 + |k_m + \kappa^2 \bar{y}^m g|^2 - 3\kappa^2 |g|^2\right] \ , \tag{10.31}$$

where we have restored the dependence on κ. We will assume that the hidden sector fields get v.e.v's of order $M_{\text{Planck}} \sim \kappa^{-1}$:

$$\langle z_i \rangle = \kappa^{-1} b_i \tag{10.32}$$

The gravitino mass is given by

$$m_{3/2} = \kappa^{-1} \langle e^{-G/2} \rangle = \kappa^2 e^{\frac{1}{2}|\kappa\langle z_i\rangle|^2} \langle h \rangle \ , \tag{10.33}$$

so that

$$\langle h \rangle = \kappa^{-2} m_{3/2} e^{-\frac{1}{2}|b_i|^2} \tag{10.34}$$

Finally we need to parametrize the expectation value of the derivative of h. In the Polonyi model we had $\langle h' \rangle = m^2 \approx \kappa^{-1} m_{3/2}$. Inspired by this result we postulate

$$\langle h_i \rangle = \bar{a}_i \kappa^{-1} m_{3/2} \ . \tag{10.35}$$

The observable sector variables, y, k and k_i have characteristic scales much below the Planck scale. Their vacuum expectation values vanish, or are completely negligible in comparison to M_{Planck}. To get the "effective" potential for the observable sector we take the limit $\kappa \to 0$, after substituting the v.e.v's of the hidden sector fields. There are some poles in κ, but only in terms that do not depend on observable sector fields. Those poles have to be canceled, which can be done by requiring that $|a_i + b_i|^2 = 3$. This is simply the requirement that the cosmological constant should vanish, and as usual this has to be arranged by hand. When the condition $|a_i + b_i|^2 = 3$ is satisfied, one finds that all constant terms cancel. The remaining terms are

$$V_{\text{obs}} = |\hat{k}_m|^2 + m_{3/2}^2 |y_m|^2 + m_{3/2} \left[y_m \hat{k}_m + (A-3)\hat{k} + \text{c.c.} \right] \tag{10.36}$$

Here $\hat{k}$ is the rescaled superpotential

$$\hat{k} = e^{\frac{1}{2}|b_i|^2} k \ , \tag{10.37}$$

and

$$A = \sum_i \bar{b}_i (a_i + b_i) \ . \tag{10.38}$$

In this calculation the scale m was put in by hand, and the cosmological constant was fine-tuned to zero by hand. Furthermore we did not worry about finding the minimum of the potential, as we did in the foregoing example. Of course it is *assumed* that $\langle z_i \rangle$ minimizes the potential

This result looks like the potential of a supersymmetric theory with soft breaking terms. The first term is the usual scalar potential. The second is a scalar mass term, one of the allowed soft supersymmetry breaking terms. The last terms are other soft correction terms.

Note that all scalars receive the same mass $m_{3/2}^2$. Equality of the masses is one of the assumptions of the MSSM. The reason it happens here is due to the choice of minimal kinetic terms in the observable sector. It is therefore not an inevitable consequence of supergravity.

The superpotential k will in general be a sum $\sum_n k^n$, where k^n contains all terms with n fields. Then

$$\sum_m y_m \partial_m k^n = n k^n \ , \tag{10.39}$$

and the complete set of correction terms has the form

$$m_{3/2} \sum_n (A - 3 + n) k^n \tag{10.40}$$

The soft supersymmetry breaking terms in the *potential* are thus given by the terms in *super-potential* with a factor $m_{3/2}(A-3+n)$ multiplying the n^{th} order terms. Cubic terms get a factor $m_{3/2}A$, quadratic one a factor $m_{3/2}(A-1)$ The value of A depends on the details of the hidden sector.

The cubic terms contains combinations of two squarks and a Higgs with the structure of a Yukawa coupling. Such terms violate continuous R-symmetry, but not R-parity. This is exactly what we assumed earlier. From the point of view of the observable sector R-symmetry looks as if it is explicitly broken. The same is true for supersymmetry. Both symmetries are however broken *spontaneously* in the hidden sector. One may thus expect a Goldstone boson of broken R-symmetry. Since this boson is built out of hidden sector field, its couplings to the observable sector are very weak, of gravitational strength. Furthermore it may happen that the R-symmetry was not exact, but has an anomaly with respect to some gauge group, in either the hidden or the observable sector. In that case the Goldstone boson is actually a pseudo-Goldstone boson, and it becomes massive. If it gets its mass from observable sector instanton effects it will be extremely light, like an invisible axion. If it gets is mass from a hidden sector gauge group it could well be very heavy (like a scaled-up η'), but since it is also very weakly coupled to us it is completely irrelevant. Note that if we add a gauge-group in the hidden sector that only couples to a gaugino, then R-symmetry automatically has an uncancelable anomaly with respect to this gauge group. Whether these alternatives can actually be realized is very model-dependent, and we merely mention them here as logical possibilities.

Now that R-symmetry is broken there is no obstruction to gaugino masses. In the particular kind of F-type breaking considered here they are not generated at tree level.

However, the gauginos couple to quark-squark loops, which generate a gaugino mass when supersymmetry is broken. The size of the mass is dominated by the top quark contribution,

$$m_a = \alpha_a \frac{m_t^2}{m_{3/2}} C, \tag{10.41}$$

where C is a numerical factor. This result holds in the limiting case $m_{3/2} \gg m_t$. If the gravitino mass is much smaller than m_t the result is proportional to $m_{3/2}$, since of course it must vanish when $m_{3/2} = 0$. In any case this value is much too small in comparison with present experimental limits on the gluino mass. As already mentioned, non-minimal couplings $f_{ab}(z)$ provide a possible way out. Gaugino masses are presumably also be generated in supergravity breaking through gaugino condensation, since the Lagrangian contains quartic gaugino interactions. However, as we have already seen, gaugino condensation also requires non-minimal couplings f_{ab}.

The universality of the gaugino masses assumed in the MSSM does not look natural from this point of view. However, if there is unification of the coupling constants there would only one gaugino mass above the unification scale. The evolution of the separate gaugino masses starts then at M_{GUT}.

Since the masses are "running" as a function of the scale, it is not quite clear at which scale we should impose the unification condition for the scalars. Many authors assume this to be the gauge unification scale. This is certainly true for those scalars that come from the same multiplet of the unified gauge group, but not for different multiplets. It seems more reasonable to assume that the boundary condition that all scalar masses are equal should be imposed at the Planck scale.

10.5 Conclusions

Once one has accepted supersymmetry as a symmetry of nature, supergravity is nearly inevitable. It is required in order to couple a supersymmetric theory to gravity, and also to avoid disastrously large contributions to the cosmological constant, inevitable in spontaneously broken global supersymmetry. Even in supergravity models the cosmological constant problem still requires a solution, but at least the existence of a solution is not *a priori* ruled out.

Supergravity also to provides the only sensible way of spontaneously breaking supersymmetry, the super-Higgs mechanism. This eliminates first of all the undesirable massless Goldstino, but also produces an indispensable contribution to the mass sum rule for broken supermultiplets.

Finally, supergravity models offer partial justification for the "unification assumptions" generally made in the MSSM, although the case is far from being convincing.

10.6 References

The references used here include some of the papers listed at the end of the supersymmetry section, plus reviews by P. van Nieuwenhuizen [31] and S. Ferrara [11].

A Spinors

In this appendix we review some properties of spinors. The action for fermions is derived using – as much as possible – only group properties of the Lorentz group.

A.1 Spinors in $SU(2)$

The three-dimensional rotation group is $SO(3)$. Its generators T^i can be chosen so that they satisfy the algebra

$$[T^i, T^j] = i\epsilon^{ijk}T^k \ , \tag{A.1}$$

with $\epsilon^{123} = 1$. The irreducible representations of $SO(3)$ have dimensions $2j + 1$, $j \in \mathbf{Z}$, and they are real. The generators of the 3-dimensional representation are $(T^i)_{jk} = -i\epsilon_{ijk}$ (upper and lower indices have no special significance here).

The elements of the group $SO(3)$ are the 3×3 orthogonal matrices with determinant 1. One can also define the group $O(3)$, by dropping the requirement that the determinant should be 1. The group $O(3)$ consist of all the elements of $SO(3)$ plus all elements times $-\mathbf{1}$, corresponding to space inversion. The group $SO(3)$ is connected (all elements can be connected in a continuous way to the identity), $O(3)$ is not.

The group manifold of $SO(3)$ can be characterized as follows. Any element of $SO(3)$ is given by by a unit vector $\hat{n}$ and a rotation ϕ around $\hat{n}$. One can choose $-\pi < \phi \leq \pi$, and represent all elements by three-vectors $\phi\hat{n}$. It follows that the group manifold is precisely the interior of the three-sphere with radius π plus part of the boundary. In order to avoid over-counting of boundary points, one has to identify antipodal points: rotating by π around $\hat{n}$ is the same as rotating by π around $-\hat{n}$.

A manifold is called simply connected if any closed loop can be contracted to a point in a continuous way. The $SO(3)$ group manifold does not have that property: a line connecting antipodal points through the interior of the sphere is a closed loop, but it cannot be contracted.

Any compact Lie group has a simply connected covering group. This is a group with the same Lie-algebra and a simply connected group manifold. In general all representations of the Lie-algebra can be exponentiated to representations of the covering group, but not always to representations of a non-simply connected group.

The covering group of $SO(3)$ is $SU(2)$. The group manifold is $SU(2)$ is the surface of a four-sphere, which is indeed simply connected. The group $SU(2)$ has more representations than $SO(3)$, since j is allowed to have half-integer values as well as integer ones. Half-integer spin representations (spinor representations) would yield opposite signs on the anti-podal points of the $SO(3)$ manifold, and hence are not representations of $SO(3)$ (this corresponds to the well-known fact that spinors change sign under a 2π rotation).

The simplest spinor representation has dimension 2, and the representation matrices are $\frac{1}{2}\tau^i$, where τ^i are the Pauli matrices (we drop the normalization factor $\frac{1}{2}$ in the remainder of this Appendix)

$$\tau^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad \tau^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}; \quad \tau^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{A.2}$$

The spinor representations are not real, but pseudo-real. This means in particular that a matrix C must exist so that $-(\tau^a)^* = C^\dagger \tau^a C$. A matrix C with that property is $C = i\tau^2$; thus $C^{\alpha\beta} = \epsilon^{\alpha\beta}$, with $\epsilon^{12} \equiv 1$. Note that $C = C^* = -C^T = -C^\dagger$.

Important invariant tensors of $SU(2)$ are C and the Pauli matrices. Suppose ψ_α and ϕ_α are spinors and V^i a vector. Then one can construct for example the following invariants

$$
\begin{array}{cc}
\psi C \phi & \vec{V} \cdot \psi C \vec{\tau} \phi \\
\psi^* \phi & \vec{V} \cdot \psi^* \vec{\tau} \phi
\end{array}
$$

with all suppressed indices contracted in the obvious way.

It is sometimes convenient to introduce quantities with upper and lower indices, related as follows

$$\phi^\alpha = \epsilon^{\alpha\beta} \phi_\beta \ . \tag{A.3}$$

For the lowering of indices we define

$$\phi_\alpha = \phi^\beta \epsilon_{\beta\alpha} \tag{A.4}$$

It is easy to check that the validity of both relations requires

$$
\begin{aligned}
\phi_\alpha &= \phi^\beta \epsilon_{\beta\alpha} \\
&= \epsilon_{\beta\alpha} \epsilon^{\beta\gamma} \phi_\gamma \ ,
\end{aligned}
$$

so that $\epsilon_{\beta\alpha}\epsilon^{\beta\gamma} = \delta_\alpha^\gamma$. This implies in particular that the following relation holds numerically*

$$\epsilon_{\alpha\beta} = \epsilon^{\alpha\beta} \ , \tag{A.5}$$

so that, for example, $\epsilon_{12} = 1$. Note that $\epsilon_{\alpha\beta}$ is also correctly obtained from $\epsilon^{\alpha\beta}$ by lowering indices:

$$\epsilon_{\alpha\beta} = \epsilon^{\gamma\delta} \epsilon_{\gamma\alpha} \epsilon_{\delta\beta} \tag{A.6}$$

Then for example the first invariant can also be written as

$$
\begin{aligned}
\psi C \phi &= \psi_\alpha \epsilon^{\alpha\beta} \phi_\beta \\
&= \psi_\alpha \phi^\alpha \\
&= -\psi^\alpha \phi_\alpha \ .
\end{aligned}
$$

Under infinitesimal rotations these quantities transform as follows

$$\delta\phi_\alpha = \delta_i \sigma^i{}_\alpha{}^\beta \phi_\beta \tag{A.7}$$

* A relation holds *numerically* if the left- and right hand side yield the same answer for the same index values, regardless of the position of the indices. If the indices have different positions, the objects transform differently as tensors under rotations, so they are not *identical*, they are just *numerically* the same.

Here δ_i are three infinitesimal $SO(3)$ parameters, and $\sigma^i{}_\alpha{}^\beta$ is numerically*equal to $\tau^i_{\alpha\beta}$. For the quantity with upper indices we find then

$$\begin{aligned}\delta\phi^\alpha &= \epsilon^{\alpha\gamma}\delta_i\sigma^i{}_\gamma{}^\beta\phi_\beta \\ &= \epsilon^{\alpha\gamma}\delta_i\sigma^i{}_\gamma{}^\beta\phi^\rho\epsilon_{\rho\beta} \\ &= \delta_i\hat{\sigma}^{i,\alpha}{}_\rho\phi^\rho\end{aligned}$$

We see that numerically

$$\hat{\sigma}^{i,\alpha}{}_\rho = \epsilon^{\alpha\gamma}\sigma^i{}_\gamma{}^\beta\epsilon_{\rho\beta} = [C\tau_i(C^T)]_{\alpha\rho} = [C\tau_i C^\dagger]_{\alpha\rho} = -(\tau^i)^*_{\alpha\rho} \tag{A.8}$$

Hence we see that ϕ^α transforms according to the conjugate representation, so that we see once again that $\phi_\alpha\phi^\alpha$ is indeed invariant. Note that $\sigma^i{}_\alpha{}^\beta$ is an invariant tensor, if we transform all its three indices correctly:

$$\begin{aligned}\delta\sigma^j{}_\alpha{}^\beta &= \delta_i\left(-i\epsilon_{ijk}\,\sigma^k{}_\alpha{}^\beta + \sigma^i_\alpha{}^\gamma\sigma^j{}_\gamma{}^\beta + \hat{\sigma}^{i,\beta}{}_\gamma\sigma^j{}_\alpha{}^\gamma\right) \\ &= \delta_i\left(-i\epsilon_{ijk}\,\sigma^k{}_\alpha{}^\beta + [\sigma^i,\sigma^j]_\alpha{}^\beta\right) = 0\end{aligned}$$

A.2 The Lorentz Group

All particles belong to representations of the Lorentz group, $SO(3,1)$, which is a non-compact real form of $SO(4)$. The Lie-algebra of $SO(4)$ is isomorphic to $SU(2)\times SU(2)$. The finite-dimensional unitary representations of $SO(4)$ are thus labeled by two numbers (s_1, s_2) which must be integer or half-integer. Here s_i is the $SU(2)$ spin. However, to agree with the notation we use for other groups, we will use the dimension $d_i = 2s_i + 1$ to denote the representations.

The finite dimensional (but not unitary) representation matrices of the non-compact form $SO(3,1)$ can be obtained from those of $SO(4)$ by choosing a suitable basis of hermitean generators (as defined in appendix B, under *real forms*) and multiplying three of the generators by i. If we write $SO(4)$ as $SU(2)\times SU(2)$ and denote the first and second set of $SU(2)$ generators as R^a and S^a, then this basis choice can be written down explicitly: $T^a = (R^a + S^a)$, $\hat{T}^a = (R^a - S^a)$. In $SO(3,1)$ the first set generates the rotations, the second, multiplied by i, the boosts. Note that the set T^a closes under commutation, whereas the commutator of two infinitesimal boosts gives an infinitesimal rotation.

The extra factor i has no influence on the choice of representations. Just as for $SO(4)$, they are labelled by two integers (d_1, d_2), the dimensions of the $SU(2)$ representations. The total dimension of the representation (d_1, d_2) is $d_1 d_2$. We can use a pair of indices (i_1, i_2), $i_1 = 1,\ldots,d_1$, $i_2 = 1,\ldots,d_2$ as labels on the representation space. Then the aforementioned matrices are explicitly

$$\begin{aligned}R^a_{(i_1,i_2),(j_1,j_2)} &= M^a_{i_1 j_1}(d_1)\times\delta_{i_2 j_2} \\ S^a_{(i_1,i_2),(j_1,j_2)} &= \delta_{i_1 j_1}\times M^a_{i_2 j_2}(d_2)\end{aligned} \tag{A.9}$$

* We use the notation σ for tensors and τ for matrices. The position of the indices of σ indicate a transformation property; the position of indices of τ has no special meaning.

where $M^a(d_i)$ is the $SU(2)$ representation matrix in the representation with dimension d_i. For example, $M^a(\mathbf{1}) = 0$ and $M^a(\mathbf{2}) = \tau^a$ (as before we omit the usual normalization factor $\frac{1}{2}$ here).

The most common $SO(3,1)$ representations are $(\mathbf{1},\mathbf{1})$ for scalars, $(\mathbf{2},\mathbf{1})$ for left-handed spinors, $(\mathbf{1},\mathbf{2})$ for right-handed ones, and $(\mathbf{2},\mathbf{2})$ for vectors. The representation matrices (R^a, S^b), $(a, b = 1, 2, 3)$, for these four representations are respectively $(R^a, S^b) = (0,0), (\tau^a, 0)$, $(0, -(\tau^b)^*)$ and $(\tau^a, -(\tau^b)^*)$. Here we have denoted the full matrices Eqn. (A.9) for simplicity by just specifying the pair of $SU(2)$ matrices $(M^a(d_1), M^b(d_2))$ out of which they are built. Note that in the second $SU(2)$ factor we use the complex conjugate representation. This is just a convention. Four-vector indices are denoted here by letters $\mu, \nu \ldots$, left-handed spinors by $\alpha, \beta, \ldots$ and right-handed ones by $\dot{\alpha}, \dot{\beta}, \ldots$.

The tensor product of $(\mathbf{2},\mathbf{1})$ and $(\mathbf{1},\mathbf{2})$ contains $(\mathbf{2},\mathbf{2})$, and therefore there must exists an invariant tensor $\sigma^\mu_{\alpha\dot{\beta}}$. It is customary to express that tensor in terms of invariant tensors of the $SO(3)$ subgroup of $SO(3,1)$ corresponding to space rotations. This subgroup is precisely the diagonal subgroup of the two $SU(2)$'s, and has generators $T^a = (R^a + S^a)$.

Under this subgroup the two kinds of spinors are transforming as

$$\delta\psi_\alpha = \delta_i \sigma^i{}_\alpha{}^\beta \psi_\beta \tag{A.10}$$

and

$$\delta\psi_{\dot{\alpha}} = \delta_i \sigma^i{}_{\dot{\alpha}}{}^{\dot{\beta}} \psi_{\dot{\beta}} \tag{A.11}$$

We use the convention that, numerically, $\sigma^i{}_\alpha{}^\beta = \tau^i_{\alpha\beta}$; $\sigma^{i,\dot{\alpha}}{}_{\dot{\beta}} = \tau^i_{\dot{\alpha}\dot{\beta}}$. Then $\sigma^{i,}{}_{\dot{\alpha}}{}^{\dot{\beta}} = -(\tau^i_{\dot{\alpha}\dot{\beta}})^*$. This is a consequence of our convention to use the (equivalent) complex conjugate representation $(\mathbf{1},\mathbf{2}^*)$ for the right-handed spinors.

In terms of the diagonal $SU(2)$ the corresponding tensor product is $\mathbf{2}\times\mathbf{2} = \mathbf{1}+\mathbf{3}$, which implies in particular that a four-vector decomposes into a three-vector and a singlet. Left- and right-handed spinors are indistinguishable with the $SU(2)$ subgroup; both become doublets, and the distinction between dotted and undotted indices disappears, provided that we remember to treat dotted upper (lower) indices as undotted lower (upper) indices.

The invariant tensor $\sigma^\mu_{\alpha\dot{\beta}}$ coupling $(\mathbf{2},\mathbf{1}) \times (\mathbf{1},\mathbf{2})$ to $(\mathbf{2},\mathbf{2})$ can now be written as

$$\sigma^\mu_{\alpha\dot{\beta}} = (\pm\mathbf{1}, \vec{\tau})^\mu_{\alpha\dot{\beta}} \ . \tag{A.12}$$

This notation indicates that the space components $\sigma^i_{\alpha\dot{\beta}}$ are numerically equal to the Pauli matrices. As tensors these are indeed invariant. This follows from Eq. (A.9), plus the fact that the dotted lower indices transform under the rotation subgroup as undotted upper indices.* The relative normalization between the space and time components does not follow from these $SO(3)$-based arguments, but can be derived by requiring that $\psi^\alpha \sigma^\mu_{\alpha\dot{\beta}} \chi^{\dot{\beta}}$ transforms as a vector V^μ. This fixes the relative factor up to a sign, which is a convention.

