Beyond the Standard Model

Lecture in WS 2024/25 at the KFU Graz

Axel Maas

Contents

Chapter 1

Introduction

At the end of 2009 the largest particle physics experiment so far has been started, the LHC at CERN. With proton-proton collisions at a center-of-mass energy of up to 14 TeV, there are two major objectives. One is to complete the current picture of the standard model of particle physics. To do so, it was required to find the Higgs boson, the last missing particle in the standard model. This has been achieved in 2012.

The second objective is to search for new physics beyond the standard model. For various reasons it is believed that there will be new phenomena appearing in particle physics at a scale of 1 TeV. Though this is not guaranteed, there is motivation for it, as will be discussed in section 3.5.

Afterwards, a number of possibilities will be presented. In particular, grand unified theories, technicolor, extended Higgs sectors or additional (possibly hidden) sectors extending the standard model in one way or another by additional forces and particles will be presented. These are candidates to resolve some of these issues. A more elaborate approach is to impose a new structure on particle physics. This is done in particular by supersymmetry as an extra (though broken) symmetry of nature. This is the one extension most commonly believed to be the candidate for beyond-the-standard-model physics. Supersymmetry is in general also an important ingredient in theories which go a step further and endow the very arena of physics, space-time, by a different structure. In particular, supergravity theories and string theories do so. An example of a (non-supersymmetric) string theory will be given at the end of this lecture. A simpler case of such an extension is given by large extra dimensions, which will be discussed beforehand.

A useful list of literature for the present lecture is given by

- Aitchison, "Supersymmetry in particle physics" (Cambridge)
- Andersen et al., "Discovering technicolor", 1104.1255

- Bambi et al. “Introduction to particle cosmology”, Springer
- Bedford, “An introduction to string theory”, 1107.3967
- Böhm, Denner, and Joos, “Gauge theories”, Teubner
- Cheng, “Introduction to extra dimensions”, 1003.1162
- Dolgov, “Cosmology and physics beyond the standard model”, Cosmology and Gravitation, American Institute of Physics
- Han et al., “Kaluza-Klein states from large extra dimensions”, hep-ph/98113504
- Hill and Simmons, “Strong dynamics and electroweak symmetry breaking”, hep-ph/0203079
- Kalka, “Supersymmetrie”, Teubner
- Lane, “Two lectures on technicolor”, hep-ph/0202255
- Maas, “Brout-Englert-Higgs physics: From foundations to phenomenology”, 1712.04721
- Morrissey et al., “New physics at the LHC”, 0912.3259
- Piai “Lectures on walking technicolor, holography, and gauge/gravity dualities”, 1004.0176
- Polchinski, “String theory”, Cambridge University Press
- Rovelli, “A dialog on quantum gravity”, hep-th/0310077
- Shifman, “Advanced topics in quantum field theory”, Cambridge
- Siegel, “Fields”, hep-th/9912205
- Weinberg, “The quantum theory of fields III” (as well as I and II) (Cambridge)
- Wess et al. “Supersymmetry and supergravity” (Princeton)

However, the topic is developing, and will be even more rapdiyl do so as soon as something is found at the LHC or elsewhere. In particular, it is not possible to cover only a serious fraction of all proposals for physics beyond the standard model. This is particularly true, as most proposals features sufficient freedom so that they can be adapted to any new observation being in conflict with them. Hence, this lecture can only present a small selection, which is necessarily both not even exhaustive on a principle level and a subjective selection by the lecturer. Still, all the more popular proposals in their most general form should be covered.

Note that this lecture has necessarily some overlap with the astroparticle physics lecture and the advanced general relativity and quantum gravity lecture, as those fields are tightly connected. However, here the emphasis will be on microscopical models and their tests and predictions in earthbound experiments, rather than on cosmological implications. Furthermore, this brings with it that the focus will be on observables accessible in colliders, and thus frequently the Higgs and/or electroweak observables.

Finally, right now new experimental results and observations from astronomy flood in on an almost weekly basis. Essentially all of them confirm our knowledge, and those which do not have often relevant uncertainties attached to them. Taking this seriously is very important, as not doing so has given rise to various false claims of new physics, as will be discussed in section 3.6.1. Therefore, adding a discussion of the current experimental situation makes no sense within these lecture notes, as it will be probably outdated by the time the lecture is actually given. I will therefore only comment orally on new developments, and of course adapt the lecture for any developing situations.

Finally, I would like to point out that the research of my group, my collaborators, and myself has in the last few years rose doubt about the standard way how new theories beyond the standard model are constructed, especially low-energetic completions. These findings do contradict at some points the models presented in this lecture. However, at the current time our results are not yet established beyond doubt, and certainly not mainstream. Also, frequently the current ideas form an integral part to understand our results, as well as it is necessary to understand mainstream research in this area and its history. I therefore present in this lecture the current mainstream ideas on new physics. I will only briefly introduce our own ideas in the section 7.7 on more recent theoretical developments.