* Here we are following the standard conventions. It might be preferable to lower and raise systematically all dotted indices on all quantities, so that the transform in the same way under rotations.

From the properties of the Pauli matrices one easily derives, in a metric-independent notation,

$$\sigma^{\mu}_{\alpha\dot\beta}\sigma^{\nu,\alpha\dot\beta} = 2\eta^{00}\eta^{\mu\nu} \quad ,\sigma^{\mu}_{\alpha\dot\beta}\sigma^{\gamma\dot\delta}_{\mu} = 2\eta^{00}\delta^{\gamma}_{\alpha}\delta^{\dot\delta}_{\dot\beta} \tag{A.13}$$

Note that these relations fix the relative normalization between the space-like and time-like components of σ^{μ}.

An important difference between $SO(4)$ and $SO(3,1)$ is the behavior of the spinors under complex conjugation. Starting with Eqn. (A.9) we can compute the six $SO(4)$ and $SO(3,1)$ representation matrices on these two-dimensional spaces. In both cases the representations $(\mathbf{2},\mathbf{1})$ and $(\mathbf{1},\mathbf{2})$ are obtained from $\vec{R}=\vec{\tau}, \vec{S}=0$ or $\vec{R}=0$, $\vec{S}=-(\vec{\tau})^{*}$ respectively (note that we use, as before, the convention that the second $SU(2)$ transforms with the complex conjugate representation). Then we take the combinations $T^a = R^a + S^a$ and $\hat{T}^a = R^a - S^a$ to get the generators of $SO(4)$, and $T^a = R^a + S^a$ and $i\hat{T}^a = i(R^a - S^a)$ to get the generators of $SO(3,1)$.

In $SO(4)$ the 6 representation matrices are $\vec{R}+\vec{S}=\vec{\tau}$ and $\vec{R}-\vec{S}=\vec{\tau}$ for the representation $(\mathbf{2},\mathbf{1})$ and $\vec{R}+\vec{S}=-\vec{\tau}^{*}$ and $\vec{R}-\vec{S}=\vec{\tau}^{*}$ for $(\mathbf{1},\mathbf{2})$. If we take the conjugate of the set of matrices the $(\vec{R}+\vec{S},\vec{R}-\vec{S})=(\vec{\tau},\vec{\tau})$ we get $(-\vec{\tau}^{*},-(\vec{\tau})^{*})$, which is equivalent to the original. Hence the spinor representations of $SO(4)$ are self-conjugate (and in fact pseudo-real). This is summarized below. The symbol $\sim$ denotes equivalence in $SU(2)$, *i.e.* $C^{\dagger}\tau^{a}C = -(\tau^{a})^{*}$.

$$\begin{array}{llll}
(\mathbf{2},\mathbf{1}): & (T,\hat{T}) = \quad (\vec{\tau},\vec{\tau}); & \text{conjugate} & (-\vec{\tau}^{*},-\vec{\tau}^{*}) \sim (\vec{\tau},\vec{\tau}) \\
(\mathbf{1},\mathbf{2}): & (T,\hat{T}) = \ (-\vec{\tau}^{*},\vec{\tau}^{*}); & \text{conjugate} & (\vec{\tau},-\vec{\tau}) \sim (-\vec{\tau}^{*},\vec{\tau}^{*})
\end{array}$$

In $SO(3,1)$ the 6 generators $\vec{R}+\vec{S}$ and $i(\vec{R}-\vec{S})$ are represented by the matrices $\vec{\tau}$ and $i\vec{\tau}$ for the $(2,1)$ representation and by the matrices $-\vec{\tau}^{*}$ and $i\vec{\tau}^{*}$ for the $(1,2)$ representation. Hence in this case the two spinor representations are conjugate to each other. So now we get the following (the $SU(2)$ equivalence relation is not needed here)

$$\begin{array}{llll}
(\mathbf{2},\mathbf{1}): & (T,i\hat{T}) = \quad (\vec{\tau},i\vec{\tau}); & \text{conjugate} & (-\vec{\tau}^{*},i\vec{\tau}^{*}) \\
(\mathbf{1},\mathbf{2}): & (T,i\hat{T}) = \ (-\vec{\tau}^{*},i\vec{\tau}^{*}); & \text{conjugate} & (\vec{\tau},i\vec{\tau})
\end{array}$$

If ψ^{α} is a spinor in the representation $(\mathbf{2},\mathbf{1})$, $(\psi^{\alpha})^{*}$ is a spinor in the representation $(\mathbf{1},\mathbf{2}^{*})$. To make the transformation properties explicit we define

$$\bar{\psi}^{\dot\alpha} = (\psi^{\alpha})^{*} \tag{A.14}$$

In other words, we introduce a new symbol $\bar{\psi}$ whose components are numerically equal to those of ψ^{*}, but we give $\bar{\psi}$ a dotted index to indicate that it transforms as the $SO(3,1)$ spinor of opposite "chirality" (by definition, the chirality is $+1$ for the representation $(2,1)$ and -1 for $(1,2)$. Now we just have to be careful about the position of the index: upper or lower. This can be deduced from the transformation of the left- and right-hand side under $SU(2)$ rotations. We have

$$\delta\psi^{\alpha} = \delta_{i}\sigma^{i,\alpha}{}_{\beta}\psi^{\beta} \tag{A.15}$$

Then the complex conjugate spinor transforms as (note the usual minus sign for infinitesimal transformations of complex conjugates)

$$\delta(\psi^{\alpha})^* = -\delta_i(\sigma^{i,\alpha}{}_{\beta})^*(\psi^{\beta})^* \tag{A.16}$$

On the other hand, the left-hand side transforms as

$$\delta\bar{\psi}^{\dot\alpha} = \delta_i\sigma^{i,\dot\alpha}{}_{\dot\beta}\bar{\psi}^{\dot\beta} \tag{A.17}$$

and we have seen earlier that $\sigma^i_{\alpha}{}^{\beta} = -(\sigma^i_{\dot\alpha}{}^{\dot\beta})^*$. The same relation holds when we replace all upper indices by lower ones and vice-versa.

Using the conjugate spinor we can write down a Lorentz invariant kinetic term

$$i\sigma^0\psi^{\alpha}\sigma^{\mu}_{\alpha\dot\beta}\partial_{\mu}\bar{\psi}^{\dot\beta} \tag{A.18}$$

Here we are allowing for some convention-dependence. A proper fermion Lagrangian must lead to a positive definite Hamiltonian; this requires the terms involving the time derivatives to have the form $i\psi^{\dagger}\frac{d}{dt}\psi$. This is why the factor $\sigma^0 \equiv \sigma^0_{11}(=\sigma^0_{22})$ is present (we assume that $\sigma^0 = \bar\sigma^0 = \pm\mathbf{1}$). The canonical choice is $\sigma^0 = \eta^{00}$, so that $\sigma_0 \equiv \sigma^0\eta_{00} = 1$ in both metrics.

We can also write down a mass term, but only by combining ψ with itself; $(\mathbf{2},\mathbf{1})$ couples with itself to a singlet, but not with $(\mathbf{1},\mathbf{2}^*)^*$ Such a mass term is of the form $m\psi^{\alpha}\psi_{\alpha}$. Explicitly this is proportional to $\psi_{\alpha}\epsilon^{\alpha\beta}\psi_{\beta}$, and this vanishes if ψ is a commuting object. Up to now it was, since we have only introduced it as a vector in a two-dimensional spinor space. In physics spin-$\frac{1}{2}$ particles should however be anti-commuting, which can be achieved either by making ψ Grassmann-valued, or by making it an anti-commuting operator in a Hilbert space. In either case we the rôle of complex conjugation is replaced by hermitean conjugation (*i.e.* $\phi^{\dagger}\phi$ is a positive definite quantity, analogous to $\phi^*\phi$ for complex numbers). Hence we define $\bar\psi^{\dot\alpha} = (\psi^{\alpha})^{\dagger}$. Then the mass term is (including the necessary hermitean conjugate term and a normalization for later purposes)

$$-\tfrac{1}{2}m(\psi^{\alpha}\psi_{\alpha} + \bar\psi_{\dot\alpha}\bar\psi^{\dot\alpha}) \ . \tag{A.19}$$

If ψ has in addition to its space-time properties non-trivial transformation properties under some local (or global) unitary symmetry, then this mass term can only be invariant under that symmetry if the representation of R can couple with itself to a singlet. Furthermore the coupling must involve a symmetric tensor, since otherwise the mass term vanishes. This means that the representation must be real. This kind of mass term is called a Majorana mass term.

One cannot write down a mass term for a spinor in a complex or pseudo-real representation R. What one can do is introduce a second spinor χ in the representation R^*. The reducible representation $R + R^*$ allows a real basis of fields $\Theta^{i_1} = \frac{1}{\sqrt{2}}(\chi^i + \psi^i)$,

* Note that the fact that R and R^* can be coupled to a singlet holds for unitary representations, as a consequence of $U^{\dagger}U = \mathbf{1}$; the $SO(3,1)$ spinor representations (and all other finite dimensional representations) are however not unitary.

$\Theta^{i_2} = \frac{i}{\sqrt{2}}(\chi^i - \psi^i)$, where i is the gauge index and the spinor indices are suppressed. We can view this as a spinor in a 2dim (R) dimensional real representation. Hence one can describe this system exactly as above, with the same kinetic terms and a Majorana mass term.

This is not the usual description, however. If we write out the full kinetic and Majorana mass term we get, suppressing all spinor indices

$$i\sigma^0\Theta^t\sigma^\mu\partial_\mu\bar{\Theta}^t - \tfrac{1}{2}m(\Theta^t\Theta^t + \text{c.c}) \tag{A.20}$$

where t stands for (i_1, i_2). Expressing this in terms of ψ and χ we get

$$i\sigma^0\psi^i\sigma^\mu\partial_\mu\bar{\psi}^i + i\sigma^0\chi^i\sigma^\mu\partial_\mu\bar{\chi}^i - m(\psi^i\chi^i + \text{c.c}) \tag{A.21}$$

Note that $\psi^i\chi^i \equiv \psi^{i,\alpha}\chi^i_\alpha = \chi^i\psi^i$ using the fact that the spinors anti-commute.

Now we re-write the kinetic terms of χ in the following way (this time suppressing the gauge index i, but showing the spinor index)

$$\begin{aligned} i\sigma^0\chi^\alpha\sigma^\mu_{\alpha\dot{\beta}}\partial_\mu\bar{\chi}^{\dot{\beta}} &= -i\sigma^0\partial_\mu\bar{\chi}^{\dot{\beta}}(\sigma^\mu)^T_{\dot{\beta}\alpha}\chi^\alpha \\ &= i\sigma^0\bar{\chi}_{\dot{\beta}}\bar{\sigma}^{\mu,\dot{\beta}\alpha}\partial_\mu\chi_\alpha \ , \end{aligned} \tag{A.22}$$

where in the first step anti-commutativity was used and in the second step integration by parts. Furthermore we introduced the new tensor

$$\bar{\sigma}^{\mu,\dot{\beta}\alpha} \equiv \epsilon^{\dot{\delta}\dot{\beta}}\epsilon^{\gamma\alpha}(\sigma^\mu)^T_{\dot{\delta}\gamma} = (C^\dagger(\sigma^\mu)^*C)^{\dot{\beta}\alpha} \ , \tag{A.23}$$

where in the last step we used Hermiticity of σ^μ. Finally, using $C^\dagger C = 1$ and $C^\dagger\vec{\sigma}^*C = -\vec{\sigma}$ we find that $\bar{\sigma}^\mu$ is numerically equal to $(\sigma^0, -\vec{\sigma})$. For two anti-commuting spinors ψ and χ one has

$$\psi\sigma^\mu\bar{\chi} = -\bar{\chi}\bar{\sigma}^\mu\psi \tag{A.24}$$

The kinetic terms, $i\sigma^0\psi^i\sigma^\mu\partial_\mu\bar{\psi}^i + i\sigma^0\bar{\chi}\bar{\sigma}^\mu\partial_\mu\chi$ can also be written as

$$i\bar{\Psi}\gamma^\mu\partial_\mu\Psi \ , \tag{A.25}$$

with

$$\Psi = \begin{pmatrix}\chi_\alpha \\ \bar{\psi}^{\dot{\alpha}}\end{pmatrix} \ , \qquad \bar{\Psi} = \begin{pmatrix}\psi^\alpha \ , & \bar{\chi}_{\dot{\alpha}}\end{pmatrix} \ , \tag{A.26}$$

and

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^0\sigma^\mu \\ \sigma^0\bar{\sigma}^\mu & 0 \end{pmatrix} \tag{A.27}$$

so that

$$\gamma_0 = \gamma^0 = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \ . \tag{A.28}$$

We define

$$\bar{\Psi} = \Psi^\dagger\gamma^0 \tag{A.29}$$

Note that all expressions involving Dirac γ matrices are metric dependent, whereas all results involving σ's are written here in metric-independent form.

The mass term can be written as

$$-m\bar{\Psi}\Psi \tag{A.30}$$

The γ-matrices satisfy the Clifford algebra

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} \ , \tag{A.31}$$

and other representations exist. In the present representation (and in fact in any commonly used representation with this choice of metric) they are hermitean ($\mu = 1, 2, 3$) or anti-hermitean ($\mu = 0$, the time direction); γ_4 is hermitean.

Even if a spinor ψ is in a real representation of all symmetry groups it is customary do introduce a Dirac spinor

$$\Psi = \begin{pmatrix} \psi_\alpha \\ \bar{\psi}^{\dot{\alpha}} \end{pmatrix} \ , \tag{A.32}$$

and to write down the standard action. Only in this case one has to include an extra factor $\frac{1}{2}$, since otherwise one would get the kinetic terms Eq. (A.18) twice. The correct action for a Majorana fermion is thus

$$\tfrac{i}{2}\bar{\Psi}\gamma^\mu\partial_\mu\Psi \ - \tfrac{1}{2}m\bar{\Psi}\Psi \tag{A.33}$$

A Majorana spinor satisfies the condition

$$\bar{\Psi} = \pm\Psi^T C \ , \tag{A.34}$$

where C is the charge conjugation matrix

$$C = \begin{pmatrix} i\sigma_2 & 0 \\ 0 & -i\sigma_2 \end{pmatrix} \ . \tag{A.35}$$

The γ-matrices derived here are in a representation of the Clifford algebra which is not the most common one. It is called the Weyl representation. For example, in [5] a different representation is used.

A Dirac spinor can be projected onto its two components using the matrix γ_5 defined as $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3$. It is hermitean, its square is $\mathbf{1}$, and it commutes with all $\gamma_\mu, \mu = 0, \ldots 4$. In the explicit representation given above one has

$$\gamma_4 = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \quad , \qquad \gamma_5 = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix} \tag{A.36}$$

The left and right chiral projection operators are

$$P_L = \frac{1}{2}(1 + \gamma_5); \quad P_R = \frac{1}{2}(1 - \gamma_5) \ . \tag{A.37}$$

They satisfy $P_RP_R = P_R; P_LP_L = P_L$ and $P_LP_R = 0$. The left and right-handed components of a field are defined as $\psi_L = P_L\psi; \quad \psi_R = P_R\psi$. Due to the projections ψ_L and ψ_R are effectively two-component spinors, called *Weyl spinors.* These are precisely the spinors χ and $\bar{\psi}$ introduced above.

Note that $\bar{\psi}_L = \bar{\psi}P_R$. The flip in chirality occurs because have to commute P_L through γ_4. Hence the non-vanishing bi-linears are $i\bar{\psi}_L\gamma_\mu\psi_L$, $\bar{\psi}_R\psi_L$, and terms with all L's and R's interchanged. Thus the vector current (to which gauge bosons couple) preserves chirality, but the mass term does not. Another combination that does not preserve chirality is $\bar{\psi}_L[\gamma_\mu, \gamma_\nu]\psi_R$, to which the magnetic moment is proportional.

For an arbitrary Dirac spinor one defines

$$\bar{\Psi} = -(\Psi^c)^T C \ , \tag{A.38}$$

where Ψ^c is the charge conjugate spinor. A Majorana spinor is thus defined by $\Psi = \pm\Psi^c$. In the absence of a mass term there is not really any difference between Majorana and Weyl spinors. We may write

$$\frac{i}{2}\bar{\Psi}\gamma^\mu\partial_\mu\Psi \ = \frac{i}{2}\bar{\Psi}_L\gamma^\mu\partial_\mu\Psi_L \ + \frac{i}{2}\bar{\Psi}_R\gamma^\mu\partial_\mu\Psi_R \ \ , \tag{A.39}$$

and then substitute $\bar{\Psi}_R = -((\Psi^c)_L)^T$ in the second term. Using the Majorana property plus a little algebra (which is done explicitly in chapter 5 one finds that the second term is now transformed into the first one. Hence for Majorana fermions

$$\frac{i}{2}\bar{\Psi}\gamma^\mu\partial_\mu\Psi \ = i\bar{\Psi}_L\gamma^\mu\partial_\mu\Psi_L \ = i\bar{\Psi}_R\gamma^\mu\partial_\mu\Psi_R \ \ . \tag{A.40}$$

The difference between Majorana and Weyl fermions becomes essential if one assigns them to representations of local or global symmetries, and writes down mass terms. For Majorana fermions the representations must be real, and masses are allowed, while for Weyl fermions the representation can be complex, but then one cannot write down an invariant mass term.

Majorana and Weyl spinors both have two on shell degrees of freedom. In other words both the Majorana condition $\bar{\Psi} = -\Psi^T C$ and the Weyl condition $\Psi = P_L\Psi$ (or $\Psi = P_R\Psi$) reduce the number of degrees of freedom of a Dirac spinor from 4 to two, but one cannot reduce the number of degrees of freedom further. This is due to the fact that the $SO(3,1)$ spinor representation is complex, and hence requires always two real degrees of freedom.

In dimensions other than 4 this can be different. For example if $D = 10$ modulo 8 the $SO(D-1,1)$ spinor representations are real, and one can impose simultaneously Weyl and Majorana conditions.

B Lie Algebras

Here we collect some formulas and conventions for Lie-algebras. This is not a review of group theory, but rather an "encyclopedic dictionary" of some relevant facts with few explanations.

B.1 Classification of Lie Algebras

The algebra. We will mainly use compact groups (see below). Their Lie algebras can be characterized by a set of dim (A) hermitean generators $T^a, a = 1, \text{dim }(A)$, where A stands for "adjoint". Provided a suitable basis choice is made, the generators satisfy the following algebra

$$[T^a, T^b] = i f^{abc} T_c \ , \tag{B.1}$$

with structure constants f_{abc} that are real and completely anti-symmetric.

Exponentiation. Locally, near the identity, the corresponding Lie group can be obtained by exponentiation

$$g(\alpha) = e^{i\alpha^a T^a} \ . \tag{B.2}$$

The global properties of the group, involving element not "close" to 1, are not fully described by the Lie-algebra alone, but will not be discussed here. The space formed by all the group elements is called the *group manifold*.

Real forms. A Lie-algebra is a vector space of dimension dim (A) with an additional operation, the commutator. An arbitrary element of the vector space has the form $\sum_a \alpha_a T^a$. In applications to physics α_a is either a real or a complex number. If the coefficients α^a are all real and the generators Hermitean, the group manifold is a compact space. For a given compact group there is a unique complex Lie-algebra, which is obtained simply by allowing all coefficients α_a to be complex. Within the complex algebra there are several real sub-algebras, called *real forms*. The generators of such a sub-algebra can be chosen so that Eq. (B.1) is satisfied with all structure constants real, but with generators that are not necessarily Hermitean. One can always obtain the real forms from the compact real form (which has hermitean generators) by choosing a basis so that the generators split into two sets, $\mathcal{H}$ and $\mathcal{K}$, so that $[\mathcal{H}, \mathcal{H}] \in \mathcal{H}$ and $[\mathcal{K}, \mathcal{K}] \in \mathcal{H}$. Then one may consistently replace all generators $K \in \mathcal{K}$ by iK without affecting the reality of the coefficients f^{abc}. The most common case in physics are the real forms $SO(n, m)$ of the compact real form $SO(n + m)$. Most of the following results hold for the compact real form of the algebra, unless an explicit statement about non-compact forms is made.

The classical Lie groups. The group $SU(N)$ is the group of unitary $N \times N$ matrices with determinant 1; $SO(N)$ is the group of real orthogonal matrices with determinant 1, and $Sp(2r)$ the group of real $2r \times 2r$ matrices S that satisfies $S^T M S = M$, where M is a matrix which is block-diagonal in term of 2×2 blocks of the form $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Mathematicians (and some physicist) write $Sp(r)$ instead of $Sp(2r)$.

Simple Lie-algebras. Lie-algebras are in general a "product" (or, more accurately, a direct sum) of a semi-simple Lie-algebra and some $U(1)$'s. The latter require no further discussion. Semi-simple algebras are a product of various simple ones; the simple Lie-algebras have been classified completely, see below.

The Cartan sub-algebra. This is the maximal set of commuting generators of the simple algebra. All such sets can be shown to be equivalent. The dimension of this space is called the *rank* (denoted r) of the algebra.

Roots. If we denote the Cartan sub-algebra generators as $H_i, i = 1, \ldots, r$, then the remaining generators can be chosen so that

$$[H_i, E_{\vec{\alpha}}] = \alpha^i E_{\vec{\alpha}} \ . \tag{B.3}$$

The eigenvalues with respect to the Cartan sub-algebra are vectors in a space of dimension r. We label the remaining generators by their eigenvalues $\vec{\alpha}$. These eigenvalues are called the *roots* of the algebra.

Positive roots. A positive root is a root whose first component α_1 is positive in some fixed basis. This basis must be chosen so that $\alpha_1 \neq 0$ for all roots.

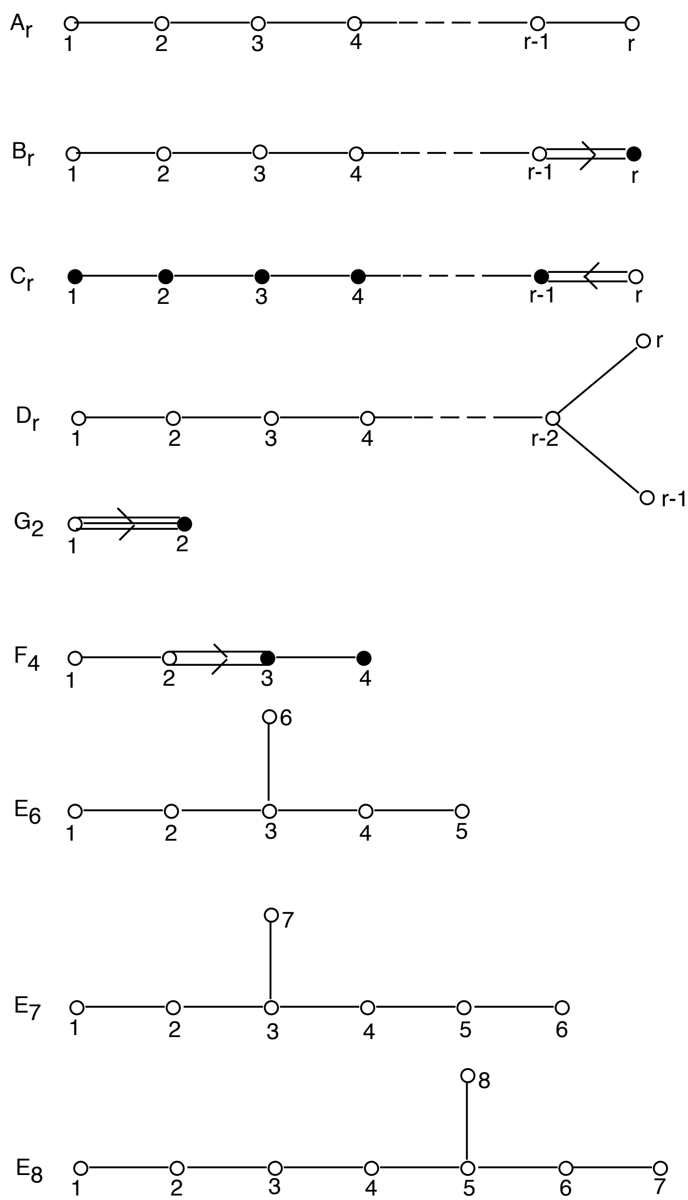
Simple roots. Simple roots are positive roots that cannot be written as positive linear combinations of other positive roots. There are precisely r of them. They form a basis of the vector space of all the roots. The set of simple roots of a given algebra is unique up to $O(r)$ rotations. In particular it does not depend on the choice of the Cartan sub-algebra or the basis choice in "root space". This set is thus completely specified by their relative lengths and mutual inner products. The inner product used here, denoted $\vec{\alpha} \cdot \vec{\beta}$, is the straightforward Euclidean one.

The Cartan matrix. The Cartan matrix is defined as

$$A^{ij} = 2\frac{\vec{\alpha}_i \cdot \vec{\alpha}_j}{\vec{\alpha}_j \cdot \vec{\alpha}_j} \ , \tag{B.4}$$

where $\vec{\alpha}_i$ is a simple root. This matrix is unique for a given algebra, up to permutations of the simple roots. One of the non-trivial results of Cartan's classification of the simple Lie algebras is that all elements of A are integers. The diagonal elements are all equal to 2 by construction; the off-diagonal ones are equal to $0, -1, -2$ or -3.

Dynkin diagrams. are a graphical representation of the Cartan matrix. Each root is represented by a dot. The dots are connected by n lines, where n is the maximum of $|A_{ij}|$ and $|A_{ji}|$. If $|A_{ij}| > |A_{ji}|$ an arrow from root i to root j is added to the line. The simple algebras are divided into 7 classes, labeled A–G , with Dynkin diagrams as shown below.



Long and short roots. If a line from i to j has an arrow, $A_{ij} \neq A_{ji}$ and hence the lengths of roots i and j are not the same. An arrow points always from a root to another root with smaller length. Lines without arrows connect roots of equal length. There is at most one line with an arrow per diagram, and therefore there are at most two different lengths. This is not only true for the simple roots, but for all roots. One frequently used convention is to give all the long roots length-squared equal to two. Then the short roots have length 1 if they are connected to the long ones by a double line, and length-squared $\frac{2}{3}$ if they are connected by a triple line. Often the short roots are labeled by closed dots, and the long ones by open dots, although this is strictly speaking superfluous. If all roots have the same length the algebra is called *simply laced*. This is true for types A,D and E.

Realizations. The compact Lie-algebras corresponding to types $A - D$ are realized by the algebras $SU(n), SO(n)$ and $Sp(n)$. The correspondence is as follows

$$\begin{array}{rc} A_r: & SU(r+1) \\ B_r: & SO(2r+1) \\ C_r: & Sp(2r) \\ D_r: & SO(2r) \end{array}$$

There is no such simple characterization for the algebras of types E, F and G, the exceptional algebras.

B.2 Representations.

A set of unitary $N \times N$ matrices satisfying the algebra (B.1) is said to form a (unitary matrix) representation of dimension N.

Equivalence of representations. If a set of hermitean generators T^a satisfy the algebra, then so do $\tilde{T}^a = U^\dagger T^a U$, if U is unitary. Then T^a and $\tilde{T}^a$ are called equivalent.

Real, complex and pseudo-real representations. The *complex conjugate representation* is the set of generators $-(T^a)^*$, which obviously satisfy the algebra if T^a does. A representation is real if a basis exists so that for all a $-(T^a)^* = T^a$ (in other words, if a $\tilde{T}^a = U^\dagger T^a U$ exists so that all generators are purely imaginary). An example of a real representation is the adjoint representation, defined below. A representation is pseudo-real if it is not real, but only real up to equivalence, *i.e.* $-(T^a)^* = C^\dagger T^a C$ for some unitary matrix C. Otherwise a representation is called complex.

A frequently occurring example of a pseudo-real representation is the two-dimensional one of $SU(2)$. The generators are the Pauli matrices, and only σ_2 is purely imaginary. However, if one conjugates with $U = i\sigma_2$ the other two matrices change sign, so that indeed $-\sigma_i^* = U^\dagger \sigma_i U$.

Irreducible representations. If a non-trivial subspace of the vector space on which a representation acts is mapped onto itself (an "invariant subspace") the representation is called reducible. Then all T's can be simultaneously block-diagonalized, and each block is by itself a representation. If there are no invariant subspaces the representation is called irreducible.

Weights. In any representation the matrices representing the Cartan sub-algebra generators H_i can be diagonalized simultaneously. The space on which the representation acts decomposes in this way into eigenspaces with a set of eigenvalues $\vec{\lambda}$, *i.e.*$H_i v_{\vec{\lambda}} = \lambda_i v_{\vec{\lambda}}$. The $\vec{\lambda}$'s, which are vectors in the vector space spanned by the roots, are called weights. The vector space is usually called weight space.

Weight space versus representation space. We are now working in two quite different vector spaces: the r-dimensional weight space, and the N dimensional space on which the representation matrices act. The former is a real space, the latter in general a complex space. Often the vectors in the latter space are referred to as "states", a terminology borrowed from quantum mechanics. Although this may be somewhat misleading in applications to classical physics, it has the advantage of avoiding confusion between the two spaces.

Weight multiplicities. In the basis in which all Cartan sub-algebra generators are simultaneously diagonal each state in a representation are characterized by some weight vector λ. However, this does not characterize states completely, since several states can have the same weight. The number of states in a representation R that have weight λ is called the multiplicity of λ in R.

Coroots. Coroots $\hat{\alpha}$ are defined as $\hat{\alpha} = \frac{2\vec{\alpha}}{\alpha\cdot\alpha}$.

Dynkin labels. For any vector λ in weight space we can define Dynkin labels l_i as $l_i = \lambda \cdot \hat{\alpha}_i = 2\frac{\lambda\cdot\alpha_i}{\alpha_i\cdot\alpha_i}$. Since the simple (co)roots form a complete basis, these Dynkin labels are nothing but the components of a weight written with respect to a different basis. The advantage of this basis is that it can be shown that for any unitary representation of the algebra the Dynkin labels are integers.

Highest weights. Every irreducible representation of a simple Lie-algebra has a unique weight λ so that on the corresponding weight vector $E_\alpha v_\lambda = 0$ for all positive roots α. Then λ is called the highest weight of the representation. Its Dynkin labels are non-negative integers. Furthermore for every set of non-negative Dynkin labels there is precisely one irreducible representation whose highest weight has these Dynkin labels.

The irreducible representations of a simple Lie-algebra. They can thus be enumerated by writing a non-negative integer next to each node of the Dynkin diagram. The states in a representation can all be constructed by acting with the generators E_α on the highest weight state. This state always has multiplicity 1.

Special Representations.

- *Fundamental representations*
  The representations with Dynkin labels $(0, 0, , \ldots, 0, 1, 0, \ldots, 0)$ are called the fundamental representations.

- *The adjoint representation*
  The adjoint representation is the set of generators $(T^a)_{bc} = -if^{abc}$; it has dimension dim (A).

- *Vector representations*
  The $N \times N$ matrices that were used above to define the classical Lie groups form the vector representation of those groups; the expansion of these matrices around the identity yields the vector representation of the corresponding Lie-algebra.
- *Fundamental spinor representations*
  They are defined only for $SO(N)$. If N is odd, they have Dynkin label $(0, 0, \ldots, 0, 1)$. If N is even there are two fundamental spinor representations with Dynkin labels $(0, 0, 0, \ldots, 1, 0)$ and $(0, 0, 0, \ldots, 0, 1)$.

The Adjoint Representation. Generically, the adjoint representation is defined by means of the action of the group G on itself, namely by

$$g \to UgU^{-1}, \text{ with } g \in G \text{ and } U \in g. \tag{B.5}$$

For unitary Lie groups this implies an action by the group on its own Lie algebra

$$T^a \to UT^aU^\dagger \tag{B.6}$$

and if we write U in infinitesimal form this implies a action of the Lie algebra on itself

$$T^a \to T^a - \epsilon^b f^{bac} T^c \tag{B.7}$$

This defines a matrix representation of dimension equal to the dimension of the Lie algebra, and matrices that are related to the structure constants, $(T^b_{\text{adj}})_{ac} = -i(f^b)^{ac}$, where a labels the generators and b and c are the matrix indices. Note that according to Eqn (2.33), for constant θ, the gauge bosons are in this representation.

Tensor Products. If V_{i_1} transforms according to some representation R_1 and W_{i_2} according to some representation R_2, then obviously the set of products $V_{i_1}W_{i_2}$ forms a representation as well. This is called the tensor product representation $R_1 \times R_2$; it has dimension dim R_1dim R_2. This representation is usually not irreducible. It can thus be decomposed into irreducible representations:

$$R_1 \times R_2 = \sum_j N_{12j} R_j \ , \tag{B.8}$$

where N_{12j} is the number of times R_j appears in the tensor product.

B.3 Traces, Dimensions, Indices and Casimir operators

Dimension formula. The dimension of a representation is given by

$$\text{dim } (\Lambda) = \prod_{\text{positive roots } \tilde{\alpha}} \frac{(\vec{\Lambda} + \vec{\rho}) \cdot \vec{\alpha}}{\vec{\rho} \cdot \vec{\alpha}} \tag{B.9}$$

where $\vec{\rho}$ is called the Weyl vector. It has Dynkin labels $(1, 1, 1, \ldots, 1, 1)$.

The Casimir eigenvalue. The operator $T_a T_a$ is called the (quadratic) Casimir operator. It commutes with all generators, and is thus constant on an irreducible representation. The eigenvalue for a representation with highest weight Lambda is proportional to the number

$$C(\Lambda) = (\vec{\Lambda} + 2\vec{\rho}) \cdot \vec{\Lambda} \tag{B.10}$$

For the adjoint representation this yields $C(A) = 2g$, where g is the *dual Coxeter number*. It is equal to the following numbers for the simple algebras

Algebra	Value of g	Adjoint dimension
$SU(N)$	N	$N^2 - 1$
$SO(N),\ N > 3$	$N - 2$	$\frac{1}{2}N(N-1)$
$Sp(2N)$	$N + 1$	$N(2N+1)$
G_2	4	14
F_4	9	52
E_6	12	78
E_7	18	133
E_8	30	248

The Standard Normalization. The basis is chosen so that Tr $T^a T^b \propto \delta^{ab}$ the proportionality constant is fixed so that

$$\mathrm{Tr}_{\mathrm{adjoint}} T^a T^b = g\delta^{ab}\ . \tag{B.11}$$

This fixes the normalization of for all other representations.

Normalization of vector representations. The vector representation matrices are now normalized as follows

$$\begin{aligned} SU(N): &\qquad \mathrm{Tr}_{\mathrm{vector}} T^a T^b = \tfrac{1}{2}\delta^{ab} \\ SO(N),\ N > 3: &\qquad \mathrm{Tr}_{\mathrm{vector}} T^a T^b = \delta^{ab} \\ Sp(2N): &\qquad \mathrm{Tr}_{\mathrm{vector}} T^a T^b = \tfrac{1}{2}\delta^{ab} \end{aligned}$$

For $SO(3)$ we use the same normalization as for $SU(2)$. The correctly normalized generators of the $SU(2)$ vector representation (which is the $SO(3)$ spinor representation) are $\frac{1}{2}\tau^i$, where τ^i are the Pauli-matrices.

Symmetrized traces. The symmetrized trace of a representation is defined as

$$\mathrm{Str}\ T^{a_1} \ldots T^{a_k} = \frac{1}{k!} \sum_{\pi}\ \mathrm{Tr}\ T^{a_{\pi(1)}} \ldots T^{a_{\pi(k)}}\ , \tag{B.12}$$

where π is a permutation of the k labels, and the sum is over all permutations.

Tensors. A tensor $V_{i_1,\ldots,i_m}$ transforms by definition as

$$V'_{i_1,\ldots,i_m} = U^1_{i_1 j_1} \ldots U^m_{i_m j_m} V_{j_1,\ldots,j_m} \ , \tag{B.13}$$

where the indices ℓ of U^ℓ, i_ℓ, j_ℓ label different irreducible representations R_ℓ, and each U_ℓ is the representation matrix of a given group element in the representation R_ℓ. The labels i_ℓ and j_ℓ take values between 1 and dim R_ℓ.

Invariant Tensors. If $V' = V$ the tensor is called and invariant tensor. Well-known examples are the tensor δ_{i_1,i_2} if R_1 and R_2 are each others complex conjugate, and the structure constants f_{abc}. Another example is the set of representation matrices T^a_{ij} for any irreducible representation R. This is an invariant tensor if one transforms i according to R, j according to R^* and of course a according to the adjoint representation. In $SU(N)$ and $SO(N)$ the rank N anti-symmetric tensor $\epsilon_{i_1,\ldots,i_N}$, where all indices are vector or conjugate vector indices, is invariant.

Relation to tensor products. For every term in Eq. (B.8) there are N_{12j} distinct invariant tensors. The invariant tensors δ_{ij} and T^a_{ij} correspond to the first two terms in the tensor product $R \times R^* = 1 + A + \ldots$.

Rank two invariant tensors. The existence of an invariant tensor with two indices implies that the two corresponding representations R_1 and R_2 contain the identity in their tensor product. For every irreducible representation R_1 there is only one representation R_2 with that property. If R_2 is not equivalent to R_1 it is the complex conjugate of R_1, and the invariant tensor is δ_{i_1,i_2} as discussed above (provided one chooses complex conjugate bases). Otherwise R_1 is either real or pseudo-real. If R_1 is real δ_{ij} is an invariant tensor; if it is pseudo-real there exists an invariant tensor $C_{ij} = -C_{ji}$.The invariance implies then that the representation matrix U is conjugated by C: $U^* = C^{-1}UC$.

Symmetric invariant adjoint tensors. For each simple algebra of rank r all fully symmetric invariant tensors with adjoint indices can be expressed in terms of r basic tensors. The ranks (number of indices) of these tensors are as follows

The tensor of rank 2 is always $d^{ab} = \delta^{ab}$. For $SU(2)$ this is the only such tensor. Consequently whenever a symmetric tensor appears with four adjoint indices, it must be proportional to $\delta^{ab}\delta^{cd} + \delta^{ac}\delta^{bd} + \delta^{ad}\delta^{bc}$.

Indices. The n index of a representation R is defined as

$$\text{Str } T^{a_1} \ldots T^{a_n} = I_n(R) d^{a_1,\ldots a_n} + \ldots \ , \tag{B.14}$$

where d is one of the basic invariant tensors, and the dots represent combinations of lower order tensors. If there is no basic tensor of rank n, the index vanishes. These indices are

Algebra	Invariant tensor ranks
A_r	$2, 3, 4, \ldots, r, r+1$
B_r	$2, 4, 6, \ldots, 2r$
C_r	$2, 4, 6, \ldots, 2r$
D_r	$2, 4, 6, \ldots, 2r-2; r$
G_2	$2, 6$
F_4	$2, 6, 8, 12$
E_6	$2, 5, 6, 8, 9, 12$
E_7	$2, 6, 8, 10, 12, 14, 18$
E_8	$2, 8, 12, 14, 18, 20, 24, 30$

defined provided one has fixed a normalization for d; this can be done by fixing it for one representation with non-zero index. The second and third indices I_2 and I_3 are defined as

$$\begin{aligned} \text{Tr } T^a T^b &= \tfrac{1}{2} I_2(R) \delta^{ab} \\ \text{Str } T^a T^b T^c &= I_3(R) d^{abc} \ , \end{aligned}$$

For the algebras $SU(N)$ a convenient normalization of d^{abc} is such that $I_3 = 1$ for the fundamental representation.

Casimir operators. One can define r basic Casimir operators as

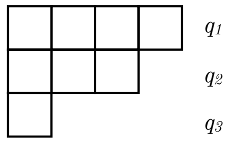
$$C_n(R) = d^{a_1,\ldots,a_n} T^{a_1} \ldots T^{a_n} \ , \tag{B.15}$$

where n is the rank of one of the basic invariant tensors. Because d is an invariant tensor, these operators commute with all the generators in any given representation. For the quadratic Casimir operator one has

$$C_2(R) = \frac{\dim A}{\dim R} I_2(R) \ . \tag{B.16}$$

B.4 Representations of $SU(N)$

The irreducible unitary representations of $SU(N)$ can be characterized in a very convenient way using Young tableaux. They are specified by a sequence of $N-1$ integers $(q_1, \ldots, q_{N-1})$, with $q_i \geq q_{i+1}$. This way of labelling representations can be derived in a straightforward way from the Dynking labelling. The sequences of integers are graphically represented by a diagram consisting of boxes forming an upside-down staircase, as in the following example



The dimension of a representation can be computed as follows. Take two copies of the picture. In the first one we write in all the boxes $N \pm j$, where j is the number of positions left or up from the diagonal (for $N + j$) or right or down from the diagonal (for $N - j$). In the second copy of the figure we put in every box the "hook length", which is the total number of boxes in the hook formed by the boxes to right and down from the box we are considering. The dimension is the product of the numbers in the first figure divided by the product in the second figure. In the example these two figures are as shown here

N	N+1	N+2	N+3
N-1	N	N+1	
N-2			

(a) numerator

6	4	3	1
4	2	1	
1			

(b) denominator

The diagrams have an interpretation in terms of symmetrization of tensor indices. One considers tensors with as many indices as there are boxes. This interpretation is straightforward for Young tableaux with just one row of length L or just one column of length M. The former corresponds to symmetric tensors of rank L, and the latter to anti-symmetric tensors of rank M. In other cases the interpretation in terms of symmetrization is more complicated.