Chapter 2

A brief reminder of known physics

2.1 The standard model

2.1.1 The sectors of the standard model

The[1] standard model of elementary particle physics is our best description of high-energy physics up to an energy of about a few hundred GeVs to one TeV[2]. Within the standard model there exists a number of sectors. One sector is the matter sector. It contains three generations, or families, of matter particles. These particles are fermions, i.e., they have spin 1/2. Each generation contains four particles, which are split into two subsets, quarks and leptons. The different particles types are called flavors.

The first family contains the up and down quarks, having masses about 2-5 MeV each, with the down quark being heavier than the up quark. Since their mass is very small compared to the scale of the strong interactions, around 1 GeV, it is very hard to measure their mass accurately, even at large energies. The leptons are the electron and the electron neutrino. The electron has a mass of 511 keV. The masses of the neutrinos will be discussed after the remaining generations have been introduced. All stable matter around us, i. e., nuclei and atoms, are just made from the first family. Particles from the other families decay to the first family on rather short time-scales, and can therefore only be generated in the laboratory, in high-energy natural processes, or virtually.

The other two families are essentially identical copies of the first one, and are only distinguished by their mass. The second family contains the strange quark, with a mass

[1]The following contains contributions from Hill and Simmons, "Strong dynamics and electroweak symmetry breaking", hep-ph/0203079, 2003 and Morrissey, Plehn, and Tait "New physics at the LHC", 0912.3259, 2009.

[2]For a detailed introduction to the standard model see also the lecture on the standard model.

between 80 and 100 MeV, and the charm quark with a mass of about 1.5 GeV. The leptons in this family are the muon with roughly 105 MeV mass, and its associated neutrino, the muon neutrino. The final, third, family contains the bottom quark with a mass of about 4.2 GeV, and the extraordinary heavy top quark with a mass of about 175 GeV. The corresponding leptons are the tau with 1777 MeV mass and its associated tau neutrino.

Of the neutrino masses only an upper limit is known, which is roughly 0.2 eV. However, it is sure that their mass, whatever it is, is not the same for all neutrinos, but the masses differ by 50 meV and 9 meV. It is, however, not clear yet, whether one of the neutrinos is massless, or which of the neutrinos is heaviest. It could be either that the one in the first family is lightest, which is called a normal hierarchy of masses, or it could be heaviest, which is called an inverted hierarchy. Experimental results favor so far a normal hierarchy, but this is not yet beyond doubt. Also, it is not yet known whether the actual masses are of the same order as the mass differences, or much larger. Both is still compatible with the data. Advanced direct measurements of the neutrino mass should help clarify at least a few of these questions until 2030.

These matter particles interact. The particles mediating the forces are called force carriers and make up the force sector. The quarks have a force, which is exclusive to them, the strong force, which binds together the nucleons in nuclei and quarks into nucleons or in general hadrons. This strong force is mediated by gluons, massless spin-1 particles. The description of the strong interactions is by a gauge theory, called quantumchromodynamics, or QCD for short. Quarks and gluons can be arranged as multiplets of the gauge group of QCD, which is SU(3). The associated charges are called color, and there are three quark colors and three anti-quark colors, as well as eight gluon colors. From a group-theoretical point of view, the (anti-)quarks appear in the (anti-)fundamental representation of SU(3) and the gluons in the adjoint representation.

All matter particles are affected by the weak force, visible in, e. g., β-decays. It is transmitted by the charged $W^{\pm}$ bosons and the neutral Z boson. These bosons also have spin 1, but, in contrast to the gluons, are massive. The W bosons have about 81 GeV mass, while the Z boson has a mass of about 90 GeV. Thus, this force only acts over short distances. This force is described by the weak interaction, again a gauge theory. The gauge group of this theory is SU(2), into which all particles can be arranged as doublets. However, this interaction violates parity maximally, and thus only couples to left-handed particles. But it is in a sense even stranger, as it not couples to the particles of the matter sector directly, but only to certain linear combinations, which also contain admixtures of right-handed particles proportional to the mass of the particles. This behavior is parametrized, though not explained, by the CKM and PMNS matrix for the quarks and for the leptons, respectively. It is mysteriously very different for both, the one for the quarks being strongly diagonal-dominant, while the one for the leptons more or less equally occupied. Both introduce also an explicit violation of CP into the standard model. The actual amount for the quarks is quite large, even though the actual process is kinematically substantially suppressed. For the leptons it is not yet firmly established, but experiments strongly hint at a non-zero, and possibly even maximal, CP violating effect. However, also for the leptons the actual consequences are strongly suppressed, in this case by the small neutrino masses.