The dimension formula continues to give correct results if we allow an N^{th} row q_N, but then the representation is the same as the one obtained by removing from the diagram all columns with N boxes. In particular, the anti-symmetric tensor of rank N is equivalent to the trivial representation of dimension 1. Complex conjugation of representations can be done by working out the "complement" of a Young tableau. Add to each column the boxes needed to complete these columns to N boxes. Then take the extra boxes, turn them around by 180° so that the lower right corner becomes the upper left corner, and the Young tableau obtained in this manner is the complex conjugate representation.

One may use these Young tableaux for various other purposes, such as computing tensor products, but this will not be discussed here. For other groups this method is far less effective. The resulting representations are not always irreducible, and some representations are not obtainable in this way (in particular spinor representations in $SO(N)$).

Note that building representations out of tensor products of vector representations is not the same as saying that the corresponding particle or field itself is somehow built out of vector representations. It is simply a mathematical method to obtain certain representations. For example, consider a particle that transforms according to the rank-2 anti-symmetric tensor representation. We can obtain that representation as a tensor product of two vector representations. The latter are just mathematical tools, and have no physical significance. It does not mean that the particle can somehow be physically "decomposed" in terms of fundamental particles in the vector representation. Indeed, if that were an option, we would have to worry what happens to the symmetric combination.

B.5 Subalgebras

Often in physics symmetries are only approximate, and hold only in special limits. Away from that limit only a sub-algebra remains as a symmetry. In the Standard Model this occurs when the Higgs mechanism breaks $SU(3) \times SU(2) \times U(1)_Y$ to $SU(3) \times U(1)_{\rm QED}$. In the high energy limit the symmetry is exact (or unbroken), whereas at low energy the subgroup is the relevant symmetry. Beyond the Standard Model this kind of situation may occur once again, with the Standard Model gauge group realized as a subalgebra of a larger algebra. The most popular option is $SU(5)$.

A subalgebra H is a set of generators written as a linear combination of the generators of an algebra G, such that their commutation relations close. One also says that H is embedded in G, and this is usually denoted as $H \subset G$. There is an analogous notion of groups and subgroups.

Particles and fields always belong to representations of symmetry groups. If the low energy symmetry group is a subgroup H of a larger group G, all particle representations of G decompose into particle representations of H. Suppose we have a particle or field ϕ^i, where i is an index on which G acts via a matrix representation. Then an infinitesimal transformation acts as follows

$$\phi_i \to \phi_i + i \sum_{a,j} \epsilon^a T^a_{ij} \phi_j \ , \tag{B.17}$$

where T^a are the generators of G in some representation R and ϵ^a is a set of infinitesimal parameters. Let us assume that the representation R acting on ϕ^i is irreducible.

The generators S^p of the subalgebra H have the form

$$S^p_{ij} = \sum_a P^p_a \, T^a_{ij} \ , \tag{B.18}$$

where P^p_a is a set of real numbers. the fact that R is irreducible means that any component of ϕ can be transformed into any other component by the action of G. If we consider a subalgebra that is not necessarily true anymore. In general we should expect that the space of fields ϕ_i splits into irreducible blocks. Each block consists of fields that are linear combinations of the ϕ_i. The fields within a block can be transformed by H transformations into each other, but not into other blocks. The original field ϕ splits into a set of fields, each forming a representation of H. The sum of the dimensions of these H representations is equal to the dimension of R. We write this as

$$R \to r_1 + \ldots + r_N \tag{B.19}$$

Some authors use the notation $\oplus$ instead of $+$.

To deal with subalgebras we have to know how the representations of G decompose into representations of H. These decompositions are called the *branching rules* of the representations with respect to the subalgebra embedding. If one knows the branching rules for a sufficiently non-trivial representation of G, this provides enough information for computing the branching rules of all other representations. The restriction to "sufficiently non-trivial" representations is necessary to exclude the trivial case, the branching rule $1 \to 1$, where 1 denotes the trivial representation with representation matrices $T^a = 0$. Obviously this trivial branching rule contains no useful information; in general there may be also exist non-trivial representations that do not contain enough information. The precise mathematical terminology for "sufficiently non-trivial" is *faithful*. We will not try to give precise definitions here, because we will focus on $SU(N)$ Lie algebras, and in that case the N-dimensional vector representation is faithful.

This implies that if we know the branching rule for the vector representation, then the decomposition of all other representations can be derived. There are several ways of doing that. The most obvious one is to explicitly block-diagonalize the representation matrices of H, starting with those of G. A simpler method is to write the representation of interest in terms of Young tableaux and construct the representation as an appropriately symmetrized tensor product of the fundamental representation. One can also use sum rules for various traces of products of generators. The simplest example is the trace over the identity, i.e. the dimension, which must match for the left- and right-handside of (B.19). One may also look up the result in tables, *e.g.* [29, 10], or use computer programs; for some examples see the reference list of [10].

B.6 Subalgebras: $SU(5)$ examples

As an example, we will now discuss the embedding of $SU(3) \times SU(2) \times U(1)$ is embedded in $SU(5)$, which used in Grand Unified Theories. First consider the group embedding. The group $SU(5)$ is defined as the set of 5×5 unitary matrices with determinant one. The subgroups is defined by the subset of matrices of the form

$$U = \begin{pmatrix} U_3 & 0 \\ 0 & U_2 \end{pmatrix} \ , \tag{B.20}$$

where U_3 and U_2 are unitary 3×3 and 2×2 matrices satisfying the relation det U_3det $U_2 = 1$. If we write $U_3 = e^{i\phi}\hat{U}_3$ and $U_2 = e^{i\chi}\hat{U}_2$ where $\hat{U}_3$ and $\hat{U}_2$ have determinant 1, then we have identified the $SU(3)$ and $SU(2)$ subgroups. The phases must satisfy $3\phi + 2\chi = 0 \bmod 2\pi$. This leaves one independent phase, corresponding to the $U(1)$.

In the following we denote $SU(5)$ representations by their dimension in bold face, and the complex conjugate representation of an $SU(5)$ representation by an asterisk. Below we will derive the decomposition of the representations of most interest, namely the **5**, the **10** and the **24**, the adjoint representation.

Decomposition of the Vector Representation. A vector V^A, $A = 1, \ldots, 5$ may be split into three components V^a, $a = 1, 2, 3$ and two components V^i, $i = 4, 5$. Under $SU(5)$ transformations V transforms as

$$V^A \to \sum_{B=1}^{5} U^{AB} V^B = \sum_{a=1}^{3} U^{Aa} V^a + \sum_{i=4}^{5} U^{Ai} V^i \tag{B.21}$$

To determine how this transforms under the $SU(3)$ color group, we take only $U^{ab} \equiv U_3^{ab} \neq 0$. The full matrix U must have determinant one, so we take $U^{ij} \equiv U_2^{ij} = \delta^{ij}$, $U^{ai} = U^{ia} = 0$, and $\det U_3 = 1$. Then we find that

$$V^a \to \sum_{b=1}^{3} U_3^{ab} V^b; \qquad V^i \to V^i \tag{B.22}$$

Hence these components are respectively a vector and a singlet under $SU(3)$. Similarly for the $SU(2)$ part of the subgroup we use $U_3^{ab} = \delta^{ab}$, $\det U_2 = 1$, and $U^{ai} = U^{ia} = 0$. Now the two components are respectively a singlet and a doublet. Finally, the $U(1)$ sub-group acts via the diagonal $SU(5)$ matrix

$$U_Y = \mathrm{diag}(e^{-2\pi i q/3}, e^{-2\pi i q/3}, e^{-2\pi i q/3}, e^{2\pi i q/2}, e^{2\pi i q/2}) \tag{B.23}$$

Now we know how the group $SU(3) \times SU(2) \times U(1)$ acts on the five components of the vector. By expanding these group elements around the identity element we obtain the action of the Lie algebra generators. It follows that the representation **5** decomposes as follows into representations of the $SU(3) \times SU(2) \times U(1)$ subgroup

$$\mathbf{5} \to (3, 1, -\tfrac{1}{3}q) + (1, 2, \tfrac{1}{2}q) \tag{B.24}$$

Here we have allowed for an arbitrary real factor q since the normalization of $U(1)$ charges is not fixed by the algebra. The $SU(3)$ and $SU(2)$ generators can simply be taken as a subset of the $SU(5)$ generators.

Decomposition of the Adjoint Representation To obtain the decomposition of the adjoint representation we may try to write it as a tensor product of vector representations. Although it is possible to obtain the adjoint as a tensor product of five vector representations, this is rather cumbersome. So we use another method: we obtain the adjoint representation as a tensor product of a vector and an anti-vector.

Observe that if ϕ^i and χ^i are $SU(N)$ vectors, then the combination $A^{ij} = \phi^i (\chi^j)^*$ transform as

$$A^{ij} \to U^{ik} (U^{jl})^* A^{kl} = (U A U^\dagger)^{ij} \tag{B.25}$$

which is the transformation rule of the adjoint representation. But the representation obtained this way is reducible. We may write A^{ij} as

$$A^{ij} = (A^{ij} - \frac{1}{N}\,\mathrm{Tr} A\, \delta^{ij}) + \frac{1}{N}\,\mathrm{Tr} A\, \delta^{ij} \tag{B.26}$$

and we see that the two terms are separately invariant. It can be shown that the remaining components are irreducible. This implies the tensor product rule

$$N \times N^* = \text{Adjoint} + 1 \ , \tag{B.27}$$

which agrees with the fact that the dimension of the adjoint representation in $SU(N)$ is N^2-1. Now we can work out the decomposition of the adjoint representation by tensoring the decomposed vectors

$$\left[(3,1,-\tfrac{1}{3}q)+(1,2,\tfrac{1}{2}q)\right] \times \left[(3^*,1,\tfrac{1}{3}q)+(1,2,-\tfrac{1}{2}q)\right] \tag{B.28}$$

We work this out term-by term, using the tensor product rule (B.27) in $SU(3)$ and $SU(2)$. Note that this produces two singlets, but we will have to remove one at the end, to account for the trace in the (B.27). Hence we get

$$\mathbf{24} \rightarrow (8,1,0)+(1,3,0)+(1,1,0)+(3,2,-\tfrac{5}{6}q)+(3^*,2,\tfrac{5}{6}q) \ . \tag{B.29}$$

Decomposition of Rank-2 Anti-Symmetric Tensor To get the contents of the **10** we can take the anti-symmetric tensor product of two **5**'s, decomposed to $SU(3) \times SU(2) \times U(1)$ representations. If ϕ^i, $i = 1, \ldots 5$ is a vector transforming according to the **5** of $SU(5)$, this means that we decompose the tensor product as $\phi^i\tilde{\phi}^j = \frac{1}{2}(\phi^i\tilde{\phi}^j - \phi^j\tilde{\phi}^i) + \frac{1}{2}(\phi^i\tilde{\phi}^j + \phi^j\tilde{\phi}^i)$. It can be shown (for any representation of any algebra) that the algebra transforms the symmetric terms into themselves, and the same with the anti-symmetric ones. Hence they form representations of the algebra. In the case under consideration here (in general for vector representations of $SU(N)$) they form in fact irreducible representations of dimension 10 and 15.

The general rule for anti-symmetric products of a direct sum $R+S$ of two representations is

$$[(R+S)^2]_a = R_a^2 + S_a^2 + RS \ , \tag{B.30}$$

where a denotes the anti-symmetric product. The same relation holds with s instead of a for symmetric products. Using this rule one may immediately derive the result, as we did above for the adjoint. However, in $SU(3)$ and $SU(2)$ the rank-2 anti-symmetric tensors are a bit special.

Let us demonstrate this explicitly by decomposing the representation matrix. Consider an anti-symmetric tensor $T^{AB} = -T^{BA}$. Because of the anti-symmetry, we can impose the condition $A < B$. Under $SU(5)$ the tensor transforms as follows

$$T^{AB} \rightarrow \sum_{C,D=1}^{5} U^{AC}U^{BD}T^{CD} \tag{B.31}$$

Note that the sum is over all C and D, but we can restrict it to the range $C < D$ using the anti-symmetry

$$T^{AB} \rightarrow \sum_{C,D=1;C<D}^{5} (U^{AC}U^{BD} - U^{AD}U^{BC})T^{CD} \tag{B.32}$$

Now we split the indices as before. Then we get

$$
\begin{aligned}
T^{AB} \;\rightarrow\; & \sum_{c,d=1;c<d}^{3} (U^{Ac}U^{Bd} - U^{Ad}U^{Bc})T^{cd} \\
+ \; & \sum_{c=1}^{3}\sum_{i=4}^{5} (U^{Ac}U^{Bi} - U^{Ai}U^{Bc})T^{ci} \\
+ \; & \sum_{i,j=4;i<j}^{5} (U^{Ai}U^{Bj} - U^{Aj}U^{Bi})T^{ij}
\end{aligned}
$$

So there are three components that transform into themselves under $SU(3)\times SU(2)\times U(1)$, and into each other under the full $SU(5)$. These components are T^{ab}, T^{ci} and T^{ij}. So let us see how the subgroup $SU(3) \times SU(2) \times U(1)$ acts on these components. The easiest one is the $U(1)$. The matrices U all take the diagonal form U_Y shown in Eqn (B.23). We see that the components T^{ab} acquire a phase $e^{2\pi i(-2q/3)}$. With the convention $q = 1$ we decided to use above this implies that the T^{ab} has charge $-\frac{2}{3}$. Similarly, T^{ij} has charge $\frac{1}{2} + \frac{1}{2} = 1$, and T^{ai} has charge $-\frac{1}{3} + \frac{1}{2} = \frac{1}{6}$.

Now consider the $SU(2)$ subgroup, choosing again $U_3^{ab} = \delta^{ab}$, det $U_2 = 1$, and $U^{ai} = U^{ia} = 0$. We see that T^{ab} transforms into itself with the matrix $\delta^{ac}\delta^{bd} - \delta^{ad}\delta^{bc} = \delta^{ac}\delta^{bd}$ because the second term vanishes if $a < b$ and $c < d$. The combination T^{aj} transforms to

$$T^{aj} \rightarrow \sum_{c=1}^{3}\sum_{i=4}^{5} (U^{ac}U^{ji} - U^{ai}U^{jc})T^{ci} = \sum_{i=4}^{5} U^{ji}T^{ai} \tag{B.33}$$

This is just a transformation by a unitary 2×2 matrix with determinant 1, *i.e.* the dimension-2 representation of $SU(2)$. Finally, the component T^{kl}, $k, l = 4, 5, k < l$ transforms as

$$T^{kl} \rightarrow \sum_{i,j=4;i<j}^{5} (U^{ki}U^{lj} - U^{kj}U^{li})T^{ij} \tag{B.34}$$

Here the only choice of indices that is possible is $k = 4, l = 5, i = 4, j = 5$, and the factor is $U^{44}U^{55} - U^{45}U^{54} = \det U_2 = 1$, hence T^{kl} is a singlet under $SU(2)$.

Finally consider $SU(3)$, *i.e.* $U^{ij} \equiv U_2^{ij} = \delta^{ij}$, $U^{ai} = U^{ia} = 0$, and det $U_3 = 1$. It is easy to see that T^{aj} transforms with an $SU(3)$ matrix U_3^{ab} and is therefore a vector, and that T^{ij} transforms into itself, and is therefore an $SU(2)$ singlet. For T^{ab} we find

$$T^{ab} \rightarrow \sum_{c,d=1;c<d}^{3} (U^{ac}U^{bd} - U^{ad}U^{bc})T^{cd} \tag{B.35}$$

To see what this implies we rewrite the three tensor components T^{12}, T^{13} and T^{23} in terms of three new variables $S^a = \frac{1}{2}\sum_{b,c} \epsilon^{abc}T^{bc}$. We see then that (from here on all sums are implicit, and unrestricted, so we drop the condition $c < d$, compensating with a factor $\frac{1}{2}$)

$$T^{ab} \rightarrow \tfrac{1}{2}U^{ac}U^{bd}(\delta^{ce}\delta^{df} - \delta^{cf}\delta^{de})T^{ef} = U^{ac}U^{bd}\epsilon^{gcd}(\tfrac{1}{2}\epsilon^{gef}T^{ef}) \tag{B.36}$$

We see now that the transformation of the components S^f is as

$$S^f \to V^{fg} S^g \tag{B.37}$$

with

$$V^{fg} = \epsilon^{fab} U^{ac} U^{bd} \epsilon^{gcd} \tag{B.38}$$

To determine the matrix V^{fg} we multiply it with U^{eg} and sum over g. Then we get

$$V^{fg} U^{eg} = \tfrac{1}{2} \epsilon^{fab} U^{eg} U^{ac} U^{bd} \epsilon^{gcd} = \tfrac{1}{2} \epsilon^{fab} \epsilon^{eab} \det U = \delta^{fe} \det U = \delta^{fe} \ .$$

It follows that $V = U^{T-1} = U^*$. So we see that the components T^{ab}, in a suitable basis, transform as the complex conjugate representation of the $SU(3)$ vector, denoted 3^*. All of this can be summarized as follows

$$\mathbf{10} \to (3^*, 1, -\frac{2}{3}q) + (1, 1, q) + (3, 2, \frac{1}{6}q) \ , \tag{B.39}$$

C Fields and Symmetries

In this appendix we collect results on the various fields we encounter throughout these lecture notes, meanwhile fixing some conventions.

C.1 Scalars

Real massive scalars have a free action

$$\mathcal{L} = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 \phi^2 \tag{C.1}$$

For complex scalars one has

$$\mathcal{L} = \partial_\mu \phi^* \partial^\mu \phi - m^2 \phi \phi^* \tag{C.2}$$

The normalizations are such that the latter expression reduces to two copies of the former if we define $\phi_1 = \frac{1}{\sqrt{2}}(\phi + \phi^*)$ and $\phi_2 = \frac{1}{\sqrt{2}}(\phi - \phi^*)$.

Scalars that transform in a complex or pseudo-real representation of any global or local symmetry must be complex, since the transformations cannot maintain their reality. Scalars in real representations may be real or complex, but in the latter case one may always decompose them into two real scalars.

C.2 Fermions

The Lagrangian for a massive fermion is

$$\mathcal{L} = i\bar{\psi} \gamma^\mu \partial_\mu \psi - m \bar{\psi} \psi \tag{C.3}$$

The γ matrices are 4×4-matrices defined by

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} \tag{C.4}$$

We can (and will) choose them in such a way that γ^0 is Hermitean and the other three are anti-Hermitean. The conjugate spinor $\bar{\psi}$ is defined as $\bar{\psi} = \psi^\dagger \gamma^0$. The matrix γ_5 is defined by

$$\gamma_5 \equiv \gamma^5 = i\gamma_0\gamma_1\gamma_2\gamma_3 \ , \tag{C.5}$$

and is Hermitean. In any representation of the γ-matrices there is a unitary matrix C so that

$$\gamma_\mu^T = -C\gamma_\mu C^{-1} \tag{C.6}$$

Left- or right-handed Weyl spinor are defined by means of the projection operators

$$P_L = \frac{1}{2}(1+\gamma_5); \quad P_R = \frac{1}{2}(1-\gamma_5) \ . \tag{C.7}$$

They satisfy $P_R P_R = P_R; P_L P_L = P_L$ and $P_L P_R = 0$. The left and right-handed components of a field are defined as $\psi_L = P_L\psi; \quad \psi_R = P_R\psi$. Due to the projections ψ_L and ψ_R are effectively two-component spinors, called *Weyl spinors*. A Dirac spinor has four complex degrees of freedom, a Weyl-spinor only two.

C.2.1 Chirality and Helicity: Conventions

In the literature one finds many different sign conventions for fermions, such as the metric, the definition of γ_5, the definition of P_L and the definition of the ϵ-tensor. Not all combinations of choices are allowed, though. Left-handed particles have their spin oriented opposite to their momentum. This is a convention-independent statement.

Classically angular momentum is defined as $\vec{L} = \vec{r}\times\vec{p}$ (*i.e.* $L^i = \epsilon^{ijk} r^j p^k$, with $\epsilon^{123} = 1$), which in quantum mechanics via the Heisenberg relation $[r^i, p^j] = i\delta^{ij}$ (which in principle involves another sign convention, though here everyone agrees; furthermore we use $\hbar = 1$.) leads to $[L^i, L^j] = i\epsilon_{ijk}L^k$. Spin operators must satisfy the same relation, and this leaves no room for sign ambiguities.

In a relativistic theory angular momentum is contained in the tensor $M^{\mu\nu} = x^\mu p^\nu - x^\nu p^\mu$. This tensor satisfies the commutation relation (with $[x^\mu, p^\nu] = -ig^{\mu\nu}$)

$$[M^{\mu\nu}, M^{\rho\sigma}] = -i(g^{\mu\rho}M^{\nu\sigma} - g^{\mu\sigma}M^{\nu\rho} + g^{\nu\sigma}M^{\mu\rho} - g^{\nu\rho}M^{\mu\sigma}) \ . \tag{C.8}$$

The operator that represents internal angular momentum on fermions must have the same commutation relations, and this identifies it uniquely as

$$\Sigma^{\mu\nu} = \tfrac{i}{4}\left[\gamma^\mu, \gamma^\nu\right] \ , \tag{C.9}$$

so that the spin operator must be

$$S^i = \tfrac{i}{4}\epsilon^{ijk}\gamma^j\gamma^k \tag{C.10}$$

To make sure that the signs are correct one can check $[S^i, S^j] = i\epsilon_{ijk}S^k$. [This is valid in the $(+---)$ metric. In the $(-+++)$ metric we have $[x^\mu, p^\nu] = +ig^{\mu\nu}$, and hence the left-hand side of Eq. (C.8) changes sign. However in *both* metrics one usually chooses $\{\gamma^\mu, \gamma^\nu\} = +2g^{\mu\nu}$. Therefore the definitions of Σ and S change sign.]

The *helicity* of a fermion is the eigenvalue of the operator $\frac{\vec{S}\cdot\vec{p}}{|\vec{p}|}$; the *chirality* is the eigenvalue of γ_5. To relate the two (for massless particles) we use the Dirac equation $p^\mu\gamma_\mu\psi = 0$. Using $p^\mu = (E, \vec{p})$ and choosing $\vec{p}$ along the x-axis, $\vec{p} = (p, 0, 0)$ we get $(E\gamma_0 + p\gamma_1)\psi = 0$, or, after raising some indices, $p\gamma^1\psi = E\gamma^0\psi$. Now consider the action of the helicity operator: $\vec{S}\cdot\vec{p}\psi = \frac{i}{2}p\gamma^2\gamma^3\psi$. A fermion with chirality + satisfies

$$\gamma_5\psi = \psi \;\rightarrow\; i\gamma_0\gamma_1\gamma_2\gamma_3\psi = \psi \;\rightarrow\; \gamma^2\gamma^3\psi = i\gamma^0\gamma^1\psi \tag{C.11}$$

Hence $\vec{S}\cdot\vec{p}\psi = -\frac{1}{2}p\gamma^0\gamma^1\psi$, which with the help of the Dirac equation becomes

$$-\frac{1}{2}p\gamma^0\gamma^1\psi = -\tfrac{1}{2}E(\gamma^0)^2\psi = -\tfrac{1}{2}|p|\psi\ . \tag{C.12}$$

This means that this fermion has its helicity opposite to its momentum and is, by definition, left-handed. We see that with our definition of γ_5 this corresponds to positive chirality, or $P_L = \frac{1}{2}(1+\gamma_5)$.

C.2.2 Majorana Fermions

For an arbitrary Dirac spinor one defines

$$\bar{\psi} = -(\psi^c)^T C\ , \tag{C.13}$$

where ψ^c is the charge conjugate spinor. A *Majorana spinor* is defined by $\psi = \pm\psi^c$. Just like the Weyl condition this reduces the number of degrees of freedom by a factor of two. However, one can impose either a Weyl condition, or a Majorana condition, but not both. The standard form of the action for a Majorana spinor is

$$\mathcal{L} = i\tfrac{1}{2}\bar{\psi}\gamma^\mu\partial_\mu\psi - \tfrac{1}{2}m\bar{\psi}\psi \tag{C.14}$$

This form obscures the fact that $\bar{\psi}$ and ψ are not to be treated as independent variables, as they are for complex fermions. Therefore it is better to express $\bar{\psi}$ in terms of ψ using the Majorana condition $\bar{\psi} = \mp(\psi^c)^T C$. Then we get

$$\mathcal{L} = \pm\tfrac{1}{2}(-i\psi C\gamma^\mu\partial_\mu\psi + m\psi C\psi) \tag{C.15}$$

In the absence of a mass term there is not really any difference between Majorana and Weyl spinors. We may write

$$i\tfrac{1}{2}\bar{\psi}\gamma^\mu\partial_\mu\psi \;= i\tfrac{1}{2}(\bar{\psi}_L\gamma^\mu\partial_\mu\psi_L \;+ \bar{\psi}_R\gamma^\mu\partial_\mu\psi_R)\ \ , \tag{C.16}$$

and then substitute $\bar{\psi}_R = -((\psi^c)_L)^T$ in the second term. Using the Majorana property plus a little algebra one finds that the second term is now transformed into the first one. Hence for Majorana fermions

$$i\tfrac{1}{2}\bar{\psi}\gamma^\mu\partial_\mu\psi \;= i\bar{\psi}_L\gamma^\mu\partial_\mu\psi_L \;= i\bar{\psi}_R\gamma^\mu\partial_\mu\psi_R\ \ . \tag{C.17}$$

The difference between Majorana and Weyl fermions becomes essential if one assigns them to representations of local or global symmetries, and writes down mass terms. For Majorana fermions the representations must be real, and masses are allowed, while for Weyl fermions the representation can be complex, but then one cannot write down an invariant mass term.

C.3 Gauge Bosons

Minimal couplings of gauge bosons are obtained by replacing ∂_μ by the covariant derivative $D_\mu = \partial_\mu - igA^a_\mu T^a$, where T^a is a generator of the Lie algebra of the gauge group in the representation of the field that D_μ is acting on. The generators are hermitean and satisfy

$$[T^a, T^b] = if^{abc}T^c \tag{C.18}$$

The action for gauge bosons is

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu,a} \ , \tag{C.19}$$

where the definition of the field strength F is determined by our definition of the covariant derivative, *i.e.*

$$-igF^a_{\mu\nu}T^a = [D_\mu, D_\nu] \ , \tag{C.20}$$

on any field. Then, in components,

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu \ . \tag{C.21}$$

Often we use Lie-algebra-valued fields instead of components. They are defined as follows

$$A_\mu \equiv -igA^a_\mu T^a; \quad F_{\mu\nu} \equiv -igF^a_{\mu\nu}T^a \ , \tag{C.22}$$

so that the covariant derivative takes the form

$$D_\mu = \partial_\mu + A_\mu \tag{C.23}$$

The generators are in the representation of the fields on which they act.

C.4 Space Inversion

Space inversion (parity) changes $x^\mu = (x^0, x^i)$ to $x^{\mu,P} = (x^0, -x^i)$. In field theory Lagrangians x^i is not a dynamical variable, but plays the rôle of a labeling of the degrees of freedom. Hence the space inversion transformation is an exchange of dynamical variables. This should be compared to a Lagrangian $L(q_1, q_2)$ which may be invariant under the interchange of q_1 and q_2. If that is the case, the also the equations of motion of q_1 and q_1 are transformed into each other. Analogously, a Lagrangian $L(\phi(\vec{x}))$ may be invariant under the replacement of $\phi(\vec{x})$ by $\phi(-\vec{x})$. In that case the solutions of the equations of motion are invariant under space inversion, or, in other words, one cannot distinguish the time evolution of fields from its mirror image*

Note that one should consider the Lagrangian (or the Hamiltonian), and not the Lagrangian density. The latter is in general not invariant, but may change from $\mathcal{L}(\vec{x})$ to $\mathcal{L}(-\vec{x})$ (this denotes the full $\vec{x}$-dependence, explicit or implicit via the fields). If this the only change, the space-integral of $\mathcal{L}(\vec{x})$ is of course invariant, and so is the Hamiltonian.

Classically the transformation we need to consider is therefore a replacement of all fields by their "mirror image"

$$\mathcal{L}(\phi(\vec{x})) \rightarrow \mathcal{L}(\phi(-\vec{x})) \ . \tag{C.24}$$

This is true in the simplest case, but the transformation may be more complicated. For fields with several components due to spin (or perhaps even external degrees of freedom) one may allow in addition to this also a transformation of these spin components, dictated by the requirement of invariance. So in this more general situation we can consider

$$\mathcal{L}(\phi^i(\vec{x})) \rightarrow \mathcal{L}(P^{ij}\phi^j(-\vec{x})) \tag{C.25}$$

where P^{ij} is some matrix. Obviously the square of P should be 1, since two space inversions equal the identity (for fermion fields P^2 may in fact be -1) Even if there is just one component P can be non-trivial, namely a sign, the *intrinsic parity* of the field.

Note that only the fields are transformed. In fact, in a local field theory there is no explicit dependence on x, so there is nothing else to transform. However, let us, for the sake of the argument, consider for a moment an example where there *is* explicit dependence on x: $\mathcal{L}(\phi(x), x) = \partial_\mu\phi(x)\partial^\mu\phi(x) + n^\mu x_\mu \phi^2(x)$. This theory is not parity invariant. For example, if $\phi_0(x)$ satisfies the equations of motion, then $\phi_0(-x)$ does not. Suppose, however, we take for parity transformation:

$$\mathcal{L}(\phi(x), x) \rightarrow \mathcal{L}(\phi(-x), -x) \tag{C.26}$$

Then the sign would disappear when we integrate L to get the action, and perform a change of integration variables $x \rightarrow -x$. This would lead to the incorrect conclusion that this theory is parity invariant. In other words: replacing x by $-x$ is only a field relabeling, and not a parity transformation.

Although explicit x-dependence of $\mathcal{L}$ never occurs, derivatives do appear, and their transformation may be a source of confusion. Again we need to keep in mind that we are interchanging dynamical variables labeled by $\vec{x}$. In one dimension, consider $\partial_x\phi(x)$. Does this transform to $\partial_x\phi(-x)$ or $-\partial_x\phi(-x)$? To get the correct answer consider the infinitesimal form of the derivative

$$\lim_{\delta\rightarrow 0} \frac{\phi(x+\delta) - \phi(x)}{\delta} \tag{C.27}$$

* This is slightly misleading language since a mirror reflects only one axis; an inversion is a reflection combined with a rotation, which we assume to be a symmetry anyway.

The symmetry transformation changes $\phi(y)$ to $\phi(-y)$ for all y, and hence the derivative changes to

$$\lim_{\delta\to 0}\frac{\phi(-x-\delta)-\phi(-x)}{\delta}=\partial_x\phi(-x)\ . \tag{C.28}$$

This is in fact just another manifestation of the fact that we transform the fields, and not $\vec{x}$.

In quantum field theory parity is represented by an operator $\mathcal{P}$ which acts on the Fock space (the multi-particle Hilbert space). If parity is a symmetry, then $\mathcal{P}$ commutes with the Hamiltonian, and the the time evolution of states $|b\rangle$ and their "mirror images" $\mathcal{P}\,|b\rangle$ are related:

$$\langle a|\,\mathcal{P}^\dagger e^{-iHt}\mathcal{P}\,|b\rangle=\langle a|\,e^{-iHt}\,|b\rangle\ . \tag{C.29}$$

The Hamiltonian is directly related to the Lagrange density, and since we usually work with the latter, we wish to check the invariance for $\mathcal{L}$ rather than H.

Thus we should consider $\mathcal{P}^\dagger\mathcal{L}\mathcal{P}$. The operators act only on the factors in the Lagrangian density that are themselves operators, *i.e.* the fields, and not on coupling constants, derivatives, group generators, gamma matrices or whatever else might appear in a Lagrangian. The parity operators change every field Φ according to the rule $\Phi\to\mathcal{P}^\dagger\Phi\mathcal{P}$. The result of this operation must correspond to the classical transformation, *i.e.*

$$\mathcal{P}^\dagger\Phi^i(x)\mathcal{P}=P^{ij}\Phi^i(x^P) \tag{C.30}$$

For scalar fields the 1×1 matrix P is either 1 or -1. In the latter case they are called *pseudo-scalars*. This is manifestly a symmetry of the scalar action (C.1). Note that it is necessary (and allowed) to replace $\partial_\mu=\frac{\partial}{\partial x_\mu}$ by $\frac{\partial}{\partial x^P_\mu}\equiv\partial^P_\mu$, since this derivative appears only contracted with another derivative.

The parity transformation for fermions involves a non-trivial matrix P. The action of a Dirac fermion is transformed to

$$\mathcal{P}^\dagger\mathcal{L}\mathcal{P}=i\bar\psi(x^P)\gamma^0P^\dagger\gamma^0\gamma^\mu\partial_\mu P\psi(x^P)-m\bar\psi(x^P)\gamma^0P^\dagger\gamma^0P\psi(x^P) \tag{C.31}$$

We want to change ∂_μ to ∂^P_μ and get rid of the matrices P. For the kinetic terms we get the requirement

$$P^\dagger\gamma^0\gamma^\mu P=\gamma^0\gamma^{\mu,P}\ , \tag{C.32}$$

and for the mass terms

$$P^\dagger\gamma^0P=\gamma^0\ . \tag{C.33}$$

The first condition for $\mu=0$ requires P to be unitary. Then the second condition, substituted into the first one yields

$$P^\dagger\gamma^\mu P=\gamma^{\mu,P}\ , \tag{C.34}$$

a *i.e.* all three space components of γ^μ should change sign. The matrix that achieves this is unique*up to a phase: $P=i\gamma^0$ (the factor i is not essential here, see however [5] for a justification). Hence

$$\mathcal{P}\psi(x)\mathcal{P}^{-1}=i\gamma^0\psi(x^P)\ , \tag{C.35}$$

* To prove uniqueness one can use the fact that the only unitary matrix that commutes with all γ^μ is the unit matrix times a phase.

where $\mathcal{P}$ is the operator acting on the Hilbert space. From this we can derive the action on chiral fermions,

$$\mathcal{P}\psi_{L,R}(x)\mathcal{P}^{-1} = i\gamma^0\psi_{R,L}(x^P) \ . \tag{C.36}$$

Obviously any matrix P that changes the sign of three of the four Dirac matrices must also change the sign of γ^5.

To summarize, the rôle of the parity transformation matrix $i\gamma^0$ is to ensure that the vector $i\bar{\psi}\gamma_\mu\psi$ transforms under parity like a vector. Parity reversal is not a symmetry of the kinetic terms of the left-handed fields alone.

For couplings of gauge fields not to destroy parity (if it is a symmetry without the coupling to gauge fields), the transformation of ∂_μ must be the same as that of D_μ. Hence the space components of A_μ must change sign, while the time component does not; in other words, A_μ must transform like a vector. A quantity transforming with an extra $-$ sign is called a pseudo-vector.

The fermion bi-linears transform as follows: $\bar{\psi}\psi$ is a scalar, $\bar{\psi}\gamma_5\psi$ a pseudo-scalar, $\bar{\psi}\gamma_\mu\psi$ a vector and $\bar{\psi}\gamma_\mu\gamma_5\psi$ a pseudo-vector.

C.5 Charge Conjugation

Lagrangians must always be real, and hence for every complex field there exists a complex conjugate field. The latter transforms in the complex conjugate representation with respect to any gauge group. In particular with respect to $U(1)$'s complex conjugate fields have opposite charge.

Charge conjugation is a symmetry that involves the interchange of fields with their complex conjugates. Just like space inversion it is an interchange of dynamical variables, which in quantum field theory is represented by a unitary operator sending every field to $\mathcal{C}\Phi\mathcal{C}^{-1}$. In addition to complex conjugating the field it may, just as parity reversal, act non-trivially on the indices of the field in order to get proper transformation rules. Hence the generic transformation is

$$\mathcal{C}^\dagger\Phi^i\mathcal{C} = C^{ij}(\Phi^j)^* \tag{C.37}$$

Since charge conjugation does not act on the argument of Φ, we will omit it. The rôle of the matrix C is to ensure that we transform the components of Φ to the proper basis of the complex conjugate representation. It may act on external as well as spin indices. For bosonic fields there is no need for any action on spin indices, since all integer spin representations are real. But this is not true for half-integer spin fields.

A complex scalar ϕ_i is transformed to

$$\phi_i \to \mathcal{C}^\dagger\phi_i\mathcal{C} = C_R^{ij}\phi_j^* \ , \tag{C.38}$$

where C_R is a unitary matrix that depends on the representation R of the scalar with respect to the gauge group under consideration. Conjugating twice we find

$$\phi_i = C_R^{ij}(C_R^{jk}\phi_k^*)^* = (C_R C_R^*)_{ij}\phi_j \tag{C.39}$$

so that $C_R C_R^* = 1$, or $C_R = C_R^T$. From the kinetic terms alone we get no further constraints on this matrix, but if we consider coupling to gauge bosons the issue of invariance becomes non-trivial.

Derivatives $\partial_\mu \phi$ transform exactly like ϕ, and hence covariant derivatives must also transform like ϕ. We have

$$D_\mu \phi \to \mathcal{C}^\dagger D_\mu \phi_i \mathcal{C} = \mathcal{C}^\dagger D_\mu \mathcal{C} C_R \phi^* \ . \tag{C.40}$$

Clearly we need now, in this representation

$$\mathcal{C}^\dagger D_\mu \mathcal{C} = C_R D_\mu^* C_R^\dagger \tag{C.41}$$

The ordinary derivative part of D_μ transforms trivially, but the operator $\mathcal{C}$ does act non-trivially on the field A_μ. From Eq. (C.41) we get

$$\mathcal{C}^\dagger A_\mu^a \mathcal{C} T_R^a = -A_\mu^a C_R (T_R^a)^* C_R^\dagger \ , \tag{C.42}$$

where T_R^a denotes a generator in the representation R. Since A_μ^a is a real field, it transforms with some real matrix C_A (where "A" stands for "adjoint"), as follows

$$\mathcal{C}^\dagger A_\mu^a \mathcal{C} = C_A^{ab} A_\mu^b \tag{C.43}$$

Hence we get the following requirement on the matrices C_R and C_A

$$T_R^a C_A^{ab} = -C_R (T_R^b)^* C_R^\dagger \ . \tag{C.44}$$

We can have scalars in many different representations, but the transformation C_A acts only on one set of fields A_μ, and hence C_A must be independent of R. Such a set of matrices always exists, as we will show below explicitly for $SU(N)$ (charge conjugation is in fact nothing but a "space-inversion" in the root space of the Lie algebra under consideration). In the special case of a Hermitean $U(1)$ generator $Q = Q^*$ the equation reads $\mathcal{C}^\dagger A_\mu \mathcal{C} = -A_\mu$, which implies that a vector boson has charge parity ("C-parity") -1.

The matrices C_R are not unique, but change under a basis transformation of the representation R. Suppose $\tilde{T}_R = S_R T_R S_R^\dagger$, for some unitary S_R. Here T_R is a generator in the representation R, and obviously so is $\tilde{T}_R$. Then (in this derivation we omit the subscript R on C, S and T)

$$\begin{aligned}
\tilde{T}^a C_A^{ab} &= S T^a S^\dagger C_A^{ab} = -SC(T^b)^* C^\dagger S^\dagger \\
&= -SCS^T S^* (T^b)^* S^T S^* C^\dagger S^\dagger = -SCS^T (\tilde{T}^a)^* S^* C^\dagger S^\dagger
\end{aligned}$$

from which we read off

$$\tilde{C}_R = S_R C_R S_R^T \tag{C.45}$$

For a complex, one-component charged scalar the condition $C_R C_R^* = 1$ (from Eq. (C.39)) still allows arbitrary phases, but then we may use Eq. (C.45) to set C_R equal to any desired phase, *e.g.* $C_R = 1$. In other words, C-parity is not defined for charged fields. For a real, one-component scalar C_R must be real, and then $C_R C_R = 1$ has two solutions, namely $C_R = \pm$. The basis transformation S_R must be real as well for real fields, so that we cannot change C_R.

Similar remarks apply to scalars in representations of non-abelian groups. Since C_R is a unitary, symmetric matrix, we can define the square root of C_R, which is also a unitary symmetric matrix. If we now set $S = C_R^{-1/2}$ we find that $\tilde{C}_R = 1$. Hence for complex fields we can always choose $C_R = 1$, provided we make a suitable choice of the generators. This may not be possible for real fields: note that S is in general a complex matrix, even if C_R is real. Then transforming by S may make the generators iT^a complex, whereas for real fields we need them to be real.

For example for $SU(N)$ all representations can be obtained as tensor products of the fundamental representation, with real projections. One usually chooses a basis so that the $N-1$ Cartan sub-algebra generators are real (and diagonal), while half of the remaining $N^2 - N$ generators are real, and the other half purely imaginary. If this choice is made in the fundamental representation, the reality properties are the same for all representations (in other words the same generators T_R^a are always either real or imaginary, if they generators are constructed using the tensor method). Then the matrix C_A^{ab} is diagonal and equal to -1 for the real generators, and 1 for the imaginary ones, so that in any representation R

$$T_R^a C_A^{ab} = -(T_R^b)^* \ . \tag{C.46}$$

Hence we have explicitly satisfied (C.44) with $C_R = 1$. Note however that the tensor procedure does not produce a real basis for real representations. Indeed, (C.46) is obviously not valid for a real basis* if C_A is non-trivial. If we transform T_R^a to a real basis we get in terms of the real generators the transformation (C.44), with a non-trivial matrix $\tilde{C}_R = SS^T$ generated by transforming to a real basis using (C.45). This shows how (C.44) can indeed be satisfied for any $SU(N)$ representation.