Finally, all electrically charged particles, and thus everything except gluons and neutrinos, are affected by the electromagnetic interactions. These are mediated by the photons, massless spin-1 particles. The corresponding theory is again a gauge theory, having gauge group U(1). It is actually entangled with the weak interactions in a certain way, and thus both theories are often taken together as the electroweak sector of the standard model.

Together with the strong interactions, the gauge group of the standard model is therefore $\mathrm{SU}(3)_{\mathrm{color}}\times\mathrm{SU}(2)_{\mathrm{weak}}\times\mathrm{U}(1)_{\mathrm{em}}$[3]. Obtaining this structure in theories beyond the standard model will be a recurring theme in this lecture. It should be noted that this group structure is not directly related to the actual group structure. In particular, the groups SU(2) and U(1) are the weak isospin and hypercharge groups, and a mixture of them finally represents the weak interactions and the electromagnetic interactions. In particular, left-handed fermions and right-handed fermions have different hypercharges while they have the same electromagnetic charges.

However, because of the parity violation of the weak interactions, the masses of the particles cannot be intrinsic properties of them, as otherwise no consistent gauge theory can be formulated. Therefore, the mass is attributed to be a dynamically generated effect. Its origin is from the dynamics of the Higgs particle, which interacts with all fields of the standard model except gluons. Still, it is often taken to be a part of the electroweak sector[4]. This particle is a scalar boson, and is by now experimentally as well established as the other particles in the standard model.

The particular self-interactions of the Higgs particle obscures the gauge group, they hide or, casually spoken, break the symmetry group of the standard model down to $\mathrm{SU}(3)_{\mathrm{color}}\times\mathrm{U}(1)_{\mathrm{em}}$. This occurs, because the Higgs field forms a condensate, very much like Cooper pairs in a superconductor. As a consequence of the interactions with this condensate the particles directly interacting with the Higgs boson acquire a mass, i. e.

[3]Actually, it is $\mathrm{S}(\mathrm{U}(3)\times\mathrm{U}(2))=(\mathrm{SU}(3)/\mathrm{Z}_3)_{\mathrm{color}}\times(\mathrm{SU}(2)/\mathrm{Z}_2)_{\mathrm{weak}}\times\mathrm{U}(1)_{\mathrm{em}}$, to be precise. This is actually not a trivial matter, and can be used as a restriction when constructing grand-unified theories in chapter 7, see e. g. O'Raifeartaigh "Group structure of gauge theories", Cambridge, 1986.

[4]See the lecture on electroweak physics from SS 2016.

all quarks and leptons and the weak gauge bosons W and Z. Only the photon remains massless, despite its coupling to the Higgs, as it endows the unbroken $\mathrm{U}(1)_{\mathrm{em}}$ symmetry.

2.1.2 The Brout-Englert-Higgs effect

This Brout-Englert-Higgs effect is a very generic process[5], and it reappears in different forms in the majority of beyond-the-standard-model (BSM) scenarios to be described in this lecture, and also in the literature. It is therefore worthwhile to detail it more for the standard model. Begin by considering the SU(2)×U(1) part of the standard model with one complex scalar field in the fundamental representation of the weak isospin group SU(2). The covariant derivative is given by

$$\begin{aligned} iD_\mu &= i\partial_\mu - g_i W_\mu^a Q_a - g_h B_\mu \frac{y}{2} \\ &= i\partial_\mu - g_i W_\mu^+ Q^- - g_i W_\mu^- Q^+ - g_i W_\mu^3 Q^3 - g_h B_\mu \frac{y}{2} \end{aligned}$$

with the charge basis expressions

$$\begin{aligned} Q^\pm &= \frac{(Q^1 \pm iQ^2)}{\sqrt{2}} \\ W_\mu^\pm &= \frac{W_\mu^1 \pm iW_\mu^2}{\sqrt{2}}. \end{aligned}$$

Note that there are two gauge coupling constants, g_i and g_h for the subgroups SU(2) and U(1), respectively, which are independent. The hypercharge y of the particles are, in the standard model, an arbitrary number, and have to be fixed by experiment. The relevance of this observation will be discussed in section 2.1.4, and in particular in chapter 7. The Q^a are the generators of the gauge group SU(2), and satisfy the algebra

$$[Q^a, Q^b] = i\epsilon_{abc} Q^c$$

within the representation t of the matter field on which the covariant derivative acts. In the standard model, these are either the fundamental representation $t = 1/2$, i. e. doublets, and thus the $Q^a = \tau^a$ are just the Pauli matrices, or singlets $t = 0$, in which case it is the trivial representation with the $Q^a = 0$.