For fermions the action of charge conjugation is slightly more complicated because of the fact that they are in a spinor representation of the space-time symmetry group. For Weyl spinors these representations are complex in four dimensions, and hence transform into an inequivalent Weyl spinor, describing a particle with opposite helicity. Dirac spinors transform into themselves since they contain two Weyl spinors of opposite parity, but there is still a non-trivial matrix in the transformation. We will consider first fermions that are in a trivial representation of any gauge group. Hence the matrix C^{ij} only acts on spin indices.

The transformation rule is

$$\mathcal{C}^\dagger \psi \mathcal{C} = C_F \psi^* \ , \tag{C.47}$$

where C_F is some matrix to be determined. Consider now some current $\bar{\psi}\Gamma\psi$, where Γ is some product of γ matrices. this current transforms to

$$\mathcal{C}^\dagger \bar{\psi}\Gamma\psi\mathcal{C} \ = \ \psi^T (C_F)^\dagger (\gamma^0) \Gamma C_F \psi^*$$

* Note that in a real basis iT^a is real, T^a purely imaginary.

$$
\begin{aligned}
&= -\psi^\dagger (C_F)^T \Gamma^T (\gamma^0)^T (C_F)^* \psi \\
&= -\bar{\psi} \gamma^0 (C_F)^T \Gamma^T (\gamma^0)^T (C_F)^* \psi
\end{aligned}
$$

In the second step we took the transpose of the entire quantity (which is a number) and introduced a $-$ sign because in the process we anti-commute two fermions*

There are two cases of interest, namely $\Gamma = \gamma^\mu$, and $\Gamma = 1$, corresponding respectively to the kinetic terms and the mass terms. Invariance of the kinetic terms leads to the requirement

$$\gamma^0 (C_F)^T (\gamma^\mu)^T (\gamma^0)^T (C_F)^* = \gamma^\mu \tag{C.48}$$

and invariance of the mass term to

$$\gamma^0 (C_F)^T (\gamma^0)^T (C_F)^* = -1 \tag{C.49}$$

[Note that in the first condition there is an extra $-$ sign due to the fact that the derivative has to be partially integrated so that it acts on ψ rather than ψ^*.]

As in the discussion of parity the $\mu = 0$ component of the first condition implies that C_F must be unitary. It is not hard to show that the unique solution (up to a phase) for C_F is then

$$C_F = C^{-1} (\gamma^0)^T \ , \tag{C.50}$$

where C is the matrix introduced earlier in this appendix, satisfying

$$C \gamma_\mu C^{-1} = -\gamma_\mu^T \ , \tag{C.51}$$

with $C^{-1} = C^\dagger$. and $C = -C^T$. The precise form of the matrix depends on the representation one chooses for the γ matrices, but it can be shown that in four dimensions it always satisfies $C = -C^T$.

Let us now consider other choices for Γ. We define

$$C \Gamma C^{-1} = \eta_\Gamma \Gamma^T \ , \tag{C.52}$$

and then we get

$$\mathcal{C}^\dagger \bar{\psi} \Gamma \psi \mathcal{C} = \eta_\Gamma \bar{\psi} \Gamma \psi \ . \tag{C.53}$$

We know already that $\eta_{\gamma^\mu} = -1$. The matrix γ_5 defined in (C.5) is transformed in the following way by the matrix C

$$C \gamma^5 C^{-1} = i \gamma_0^T \gamma_1^T \gamma_2^T \gamma_3^T \ , \tag{C.54}$$

which equals γ_5^T after one re-orders the four factors (which does not produce a sign flip). Hence $\eta_{\gamma_5} = 1$. Then $\eta_{\gamma^\mu \gamma_5} = 1$ as well. Hence the axial vector $\bar{\psi} \gamma_\mu \gamma_5 \psi$ transforms into

* The attentive reader may be confused by the apparent contradiction between the following operations: when proving Hermiticity of the Lagrangian we have $(\psi^\dagger \gamma^0 \psi)^\dagger = \psi^\dagger (\gamma^0)^\dagger \psi$ (no sign change), whereas here we have $(\psi^T (\gamma^0)^T \psi^*)^T = -\psi^\dagger \gamma^0 \psi$ (sign change). This difference is due to the fact that the $\dagger$ in the first expression is a hermitean conjugate in Hilbert space, which takes $(AB)^\dagger$ to $B^\dagger A^\dagger$ irrespective of there commutation relations, whereas the T in the second expression acts only on the spinor labels.

itself, and hence the vector current and the axial vector current transform with opposite sign under charge conjugation. Therefore the chiral action $i\bar{\psi}_L\gamma^\mu\partial_\mu\psi_L$ is not invariant, but transforms into the same expression with "L" replaced by "R".

Apart from this helicity flip, the coupling of fermions to gauge bosons is invariant under $\mathcal{C}$. The gauge current $\bar{\psi}\gamma_\mu T^a\psi_j$ transforms to $-\bar{\psi}\gamma_\mu(T^a)^*\psi$, where we used the fact that the generators are hermitean. Hence the action for a minimally coupled Dirac fermion transforms to

$$i\mathcal{C}^\dagger\bar{\psi}\gamma_\mu D_\mu\psi\mathcal{C} = i\bar{\psi}\gamma_\mu(\partial_\mu + ig(T^a)^*\mathcal{C}^\dagger A^a_\mu\mathcal{C})\psi \ . \tag{C.55}$$

The rest of the discussion is identical to the earlier discussion of the transformation of covariant derivatives for scalar fields: in addition to the matrix C_F acting on the spinor indices, we need a second, representation dependent matrix C_R, satisfying (C.44). If we define

$$\mathcal{C}^\dagger\psi\mathcal{C} = C_F C_R\psi^* \ , \tag{C.56}$$

then the transformed interaction term becomes

$$i\mathcal{C}^\dagger\bar{\psi}\gamma_\mu D_\mu\psi\mathcal{C} = i\bar{\psi}(C_R)^T\gamma_\mu(\partial_\mu + ig(T^a)^*C^{ab}A^b_\mu)(C_R)^*\psi \ . \tag{C.57}$$

Now we complex conjugate (C.44), and make use of the fact that C_A is real (since it transforms the real fields A_μ). Hence

$$(T^a)^*C^{ab} = -C_R^*T^bC_R^T \tag{C.58}$$

If we substitute this we get the original action back.

C.6 Time Reversal

Time reversal in quantum field theory requires anti-unitary operators, just like in quantum mechanics. In general we have

$$\mathcal{T}^\dagger\Phi^i(x)\mathcal{T} = T^{ij}\Phi^j(x^T) \ , \tag{C.59}$$

where $\mathcal{T}$ is an anti-unitary operator and T a matrix in the internal (spin) and/or external space of degrees of freedom of the field. Here x^T is the four-vector $-x^0, x^i$.

The theory is invariant under time reversal if (and only if)

$$\mathcal{T}^\dagger\mathcal{L}(x^0, x^i)\mathcal{T} = \mathcal{L}(-x^0, x^i) \tag{C.60}$$

Then for the Hamiltonian we have

$$\mathcal{T}^\dagger H(t)\mathcal{T} = H(-t) \tag{C.61}$$

and for a time evolution of time-reversed states we find the following. If we start with an amplitude

$$A(t) = \langle a| e^{-iH(t)t} |b\rangle \ , \tag{C.62}$$

then for the time reversed states the time evolution amplitude is

$$\langle a| \mathcal{T}^\dagger e^{-iH(t)t} \mathcal{T} |b\rangle = \langle a| e^{iH(-t)t} |b\rangle = A(-t) \tag{C.63}$$

From the Hamiltonian point of view space and time reversal are not treated symmetrically, since time plays a special rôle. However, from the point of view of the action the concept of symmetry is the same in both cases: the action is invariant if $\mathcal{L}$ satisfies Eq. (C.60).

Because $\mathcal{T}$ is *anti*-unitary, it is not true anymore that only the fields are transformed. All other objects in the Lagrangian density are also changed, namely complex conjugated.

For free Dirac fermions we get

$$\mathcal{T}^\dagger \psi(x) \mathcal{T} = T_F \psi(x^T) \ , \tag{C.64}$$

where T_F is a matrix in spinor space. The Dirac Lagrangian transforms as follows

$$\begin{aligned}\mathcal{T}^\dagger[-i\bar{\psi}\gamma^\mu\partial_\mu\psi(x) - m\bar{\psi}\psi(x)]\mathcal{T} &= -i\psi^\dagger(x^T)T^\dagger(\gamma^0)^*(\gamma^\mu)^*\partial_\mu T\psi(x^T) \\ &\quad -m\psi^\dagger(x^T)T^\dagger(\gamma^0)^*T\psi(x^T) \ .\end{aligned}$$

Again we consider first the time component of the first term. This term should change sign, which it does precisely if T is unitary. For the space components and the mass we find then

$$-iT^\dagger(\gamma^0)^*(\gamma^i)^*T = i\gamma^0(\gamma^i) \tag{C.65}$$

and

$$T^\dagger(\gamma^0)^*T = \gamma^0 \tag{C.66}$$

Substituting the last equation into the first gives

$$-iT^\dagger(\gamma^i)^*T = i(\gamma^i) \tag{C.67}$$

The solution is, up to a phase

$$T = C\gamma^5 \ . \tag{C.68}$$

Note that unlike parity and charge conjugation time reversal is a symmetry for Weyl fermions. The physical reason is that time reversal changes the direction of both spin and momentum and hence the helicity is conserved. Parity flips the momentum, but not the spin (spin transforms like orbital angular momentum, $\vec{r} \times \vec{p}$), whereas for half-integer spin particles charge conjugation flips spin but not momentum (technically this happens because spinors are in complex representations of $SO(3,1)$).

D Supersymmetry

D.1 Notation

We will use here the (dotted) index notation for spinors introduced in appendix A. Implicit contraction for indices are as follows

$$\chi\psi \equiv \chi^\alpha\psi_\alpha \ ; \qquad \bar{\chi}\bar{\psi} \equiv \bar{\chi}_{\dot{\alpha}}\bar{\psi}^{\dot{\alpha}} \tag{D.1}$$

The following relations hold for anti-commuting spinors

$$\chi\psi = \psi\chi \;; \qquad \bar{\chi}\bar{\psi} = \bar{\psi}\bar{\chi} \tag{D.2}$$

$$\psi\sigma^\mu\bar{\chi} = -\bar{\chi}\bar{\sigma}^\mu\psi \tag{D.3}$$

$$\chi^\alpha\psi_\alpha = -\chi_\alpha\psi^\alpha \tag{D.4}$$

$$(\psi\chi)^\dagger = \bar{\chi}\bar{\psi} = \bar{\psi}\bar{\chi} \tag{D.5}$$

$$(\psi\sigma^\mu\bar{\chi})^\dagger = \chi\sigma^\mu\bar{\psi}\ . \tag{D.6}$$

To derive the last two, note that for operators $(AB)^\dagger = B^\dagger A^\dagger$. Hence $(\psi\chi)^\dagger = (\psi_\alpha)^\dagger(\chi^\alpha)^\dagger = (\bar{\psi}_{\dot{\alpha}})(\bar{\chi}^{\dot{\alpha}}) = \bar{\psi}\bar{\chi}$. For the last one, $(\psi\sigma^\mu\bar{\chi})^\dagger = \chi^\beta(\sigma^\mu_{\alpha\dot{\beta}})^*\bar{\psi}^{\dot{\alpha}}$; then use Hermiticity of σ^μ and the fact that, numerically, $\beta = \dot{\beta}$ and $\alpha = \dot{\alpha}$. We will use a metric-independent notation, *i.e.* we will write all formulas in such a way that they are correct for two metrics, $(+---)$ as well as $(-+++)$. This is done by explicit factors η^{00} and σ^0.

D.2 The Wess-Zumino Model

The Wess-Zumino is the simplest example of a four-dimensional supersymmetric field theory. There are just two free fields, a complex boson and a Weyl fermion. The Lagrangian is

$$\mathcal{L} = \mathcal{L}_{\text{boson}} + \mathcal{L}_{\text{fermion}} \tag{D.7}$$

where

$$\mathcal{L}_{\text{boson}} = \eta^{00}\eta^{\mu\nu}\partial_\mu\phi^\dagger\partial_\nu\phi \tag{D.8}$$

The fermion Lagrangian has the form (see Eq. (A.22))

$$\mathcal{L}_{\text{fermion}} = i\sigma^0\bar{\psi}\bar{\sigma}^\mu\partial_\mu\psi \tag{D.9}$$

This Lagrangian is supposed to have a symmetry transforming bosons into fermions and vice-versa. The transformation of the scalars is as follows

$$\delta_\varepsilon\phi = \sqrt{2}\varepsilon\psi \equiv \sqrt{2}\varepsilon^\alpha\psi_\alpha \tag{D.10}$$

Here ε is a spinor, which is assumed to anti-commute with all other spinors in the problem. The factor $\sqrt{2}$ is the standard convention used in the literature. The conjugate of the scalar transforms as

$$\delta_\varepsilon\phi^\dagger = \sqrt{2}(\varepsilon\psi)^\dagger = \sqrt{2}\bar{\varepsilon}\bar{\psi} \tag{D.11}$$

As a result of this transformation, the scalar Lagrangian transforms as follows

$$\delta_\varepsilon\mathcal{L}_{\text{boson}} = \sqrt{2}\eta^{00}\left(\varepsilon\partial^\mu\psi\partial_\mu\phi^\dagger + \bar{\varepsilon}\partial^\mu\bar{\psi}\partial_\mu\phi\right) \tag{D.12}$$

These terms have to be canceled by the variation of the fermionic terms. An educated guess for the fermion transformation is (λ is a real parameter to be determined later)

$$\delta_\varepsilon\psi_\alpha = i\lambda(\sigma^\mu\bar{\varepsilon})_\alpha\partial_\mu\phi = i\lambda\sigma^\mu_{\alpha\dot{\beta}}\bar{\varepsilon}^{\dot{\beta}}\partial_\mu\phi \tag{D.13}$$

Hence for the conjugate field we get

$$\delta_\varepsilon \bar{\psi}_{\dot{\alpha}} = (\delta_\varepsilon \psi_\alpha)^\dagger = -i\lambda(\bar{\varepsilon}^{\dot{\beta}})^\dagger(\sigma^\mu_{\alpha\dot{\beta}})^*\partial_\mu\phi^\dagger = -i\lambda(\bar{\varepsilon}^{\dot{\beta}})^\dagger(\sigma^\mu_{\dot{\beta}\alpha})^\dagger\partial_\mu\phi^\dagger = -i\lambda\varepsilon^\beta\sigma^\mu_{\beta\dot{\alpha}}\partial_\mu\phi^\dagger \qquad \text{(D.14)}$$

Note that σ^μ under this Hermitean conjugation must be treated as a set of numbers, and is not an operator. Substituting this into the fermion Lagrangian we get

$$\begin{aligned} \delta_\varepsilon \mathcal{L}_{\text{fermion}} &= \lambda\sigma^0(\varepsilon\sigma^\mu\partial_\mu\phi^\dagger\bar{\sigma}^\nu\partial_\nu\psi - \bar{\psi}\bar{\sigma}^\mu\sigma^\nu\bar{\varepsilon}\partial_\mu\partial_\nu\phi) \\ &= \lambda\sigma^0(-\varepsilon\sigma^\mu\bar{\sigma}^\nu\psi\partial_\mu\partial_\nu\phi^\dagger - \bar{\psi}\bar{\sigma}^\mu\sigma^\nu\bar{\varepsilon}\partial_\mu\partial_\nu\phi) \ , \end{aligned}$$

up to total derivatives, which are irrelevant. Because of the symmetric appearance of the derivatives we may replace $[\sigma^\mu\bar{\sigma}^\nu]_\alpha^\beta$ by

$$\tfrac{1}{2}[\sigma^\mu\bar{\sigma}^\nu + \sigma^\nu\bar{\sigma}^\mu]_\alpha^\beta = \eta^{00}\eta^{\mu\nu}\delta_\alpha^\beta \ , \qquad \text{(D.15)}$$

a relation that can easily be checked explicitly. A similar relation holds with bars interchanged and dots on the spinor indices. Integrating once more by parts, we find then that the two variations cancel each other if

$$\lambda = -\sqrt{2}\sigma^0 \qquad \text{(D.16)}$$

One may introduce an operator Q_α that generates the transformation in the quantum theory. Since the result has terms proportional to ε and $\bar{\varepsilon}$ we actually use the combination $\varepsilon Q + \bar{Q}\bar{\varepsilon}$. One can derive this operator as the charge of the Noether current of supersymmetry. It is in general some bi-linear expression in terms of the quantum fields. Here we will simply define it by its transformation properties, namely

$$\left(\varepsilon Q + \bar{Q}\bar{\varepsilon}\right) X = \delta_\varepsilon X \ , \qquad \text{(D.17)}$$

where X denotes any field. Since Q has a spinor index, it is natural to take it to be anti-commuting, which indeed it turns out to be. Then $\varepsilon Q + \bar{Q}\bar{\varepsilon}$ is a bosonic operator.

As usual with generators of a symmetry, it is interesting to study their commutator. Consider

$$\left[\varepsilon_1 Q + \bar{Q}\bar{\varepsilon}_1, \varepsilon_2 Q + \bar{Q}\bar{\varepsilon}_2\right] . \qquad \text{(D.18)}$$

To see what the result is, it is easiest to make it act on the generic field X.

$$\left[\varepsilon_1 Q + \bar{Q}\bar{\varepsilon}_1, \varepsilon_2 Q + \bar{Q}\bar{\varepsilon}_2\right] X = (\delta_{\varepsilon_1}\delta_{\varepsilon_2} - \delta_{\varepsilon_2}\delta_{\varepsilon_1})X \qquad \text{(D.19)}$$

This is most easily computed if we take a scalar ϕ:

$$\delta_{\varepsilon_1}\delta_{\varepsilon_2}\phi - \delta_{\varepsilon_2}\delta_{\varepsilon_1}\phi = -2i\sigma^0(\varepsilon_2\sigma^\mu\bar{\varepsilon}_1 - \varepsilon_1\sigma^\mu\bar{\varepsilon}_2)\partial_\mu\phi \qquad \text{(D.20)}$$

If indeed this holds also for other choices of X (as we will check in a moment), then we have

$$\left[\varepsilon_1 Q + \bar{Q}\bar{\varepsilon}_1, \varepsilon_2 Q + \bar{Q}\bar{\varepsilon}_2\right] = -2i\sigma^0(\varepsilon_2\sigma^\mu\bar{\varepsilon}_1 - \varepsilon_1\sigma^\mu\bar{\varepsilon}_2)\partial_\mu \qquad \text{(D.21)}$$

Now we expand both sides in ε and compare the terms. We see then that $\{Q_\alpha, Q_\beta\} = \{\bar{Q}_{\dot\alpha}, \bar{Q}_{\dot\beta}\} = 0$ and furthermore,

$$\left\{Q_\alpha, \bar{Q}_{\dot\alpha}\right\} = 2i\sigma^0\sigma^\mu_{\alpha\dot\alpha}\partial_\mu \tag{D.22}$$

For the operators Q this means that the corresponding commutator must yield the operator that generates translations on the fields, *i.e.* the momentum operator. Indeed, if one does the explicit computation one gets

$$\left\{Q_\alpha, \bar{Q}_{\dot\alpha}\right\} = 2\sigma^0\eta^{00}\sigma^\mu_{\alpha\dot\alpha}P_\mu = 2\sigma^0\eta^{00}\sigma_{\mu,\alpha\dot\alpha}P^\mu \ . \tag{D.23}$$

To see that the sign is correct, note the following. Independent of the metric we have $p^\mu = (E, \vec{p})$, $q^\mu = (t, \vec{x})$, and $[p^i, x^j] = -i\delta^{ij}$. Therefore $[p^\mu, q^\nu] = i\eta^{00}\eta^{\mu\nu}$. Hence $[p_\mu, q^\nu] = i\eta^{00}\delta^\nu_\mu$. This implies that $p_\mu = i\eta^{00}\partial_\mu$, so that the relation between Eqs. (D.22) and (D.23) is correct.

If only the time-like components $P^0 = H$ contribute the right-hand side is

$$2\sigma^0\eta^{00}\sigma_{0,\alpha\dot\alpha}H = 2\sigma^0\sigma^0_{\alpha\dot\alpha}H = H\delta_{\alpha\dot\alpha} \tag{D.24}$$

This shows that any dependence on the metric and the choice of σ^0 nicely cancels. The overall sign is the right one: the anti-commutator has non-negative expectation values between states, consistent with positivity of the spectrum of H.

Now we must still consider the commutator of two supersymmetries on the fermion field. The result is

$$\begin{aligned} \delta_{\varepsilon_1}\delta_{\varepsilon_2}\psi_\alpha &= -i\sqrt{2}\sigma^0(\sigma^\mu\bar\varepsilon_2)_\alpha\partial_\mu\delta_{\varepsilon_1}\phi \\ &= -2i\sigma^0(\sigma^\mu\bar\varepsilon_2)_\alpha\varepsilon_1\partial_\mu\psi \end{aligned} \tag{D.25}$$

This does not have exactly the right form, but we may use the following Fiertz identity

$$\chi_\alpha(\xi\eta) = -\xi_\alpha(\eta\chi) - \eta_\alpha(\chi\xi) \ , \tag{D.26}$$

which can be proved by writing out both sides of the identity. Applying it to Eq. (D.25) yields

$$\begin{aligned} \delta_{\varepsilon_1}\delta_{\varepsilon_2}\psi_\alpha &= 2i\sigma^0((\varepsilon_1)_\alpha\partial_\mu\psi(\sigma^\mu\bar\varepsilon_2) + \partial_\mu\psi_\alpha(\sigma^\mu\bar\varepsilon_2)\varepsilon_1) \\ &= 2i\sigma^0(-(\varepsilon_1)_\alpha\bar\varepsilon_2\bar\sigma^\mu\partial_\mu\psi + \partial_\mu\psi_\alpha\varepsilon_1(\sigma^\mu\bar\varepsilon_2)) \end{aligned} \tag{D.27}$$

The second term has precisely the right form, but the first one does not. However, it vanishes if we assume that ψ satisfies the equation of motion (the Dirac equation) $\bar\sigma^\mu\partial_\mu\psi = 0$. This implies that the super algebra Eq. (D.23) only holds "on-shell", *i.e.* for fields that satisfy the equation of motion.

This is an annoying feature in a quantum field theory where off-shell degrees of freedom do play a rôle in virtual processes. Of course it is not a fundamental problem, since we can simply compute scattering amplitudes ignoring the supersymmetry, but it becomes then very difficult to show that supersymmetry is preserved in such calculations.

The problem can be circumvented by introducing an auxiliary field F. It is not a dynamical degree of freedom, as is clear from its action:

$$\mathcal{L}_{\rm aux} = F^\dagger F \tag{D.28}$$

The equation of motion for F is $F = 0$, which eliminates F on-shell. However, we may use F by changing the transformation of ψ to produce a term involving F, and adding a transformation of F that involves ψ, so that in the combination of these two transformations the undesired term cancels.

Since F vanishes on-shell, its transformation must also vanish on-shell. Hence we try

$$\delta_\varepsilon F = \xi\bar{\varepsilon}\bar{\sigma}^\mu\partial_\mu\psi; \qquad \delta_\varepsilon F^\dagger = \xi^*\partial_\mu\bar{\psi}\bar{\sigma}^\mu\varepsilon \tag{D.29}$$

Then

$$\delta_\varepsilon\mathcal{L}_{\rm aux} = \xi^*\partial_\mu\bar{\psi}\bar{\sigma}^\mu F + F^\dagger\xi\bar{\varepsilon}\bar{\sigma}^\mu\partial_\mu\psi \tag{D.30}$$

These terms are canceled if we transform the fermions as

$$\delta_\varepsilon\psi_\alpha = -i\sqrt{2}\sigma^0(\sigma^\mu\bar{\varepsilon})_\alpha\partial_\mu\phi + \omega\varepsilon_\alpha F \tag{D.31}$$

$$\delta_\varepsilon\bar{\psi}_{\dot{\alpha}} = i\sqrt{2}\sigma^0(\varepsilon\sigma^\mu)_{\dot{\alpha}}\partial_\mu\phi^\dagger + \omega^*\bar{\varepsilon}F^\dagger \tag{D.32}$$

with $\omega = -i\sigma^0\xi^*$. The second condition ξ and ω have to satisfy is that the extra terms in the commutator Eq. (D.27) cancel. This leads to the condition $\omega\xi = -2i\sigma^0$. Combining the two conditions we find $\xi^*\xi = 2$. The phase of ξ is not determined, and this is not surprising since it can be absorbed in F. We will choose $\xi = -i\sqrt{2}\sigma^0$, so that $\omega = \sqrt{2}$.

The last thing to check is that the commutator acting on F gives the same answer as on ϕ and ψ. This is true without further conditions. The commutator on F produces a term involving ϕ, but this term has the form $\bar{\varepsilon}_2\bar{\sigma}^\mu\sigma^\nu\bar{\varepsilon}_1\partial_\mu\partial_\nu\phi - (1\leftrightarrow 2)$. This is proportional to $\bar{\varepsilon}_1\bar{\varepsilon}_2 - (1\leftrightarrow 2) = 0$.

The transformations we have obtained are thus

$$\begin{aligned}
\delta_\varepsilon\phi &= \sqrt{2}\varepsilon\psi \\
\delta_\varepsilon\psi_\alpha &= -i\sqrt{2}\sigma^0(\sigma^\mu\bar{\varepsilon})_\alpha\partial_\mu\phi + \sqrt{2}\varepsilon_\alpha F \\
\delta_\varepsilon F &= -i\sqrt{2}\sigma^0\bar{\varepsilon}\bar{\sigma}^\mu\partial_\mu\psi
\end{aligned}$$

The physical reason why we need auxiliary fields is that the off-shell count of degrees of freedom between bosons and fermions does not match. Off-shell a complex scalar has one complex degrees of freedom, and a Weyl spinor has two complex degrees of freedom. The equation of motion for a scalar do not modify the number degrees of freedom. In momentum space, they only impose the constraint $k^2 = 0$ (if the scalar is massless). For fermions they impose the same constraint, but also the stronger constraint $k^\mu\gamma_\mu\psi = 0$. This is a matrix condition that only half the components can satisfy. The other half is eliminated on-shell. Hence on-shell a Weyl spinor has one complex degree of freedom, the same as the scalar. This is what makes the existence of on-shell supersymmetry possible. To realize it off-shell we need to introduce the "missing" bosonic degrees of freedom in the form of the complex auxiliary field F.

D.3 Superfields

The model studied in the previous section has free fields only. Now we have to find out how to write down supersymmetric interactions, and we also need to consider fields of higher spin, to accommodate gauge bosons (and also gravitons).

The construction of Lagrangians for theories with $N = 1$ supersymmetry is most conveniently done using superfields. One introduces anti-commuting parameters θ^α which transform according to the $SO(3,1)$ representation (2,1). Their Hermitean conjugates transform then as (1,2), and are denoted as $\bar{\theta}^{\dot{\alpha}}$. These parameters anti-commute with each other for any choice of indices. They also anti-commute with any other fermionic field or operator.

The supersymmetry algebra can now be written as,

$$[\theta Q, \bar{Q}\bar{\theta}] = 2(\sigma^0\eta^{00})\theta\sigma^\mu\bar{\theta}P_\mu \ . \tag{D.33}$$

To simplify the notation somewhat we introduce $\rho^\mu = \sigma^0\sigma^\mu$.

D.4 Translations in Superspace

The idea is now to realize this algebra by means of differential operators acting on a space which is ordinary space-time plus two anti-commuting coordinates θ and $\bar{\theta}$. The analogy one may keep in mind is the realization of the angular momentum algebra in terms of the differential operator $\epsilon_{ijk}x_i\partial_j$. In that case representations of the algebra can be constructed in terms of spherical harmonics $Y^l_m(x)$. Here the representations of the super-algebra will be constructed out of fields that depend on x^μ, θ and $\bar{\theta}$. These fields are called superfields.

A superfield is in general a function of x, θ and $\bar{\theta}$. Its Taylor expansion in θ and $\bar{\theta}$ has only a finite number of terms, due to the anti-commutativity of these parameters. The coefficients are functions of x, which we identify with the fields (or the auxiliary fields). We may define a generalized "translation" operator in superspace that has the form

$$G(x, \theta, \bar{\theta}) = e^{i(\eta^{00}P^\mu x_\mu + \theta Q + \bar{Q}\bar{\theta})} \tag{D.34}$$

From the commutator of the operators Q we can derive a product rule, using the Baker-Campbell-Hausdorff formula

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]+\ldots} \ , \tag{D.35}$$

which in this case is exact to this order, since all higher order commutators vanish. We find

$$G(y, \xi, \bar{\xi})G(x, \theta, \bar{\theta}) = G(x + y + i\xi\rho\bar{\theta} - i\theta\rho\bar{\xi}, \xi + \theta, \bar{\xi} + \bar{\theta}) \tag{D.36}$$

This defines an action on the coordinates of superspace,

$$x^\mu \ \rightarrow \ x^\mu + y^\mu + i\xi\rho^\mu\bar{\theta} - i\theta\rho^\mu\bar{\xi}$$

$$\begin{array}{rcl}\theta & \to & \xi+\theta \\ \bar\theta & \to & \bar\xi+\bar\theta\end{array}$$

We now introduce differential operators that yield the same action on the coordinates. These differential operators are denoted as $Q^{\rm diff}$ and $\bar{Q}^{\rm diff}$. These operators are required to reproduce the ξ and $\bar\xi$ terms in the transformation of the super-coordinates:

$$\begin{array}{rcl}\delta_\xi(x,\theta,\bar\theta) & = & \xi Q^{\rm diff}(x,\theta,\bar\theta) \\ \delta_{\bar\xi}(x,\theta,\bar\theta) & = & \bar{Q}^{\rm diff}\bar\xi(x,\theta,\bar\theta)\end{array}$$

It is easy to check that the following operators do the job

$$\begin{array}{rcl}Q^{\rm diff}_\alpha & = & \partial_\alpha + i\rho^\mu_{\alpha\dot\beta}\bar\theta^{\dot\beta}\partial_\mu \\ \bar{Q}^{\rm diff}_{\dot\alpha} & = & -\bar\partial_{\dot\alpha} - i\theta^\beta\rho^\mu_{\beta\dot\alpha}\partial_\mu\end{array}$$

Here $\partial_\mu = \frac{\partial}{\partial x^\mu}$. The reason for using a lower index on ∂ on the left and an upper one for x on the left is that in this way $\partial_\mu x^\nu = \delta^\nu_\mu$ is a proper Lorentz invariant tensor while $\delta_{\mu\nu}$ is not (the only is a Lorentz invariant tensor with two lower indices is $\eta_{\mu\nu}$). The indices on the other partial derivatives follow a similar logic:

$$\partial_\alpha = \frac{\partial}{\partial\theta^\alpha}\ ; \qquad \bar\partial_{\dot\alpha} = \frac{\partial}{\partial\bar\theta^{\dot\alpha}}\ , \tag{D.37}$$

where the following sign changes should be noted

$$\partial^\alpha = -\frac{\partial}{\partial\theta_\alpha}\ ; \qquad \bar\partial^{\dot\alpha} = -\frac{\partial}{\partial\bar\theta_{\dot\alpha}}\ . \tag{D.38}$$

These are simply a consequence of raising and lowering indices with the ϵ tensor, *c.f.* Eq. (A.7):

$$\partial^\alpha\theta_\gamma = \epsilon^{\alpha\beta}\partial_\beta\theta^\delta\epsilon_{\delta\gamma} = \epsilon^{\alpha\beta}\epsilon_{\delta\gamma}\delta^\delta_\beta = \epsilon^{\alpha\beta}\epsilon_{\beta\gamma} = -\delta^\alpha_\gamma \tag{D.39}$$

To understand the sign of the first term of $\bar{Q}^{\rm diff}$ note that the derivatives ∂_α and $\bar\partial_{\dot\alpha}$ anti-commute with all other fermionic objects. Hence (ignoring the second term) we have

$$\bar{Q}^{\rm diff}\bar\xi = \bar\xi\bar{Q}^{\rm diff} = -\bar\xi_{\dot\alpha}\partial^{\dot\alpha} = \bar\xi_{\dot\alpha}\frac{\partial}{\partial\bar\theta_{\dot\alpha}} = \bar\xi^{\dot\alpha}\partial_{\dot\alpha} = \bar\xi^{\dot\alpha}\frac{\partial}{\partial\bar\theta^{\dot\alpha}} \tag{D.40}$$

Hence this operator acts correctly as a shift operator on both $\bar\theta^{\dot\alpha}$ and $\bar\theta_{\dot\alpha}$.

The commutator of two of these differential operators yields

$$[\theta Q^{\rm diff}, \bar{Q}^{\rm diff}\bar\theta] = -2i\theta\rho^\mu\bar\theta\partial_\mu\ , \tag{D.41}$$

which differs by a sign from the corresponding quantum operators, Eq. (D.22). For an explanation of this sign see [2]. Here we simply note that for the commutator of the quantum operators the right-hand side yields the Hamiltonian, and the sign is the fixed by the requirement of positivity of that Hamiltonian. Here we simply get a time derivative, and the fact that it appears with the "opposite" sign compared to our (too naive) expectations is not a problem.

D.5 Different Realizations

There are other differential operator realizations of the super-algebra, which act (as we will see) on a slightly modified superspace. The following three sets are used

$$\begin{array}{llcl} \mathrm{S}: & Q_\alpha^{\mathrm{diff}} & = & \partial_\alpha + i\rho^\mu_{\alpha\dot\beta}\bar\theta^{\dot\beta}\partial_\mu \\ & \bar Q_{\dot\alpha}^{\mathrm{diff}} & = & -\bar\partial_{\dot\alpha} - i\theta^\beta\rho^\mu_{\beta\dot\alpha}\partial_\mu \end{array}$$

$$\begin{array}{llcl} \mathrm{L}: & Q_\alpha^{\mathrm{diff}} & = & \partial_\alpha \\ & \bar Q_{\dot\alpha}^{\mathrm{diff}} & = & -\bar\partial_{\dot\alpha} - 2i\theta^\beta\rho^\mu_{\beta\dot\alpha}\partial_\mu \end{array}$$

$$\begin{array}{llcl} \mathrm{R}: & Q_\alpha^{\mathrm{diff}} & = & \partial_\alpha + 2i\rho^\mu_{\alpha\dot\beta}\bar\theta^{\dot\beta}\partial_\mu \\ & \bar Q_{\dot\alpha}^{\mathrm{diff}} & = & -\bar\partial_{\dot\alpha} \end{array}$$

The representations are denoted by S for "Symmetric", L for "Left-handed" and R for "Right-handed". From here on we will only use the differential operators, and since no confusion is possible we drop the superscript "diff".

D.6 Action on superfields

Given these representations we can consider the action of infinitesimal supersymmetry generators $\varepsilon Q + \bar Q\bar\varepsilon$ on fields $\Phi(x^\mu, \theta, \bar\theta)$. The result is

$$\begin{aligned} \delta_\varepsilon\Phi_S &= \left[\varepsilon\partial_\theta - \partial_{\bar\theta}\bar\varepsilon + i(\varepsilon\rho^\mu\bar\theta + \theta\rho^\mu\bar\varepsilon)\partial_\mu\right]\Phi_S \\ \delta_\varepsilon\Phi_L &= \left[\varepsilon\partial_\theta - \partial_{\bar\theta}\bar\varepsilon - 2i\theta\rho^\mu\bar\varepsilon\partial_\mu\right]\Phi_L \\ \delta_\varepsilon\Phi_R &= \left[\varepsilon\partial_\theta - \partial_{\bar\theta}\bar\varepsilon + 2i\varepsilon\rho^\mu\bar\theta\partial_\mu\right]\Phi_R \end{aligned}$$

D.7 Changes of Representation

Any field can be written down in any of these representations. The choice one makes is merely a matter of convenience. One can go from one representation to any other in the following way

$$\Phi_S(x^\mu, \theta, \bar\theta) = \Phi_L(x^\mu - i\theta\rho^\mu\bar\theta, \theta, \bar\theta) = \Phi_R(x^\mu + i\theta\rho^\mu\bar\theta, \theta, \bar\theta)\ . \tag{D.42}$$

To show this one must verify, for example

$$Q^S_\alpha\Phi_L(x^\mu - i\theta\rho^\mu\bar\theta, \theta, \bar\theta) = Q^L_\alpha\Phi_L(x^\mu, \theta, \bar\theta) \tag{D.43}$$

and the same for $\bar Q$. Since the operators Q and $\bar Q$ act in any case the same on the explicit θ-dependence (the second and third argument), all that remains to be shown is

$$Q^S_\alpha\Phi_L(x^\mu - i\theta\rho^\mu\bar\theta) = Q^L_\alpha\Phi_L(x^\mu) \tag{D.44}$$

This is straightforward.

D.8 Product Representations and Supersymmetry Invariants

The product of two superfields in the same representation is again a superfield in that representation. This means that the transformations of the product under supersymmetry are given by the same formulas as for a single superfield, *i.e.* the formulas of the previous paragraph.

Since there are only two anti-commuting variables θ^1 and θ^2 the expansion of a superfield in $\theta, \bar{\theta}$ is finite and stops at the maximal order, *i.e.* with a term $\theta^2\bar{\theta}^2$. The variations of the component fields (*i.e.* the coefficients of the various θ combinations) can be read of by expanding Φ and $\delta_\varepsilon\Phi$ in $\theta, \bar{\theta}$. Consider the highest component in Φ, *i.e.* the coefficient of $\theta^2\bar{\theta}^2$. The corresponding term in $\delta_\varepsilon\Phi$ is definitely not generated by the derivatives ∂_α or $\bar{\partial}_{\dot{\alpha}}$, because to produce $\theta^2\bar{\theta}^2$ they would have to act on $\theta^3\bar{\theta}^2$ or $\theta^2\bar{\theta}^3$, neither of which can occur in Φ. A term proportional to $\theta^2\bar{\theta}^2$ in $\delta_\varepsilon\Phi$ can therefore only arise from the action of the terms involving ∂_μ. This means that the variation of the highest component in Φ transforms into a total derivative. Hence, when integrated over space-time, the highest component of Φ is invariant under supersymmetry. This is the principle which is always used to build supersymmetric actions.

These two principles give the superspace method its power: products of superfields are again superfields, and the "highest" component in the $\theta, \bar{\theta}$ expansion is a supersymmetry invariant. A general superfield has still a large number of component fields (nine, to be precise). In the previous section we have seen an example of a set of three fields (φ, ψ and F) that formed a closed representation of supersymmetry. Hence there should exist ways to restrict the number of component fields. To do this we need yet another set of differential operators.

D.9 Covariant Derivatives

We define partial derivatives D_α and $\bar{D}_{\dot{\alpha}}$ that anti-commute with δ_ε, analogous to covariant derivatives in gauge theories, which commute with gauge transformations. In the three representations these covariant derivatives are

$$\mathrm{S}: \qquad \begin{aligned} D_\alpha &= \partial_\alpha - i\rho^\mu_{\alpha\dot{\beta}}\bar{\theta}^{\dot{\beta}}\partial_\mu \\ \bar{D}_{\dot{\alpha}} &= -\bar{\partial}_{\dot{\alpha}} + i\theta^\beta\rho^\mu_{\beta\dot{\alpha}}\partial_\mu \end{aligned}$$

$$\mathrm{L}: \qquad \begin{aligned} D_\alpha &= \partial_\alpha - 2i\rho^\mu_{\alpha\dot{\beta}}\bar{\theta}^{\dot{\beta}}\partial_\mu \\ \bar{D}_{\dot{\alpha}} &= -\bar{\partial}_{\dot{\alpha}} \end{aligned}$$

$$\mathrm{R}: \qquad \begin{aligned} D_\alpha &= \partial_\alpha \\ \bar{D}_{\dot{\alpha}} &= -\bar{\partial}_{\dot{\alpha}} + 2i\theta^\beta\rho^\mu_{\beta\dot{\alpha}}\partial_\mu \end{aligned}$$

D.10 Chiral Superfields

There are only a few representations of the super algebra that we need to consider. At first sight the fields and the invariant actions do not look very natural, but a very large amount of work is quite simply summarized by these rules.

Fields $\phi(x,\theta,\bar{\theta})$ satisfying $\bar{D}_{\dot{\alpha}}\phi = 0$ are called left-handed chiral superfields (also scalar superfields). The reason that this is an interesting restriction is that $\bar{D}$ anti-commutes with the supersymmetry transformation. Therefore the property $\bar{D}_{\dot{\alpha}}\phi = 0$ is preserved by supersymmetry. This implies that chiral superfields form all by themselves representations of supersymmetry; without the restriction $\bar{D}_{\dot{\alpha}}\phi = 0$ the superfield ϕ has more components then necessary. The restriction in the number of components is most clearly seen in the left-handed representation, since $\bar{D}$ is simplest in that representation. Then the requirement is simply that ϕ should not depend on $\bar{\theta}$. Hence its expansion in terms of θ can go at most to second order:

$$\phi_L(x,\theta) = \varphi(x) + \sqrt{2}\theta\psi(x) + \theta^2 F(x) \ , \tag{D.45}$$

where, according to a previous convention, $\theta^2 = \theta^\alpha\theta_\alpha$.

The supersymmetric variation of this field is

$$\delta_\varepsilon\phi_L \equiv (\epsilon Q + \bar{Q}\bar{\epsilon})\phi_L = \sqrt{2}\varepsilon\psi + 2\varepsilon\theta F - 2i\theta\rho^\mu\bar{\varepsilon}\partial_\mu\varphi + i\sqrt{2}\theta^2\partial_\mu\psi\rho^\mu\bar{\varepsilon} \ , \tag{D.46}$$

where one has to use the identity $\theta^\alpha\theta^\beta\sigma^\mu_{\beta\dot{\alpha}} = -\frac{1}{2}\sigma^{\mu\alpha}{}_{\dot{\alpha}}\theta^2$. In terms of components these transformations are

$$\begin{aligned}
\delta_\varepsilon\varphi &= \sqrt{2}\varepsilon\psi \\
\delta_\varepsilon\psi &= \sqrt{2}\varepsilon F - i\sqrt{2}\rho^\mu\bar{\varepsilon}\partial_\mu\varphi \\
\delta_\varepsilon F &= -i\sqrt{2}\bar{\varepsilon}\bar{\rho}^\mu\partial_\mu\psi \quad ,
\end{aligned}$$

precisely the transformation we obtained in the previous section. Note that if ϕ is a left-handed superfield, all its powers are left-handed superfields as well.

Left-handed superfields have the property that the transformation of the factor of θ^2 is a total derivative. The reason is very similar to the argument given before for an unrestricted superfield. The derivatives ∂_α and $\bar{\partial}_{\dot{\alpha}}$ cannot contribute to θ^2 terms, because there are no terms of third order in θ. Hence only the derivative term can give a contribution. Hence the factor of θ^2 in a left-handed superfield is a good candidate for terms in the Lagrangian.

Fields satisfying $D\phi = 0$ are called right-handed superfields. Note that if ϕ is a left-handed superfield in the left-handed representation, then $\phi^\dagger$ is a right-handed superfield, but in the right-handed representation.

D.11 Vector Superfields

Finally there are real fields (also called "vector superfields") $V^\dagger = V$. This restriction is preserved in the symmetric representation. These fields contain both θ and $\bar{\theta}$, and hence there are many terms in the expansion in θ. However, life can be simplified considerably by a "gauge choice" called the Wess-Zumino gauge. The complete expression is

$$V(x,\theta,\bar{\theta}) = -\theta\rho_\mu\bar{\theta}V^\mu + i\theta^2\bar{\theta}\bar{\lambda} - i\bar{\theta}^2\theta\lambda + \tfrac{1}{2}\theta^2\bar{\theta}^2 D + \dots \ , \tag{D.47}$$

where the dots represent additional terms, which are absent in Wess-Zumino gauge.

D.12 Invariant Actions

Invariant actions are constructed by means of θ integrals. The rules are

$$\int d\theta = 0 \ , \qquad \int \theta d\theta = 1 \ , \tag{D.48}$$

for each component. Furthermore we define $d^2\theta$ in such a way that $\int d^2\theta\theta^2 = 1$, and the same for $\bar{\theta}$. These integrals are nothing else than a fancy way of taking the highest component of a superfield. Formally, they make the action look as an integration not just over space, but over all of superspace. This does not have profound implications for the structure of space-time, however. Superspace should simply be regarded as a convenient bookkeeping device.

The most general supersymmetric action is

$$\int d^4x(d^2\theta\mathcal{L}_F + \text{c.c}) + \int d^4x d^4\theta\mathcal{L}_D \ , \tag{D.49}$$

where $\mathcal{L}_F$ satisfies the conditions for a left-handed chiral superfield and $\mathcal{L}_D$ those of a vector superfield. One usually writes $d^4\theta$ instead of $d^2\theta d^2\bar{\theta}$. We define the normalization so that $\int d^4\theta\theta^2\bar{\theta}^2 = 1$.

The terms $\mathcal{L}$ are built out of elementary superfields describing single particles. The only terms surviving the integration are those corresponding to F and D auxiliary fields, hence the notation. Often one writes

$$\int d^2\theta X \equiv [X]_F, \quad \int d^4\theta X \equiv [X]_D \ . \tag{D.50}$$

The reason that the resulting Lagrangian is invariant under supersymmetry transformations is that the F and D terms in any superfield (whether composite or elementary) transforms into a total derivative. Hence the Lagrangian transforms into a total derivative as well, and the action is invariant.

Consider first the scalar superfields. The kinetic terms come from terms $\phi^\dagger\phi$ in $\mathcal{L}_D$. As observed above, $\phi^\dagger$ is a right-handed superfield in the right-handed representation. If one multiplies two superfields in different representations, the supersymmetry has no meaningful action on the product. One way to deal with this is to write both in the symmetric representation; then $\phi_S^\dagger\phi_S$ transforms as a vector superfield under the S-representation. To got to the S-representation we have to shift the argument x:

$$\phi_S(x,\theta,\bar{\theta}) = \phi_L(x - i\theta\rho\bar{\theta},\theta,\bar{\theta}) \tag{D.51}$$

Now we expand $\phi_S^\dagger \phi_S$ in a Taylor series in $i\theta\rho\bar{\theta}$, and we keep only terms of order $\theta^2\bar{\theta}^2$. Alternatively we may work entirely in the left-handed representation, but then we have to shift the argument of $\phi_L^\dagger$ by twice as much. It is easy to see that the result is the same. The result is

$$\int d^4\theta \phi_L^\dagger(x + 2i\theta\rho\bar{\theta}, \theta)\phi_L(x, \theta, \bar{\theta}) \tag{D.52}$$

consider first the scalar terms. Define $a^\mu = 2i\theta\rho^\mu\bar{\theta}$. The expansion yields, for the scalar component

$$\varphi(x-a) = \varphi(x) + a^\mu\partial_\mu\varphi(x) + \tfrac{1}{2}a^\mu a^\nu\partial_\mu\partial_\nu\varphi(x) + \ldots \tag{D.53}$$

The higher order terms vanish in this case, because they have two many θ's. The contribution of second order in θ and $\bar{\theta}$ to the action is thus

$$\tfrac{1}{2}(a^\mu a^\nu\partial_\mu\partial_\nu)\varphi(x)^\dagger\varphi(x) \tag{D.54}$$

Now we make use of the identity

$$\theta\rho^\mu\bar{\theta}\theta\rho^\nu\bar{\theta} = \tfrac{1}{2}\eta^{00}\eta^{\mu\nu}\theta^2\bar{\theta}^2 \tag{D.55}$$

The θ factors are removed by the integration. The result is

$$\mathcal{L}_{\text{scalar}} = \tfrac{1}{2}(2i)^2\tfrac{1}{2}\eta^{00}\eta^{\mu\nu}\partial_\mu\partial_\nu\varphi(x)^\dagger\varphi(x) \tag{D.56}$$

By partial integration we bring one derivative to φ, and we find

$$\mathcal{L}_{\text{scalar}} = \eta^{00}\eta^{\mu\nu}\partial_\mu\varphi(x)^\dagger\partial_\nu\varphi(x) \tag{D.57}$$

This is indeed the correct form of the scalar kinetic terms.

Now consider the kinetic terms for the fermions. Here we need to expand only to first order in a:

$$\begin{aligned}\mathcal{L}_{\text{fermion}} &= (\sqrt{2}\bar{\psi}_{\dot{\alpha}}(x+a)\bar{\theta}^{\dot{\alpha}})(\sqrt{2}\theta^\alpha\psi_\alpha(x)) \\ &= 4i(\theta\rho^\mu\bar{\theta})\partial_\mu\bar{\psi}_{\dot{\alpha}}\bar{\theta}^{\dot{\alpha}}\theta^\alpha\psi_\alpha \\ &= 4i(\theta\rho^\mu\bar{\theta})\bar{\theta}_{\dot{\alpha}}\theta_\alpha\partial_\mu\bar{\psi}^{\dot{\alpha}}\psi^\alpha\end{aligned}$$

Now we need the following identity

$$(\theta\rho^\mu\bar{\theta})\bar{\theta}_{\dot{\alpha}}\theta_\alpha = -\tfrac{1}{4}\bar{\rho}^\mu_{\dot{\alpha}\alpha}\theta^2\bar{\theta}^2 \tag{D.58}$$

Integrating out the θ's we find then

$$\mathcal{L}_{\text{fermion}} = -i\partial_\mu\bar{\psi}^{\dot{\alpha}}\bar{\rho}^\mu_{\dot{\alpha}\alpha}\psi^\alpha = i\bar{\psi}^{\dot{\alpha}}\bar{\rho}^\mu_{\dot{\alpha}\alpha}\partial_\mu\psi^\alpha \tag{D.59}$$

Finally, the quadratic terms for the auxiliary field comes out immediately as

$$\mathcal{L}_{\text{aux}} = F^*F \tag{D.60}$$

This completes the discussion of the kinetic terms. We see that they have precisely the form we started with in the previous section.

All other terms in the scalar superfield Lagrangian are F-terms. Any polynomial built out of left-handed chiral superfields is manifestly a left-handed chiral superfield as well. It is a bit more difficult to see (but true) that this is the only way to build chiral superfields. It turns out that to get a renormalizable theory one can allow terms of at most third order in the superfields. For a single superfield ϕ the most general polynomial is thus

$$\mathcal{L}_F = \tfrac{1}{2}m\phi^2 + \tfrac{1}{3}\lambda\phi^3 \equiv W(\phi) \tag{D.61}$$

this is called the *superpotential*. It is straightforward to expand it to second order in θ. The result is

$$\begin{aligned}\mathcal{L} &= \int d^4\theta\phi^\dagger\phi + \left[\int d^2\theta \; (\tfrac{1}{2}m\phi^2 + \tfrac{1}{3}\lambda\phi^3) + \text{c.c}\right] \\ &= -\partial_\mu\varphi\partial^\mu\varphi + i\psi\rho^\mu\partial_\mu\bar{\psi} + FF^* + \left[m\left(\varphi F - \tfrac{1}{2}\psi^2\right) + \lambda\left(F\varphi^2 - \varphi\psi^2\right) + \text{c.c}\right] \; .\end{aligned}$$

The field F appears without kinetic terms and can thus be eliminated using the equations of motion (hence the name auxiliary field). Clearly

$$F = -m\varphi^* - \lambda^*(\varphi^*)^2 \tag{D.62}$$

Substituting this back into the action we get

$$\mathcal{L} = -\partial_\mu\varphi\partial^\mu\varphi + i\psi\rho^\mu\partial_\mu\bar{\psi} - \tfrac{1}{2}m(\psi^2 + \bar{\psi}^2) - \lambda\varphi\psi^2 - \lambda^*\varphi^*\bar{\psi}^2 - |m\varphi + \lambda\varphi^2|^2 \tag{D.63}$$

The last term is $-V_F$, where V_F is the contribution to the scalar potential due to F terms.

This is very easily generalized to situations with more than one superfield, and more general superpotentials $W(\phi_i)$. For each term in the polynomial we only need to find the θ^2 terms. If we consider a term $\phi_1 \ldots \phi_k$ we get two kinds of θ^2 terms: one kind consists of F_i and factors $\varphi_{i'}$ for all terms with $i \neq i'$, and the other kind comes from $\sqrt{2}\theta\psi_i\sqrt{2}\theta\psi_j$ times factors φ_k for all $k \neq i, k \neq j$. Hence, including the $F^\dagger F$ terms we get (note that $(\theta\psi_i)(\theta\psi_j) = -\frac{1}{2}(\psi_i\psi_j)\theta^2$)

$$\sum_i F_i^\dagger F_i + \left(\tfrac{1}{2}\sum_{i,j}\psi_i\psi_j\frac{\partial^2 W(\varphi)}{\partial_{\varphi_i}\partial_{\varphi_j}} + \sum_i F_i\frac{\partial W(\varphi)}{\partial_{\varphi_i}} + c.c\right) \tag{D.64}$$

The equation for F is

$$F_i^\dagger = -\partial_{\varphi_i}W(\varphi) \equiv W_i \tag{D.65}$$

The bosonic part of the action has the form

$$(F_i^\dagger - W_i)^\dagger(F_i^\dagger - W_i) - W_i^\dagger W_i \tag{D.66}$$

The equations of motion for F remove the first term, and hence we are left with

$$V = \sum_i\left|\frac{\partial W}{\partial\varphi_i}\right|^2 \tag{D.67}$$

Note in particular that this potential is positive definite, a consequence of supersymmetry.

The Lagrangian for vector fields is more complicated to derive in superfield formalism. If one has a non-abelian gauge group there will be an adjoint multiplet of vector superfields V^a. To write down the coupling to a chiral superfield ϕ one contracts them with the generators T^a of the gauge group in the representation of ϕ. The minimal coupling to the chiral superfield is then the D term in $\phi^\dagger e^{2gV}\phi$. Expanding this in components yields

$$-|D_\mu\varphi|^2 - i\psi\rho^\mu D_\mu\bar{\psi} + 2ig[\varphi^*\lambda\psi - \varphi\bar{\lambda}\bar{\psi}] + FF^* + g\varphi^* D\varphi \ . \tag{D.68}$$

The explicit indices have been suppressed, but are uniquely determined by gauge invariance. For example the third term is explicitly $i\varphi_i^*\lambda^a T^a_{ij}\psi_j +$c.c, and the last one, involving the auxiliary field is

$$g\varphi_i^* T^a_{ij} D^a \varphi_j \ . \tag{D.69}$$

To write down the gauge kinetic terms one introduces a chiral superfield with a spinor index

$$W_\alpha = \bar{D}\bar{D}e^{-gV}D_\alpha e^{gV} \tag{D.70}$$

The supersymmetric and gauge-invariant Lagrangian is

$$\mathcal{L}_{\text{gauge}} = \frac{1}{32g^2}[W^\alpha W_\alpha]_{\text{F}} + \text{c.c} = -\tfrac{1}{4}(F^a_{\mu\nu})^2 - \tfrac{1}{2}\bar{\Lambda}\gamma^\mu D_\mu \Lambda + \tfrac{1}{2}(D^a)^2 \ , \tag{D.71}$$

where Λ is a four component Majorana spinor built out of the spinors λ and $\bar{\lambda}$ in the superfield V:

$$\Lambda = \begin{pmatrix} \lambda^\alpha \\ \bar{\lambda}_{\dot{\alpha}} \end{pmatrix} \ . \tag{D.72}$$

Of course Λ has both a Dirac index and an adjoint gauge index, and D_μ is the gauge covariant derivative in the adjoint representation.

In the absence of any matter multiplets, the auxiliary field must vanish; in the presence of matter it satisfies the field equation

$$D^a = -g\varphi_i^* T^a_{ij}\varphi_j \ , \tag{D.73}$$

because of the D-term contribution (D.69). Substituting this back into the action we get another contribution to the scalar potential, this time associated with the D-terms of the scalars:

$$V_D = \tfrac{1}{2}D^a D^a \ , \tag{D.74}$$

with D^a given by Eq. (D.73).

There is one additional term that can appear in the action, namely

$$\int d^2\theta d^2\bar{\theta} 2\xi V \tag{D.75}$$

which is gauge invariant if and only if V is a $U(1)$ gauge field. The only effect this term has is to add to the Lagrangian a term ξD, where D is the auxiliary field in V. This changes the equations of motion for D, and instead of Eq. (D.73) we get

$$D = -g'\varphi_i^* Q_i \varphi_i - \xi \ , \tag{D.76}$$

where g' is the $U(1)$ coupling constant and Q_i the charge of the scalar i. The action, expressed in terms of the auxiliary fields is still given by Eq. (D.74), where the implicit sum now includes the $U(1)$ factor.

Acknowledgements

I would like to thank all the students who have contributed to these notes by informing me about typos and errors. Special thanks to Rob Verheyen, Melissa van Beekveld, Leon Groenewegen, Gillian Lustermans, John van de Wetering, Stan Jacobs, Marrit Schutten and Chris Ripken.

References

[1] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, and J. March-Russell. String Axiverse. *Phys.Rev.*, D81:123530, 2010.

[2] J. Bagger and J. Wess. *Supersymmetry and Supergravity.* Princeton University Press, 1992.

[3] S. M. Bilenky, J. Hosek, and S. Petcov. On Oscillations of Neutrinos with Dirac and Majorana Masses. *Phys.Lett.*, B94:495, 1980.

[4] A. Buras, J. R. Ellis, M. Gaillard, and D. V. Nanopoulos. Aspects of the Grand Unification of Strong, Weak and Electromagnetic Interactions. *Nucl.Phys.*, B135:66–92, 1978.

[5] B. de Wit and J. Smith. *Field Theory and Particle Physics.*

[6] M. Dine, W. Fischler, and M. Srednicki. A Simple Solution to the Strong CP Problem with a Harmless Axion. *Phys.Lett.*, B104:199, 1981.

[7] M. Doi, T. Kotani, H. Nishiura, K. Okuda, and E. Takasugi. CP Violation in Majorana Neutrinos. *Phys.Lett.*, B102:323, 1981.

[8] P. D. G. C. P. et. al.(. Review of particle physics (2016). *Chin. Phys.*, C40, 2016.

[9] P. Fayet. Introduction to supersymmetric theories of particles and interactions. pages 103–155, 1993.

[10] R. Feger and T. W. Kephart. LieART - A Mathematica Application for Lie Algebras and Representation Theory. *Comput. Phys. Commun.*, 192:166–195, 2015.

[11] S. Ferrara. Tensor Calculus and the breaking of Local Supersymmetry. 1986.

[12] C. Ford, D. Jones, P. Stephenson, and M. Einhorn. The Effective potential and the renormalization group. *Nucl.Phys.*, B395:17–34, 1993.

[13] H. Georgi and S. Glashow. Unity of All Elementary Particle Forces. *Phys.Rev.Lett.*, 32:438–441, 1974.

[14] H. Georgi, H. R. Quinn, and S. Weinberg. Hierarchy of Interactions in Unified Gauge Theories. *Phys. Rev. Lett.*, 33:451–454, 1974.

[15] A. H. Guth. The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems. *Phys.Rev.*, D23:347–356, 1981.

[16] H. E. Haber. Introductory low-energy supersymmetry. 1993.

[17] H. E. Haber and G. L. Kane. The Search for Supersymmetry: Probing Physics Beyond the Standard Model. *Phys.Rept.*, 117:75–263, 1985.

[18] L. E. Ibanez. Grand Unification with Local Supersymmetry. *Nucl.Phys.*, B218:514, 1983.

[19] J. E. Kim. Weak Interaction Singlet and Strong CP Invariance. *Phys.Rev.Lett.*, 43:103, 1979.

[20] P. Langacker. Grand Unified Theories and Proton Decay. *Phys.Rept.*, 72:185, 1981.

[21] L. Lehman. Extending the Standard Model Effective Field Theory with the Complete Set of Dimension-7 Operators. *Phys. Rev.*, D90(12):125023, 2014.

[22] Z. Maki, M. Nakagawa, and S. Sakata. Remarks on the unified model of elementary particles. *Prog. Theor. Phys.*, 28:870–880, 1962.

[23] H. P. Nilles. Supersymmetry, Supergravity and Particle Physics. *Phys.Rept.*, 110:1–162, 1984.

[24] H. P. Nilles. Beyond the standard model. TASI Lectures, 1990.

[25] R. Oerter. The theory of almost everything: The standard model, the unsung triumph of modern physics. 2006.

[26] B. Pontecorvo. Inverse beta processes and nonconservation of lepton charge. *Sov. Phys. JETP*, 7:172–173, 1958. [Zh. Eksp. Teor. Fiz.34,247(1957)].

[27] M. Sher. Electroweak Higgs Potentials and Vacuum Stability. *Phys.Rept.*, 179:273–418 (Phys.Lett. B317 (1993) 159–163, Addendum–ibid. B331 (1994) 448), 1989.

[28] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov. Can Confinement Ensure Natural CP Invariance of Strong Interactions? *Nucl. Phys.*, B166:493–506, 1980.

[29] R. Slansky. Group Theory for Unified Model Building. *Phys.Rept.*, 79:1–128, 1981.

[30] C. Vafa and E. Witten. Parity Conservation in QCD. *Phys. Rev. Lett.*, 53:535, 1984.

[31] P. Van Nieuwenhuizen. Supergravity. *Phys.Rept.*, 68:189–398, 1981.

[32] S. Weinberg. A New Light Boson? *Phys.Rev.Lett.*, 40:223–226, 1978.

[33] S. Weinberg. Baryon and Lepton Nonconserving Processes. *Phys. Rev. Lett.*, 43:1566–1570, 1979.

[34] F. Wilczek. Problem of Strong p and t Invariance in the Presence of Instantons. *Phys.Rev.Lett.*, 40:279–282, 1978.

[35] E. Witten. An SU(2) Anomaly. *Phys. Lett.*, B117:324–328, 1982.

[36] A. P. Zhitnitskii. *Sov. J. Nucl.*, 31:260, 1980.

[37] F. Zwirner. The quest for low-energy supersymmetry and the role of high-energy e+ e- colliders. 1991.

[38] F. Zwirner. Supersymmetric Higgs bosons: A Theoretical introduction. 1993.