Returning to the gauge bosons, linear combinations

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

[5]This presentation is quite simplified, but the standard view. A more accurate quantum-field-theoretical description will be given in section 7.7.

can be written where Z_μ (A_μ) is the Z-boson (photon). Then the electromagnetic coupling constant e is defined as

$$g_i \sin\theta_W = e = g_h \cos\theta_W, \tag{2.1}$$

implying the relation

$$\frac{1}{e^2} = \frac{1}{g_i^2} + \frac{1}{g_h^2}.$$

This definition (2.1) introduces the weak mixing, or Weinberg

$$\tan\theta_W = \frac{g_h}{g_i}.$$

The conventional electric charge, determining the strength of the coupling to the photon field A_μ, is thus defined as

$$eQ = e\left(Q^3 + \frac{y}{2}\underline{1}\right), \tag{2.2}$$

where $\underline{1}$ is the unit matrix in the appropriate representation of the field, i. e. either the number one or the two-dimensional unit matrix.

The total charge assignment for the standard model particles is then

- Left-handed neutrinos: $t = 1/2$, $t_3 = 1/2$, $y = -1$ ($Q = 0$), color singlet
- Left-handed leptons: $t = 1/2$, $t_3 = -1/2$, $y = -1$ ($Q = -1$), color singlet
- Right-handed neutrinos: $t = 0$, $y = 0$ ($Q = 0$), color singlet
- Right-handed leptons: $t = 0$, $y = -2$ ($Q = -1$), color singlet
- Left-handed up-type (u, c, t) quarks: $t = 1/2$, $t_3 = 1/2$, $y = 1/3$ ($Q = 2/3$), color triplet
- Left-handed down-type (d, s, b) quarks: $t = 1/2$, $t_3 = -1/2$, $y = 1/3$ ($Q = -1/3$), color triplet
- Right-handed up-type quarks: $t = 0$, $y = 4/3$ ($Q = 2/3$), color triplet
- Right-handed down-type quarks: $t = 0$, $y = -2/3$ ($Q = -1/3$), color triplet
- W^+: $t = 1$, $t_3 = 1$, $y = 0$ ($Q = 1$), color singlet
- W^-: $t = 1$, $t_3 = -1$, $y = 0$ ($Q = -1$), color singlet
- Z: $t = 1$, $t_3 = 0$, $y = 0$ ($Q = 0$), color singlet
- γ: $t = 0$, $y = 0$ ($Q = 0$), color singlet

- Gluon: $t = 0$, $y = 0$ ($Q = 0$), color octet

- Higgs: $t = 1/2$, $t_3 = \pm 1/2$, $y = 1$ ($Q = 0, +1$), color singlet

Note that the right-handed neutrinos have no charge, and participate in the gauge interactions only by neutrino oscillations, i. e., by admixtures due to the leptonic CKM matrix and their interaction with the Higgs boson. Any theory beyond the standard model has to reproduce this assignment.

It is now possible to discuss the Brout-Englert-Higgs effect in more detail. The complex doublet scalar Higgs-boson can be written as

$$H = \begin{pmatrix} \phi^0 \\ \phi^- \end{pmatrix} \tag{2.3}$$

and the Lagrangian for H takes the form

$$\mathcal{L}_H = (D_\mu H)^\dagger (D^\mu H) - V(H) \tag{2.4}$$

with some (renormalizable) potential V. To generate the masses in the standard model it must be assumed that (the quantum version of) the Higgs potential has an unstable extremum for $H = 0$ and a nontrivial minimum, e. g.

$$V(H) = \frac{\lambda}{2}(H^\dagger H - v^2)^2 \tag{2.5}$$

The Higgs boson then develops a vacuum expectation value v, the Higgs condensate. It is always possible to find a gauge, e. g. the 't Hooft gauge, in which v is real and oriented along the upper component, and thus to be annihilated by the electric charge to make it neutral,

$$\langle H \rangle = \begin{pmatrix} v \\ 0 \end{pmatrix}.$$

In the conventions used here, the value of v is $v = (2G_F)^{-1/2} \approx 250$ GeV, where G_F is Fermi's constant. Note that the operator Q defined by (2.2) acting on the Higgs vacuum expectation value yields zero, which implies that the condensate is uncharged, and this implies that the photon remains massless.

Inserting the decomposition of H into vacuum expectation value v and quantum fluctuations $h = H - v$ into (2.4) generates the masses of the weak gauge bosons as

$$\begin{aligned}
\mathcal{L}_{\text{mass}} &= 1/2(\partial h)^\dagger \partial h + 1/2M_W^2 W_\mu^+ W^{\mu -} + 1/2M_Z^2 Z_\mu Z^\mu - 1/2M_H^2 hh^\dagger \\
&\quad - \frac{\sqrt{\lambda}}{2} M_H (h^2 h^\dagger + (h^\dagger)^2 h) - \frac{1}{8}\lambda (hh^\dagger)^2 \\
&\quad + 1/2 \left(hh^+ + \frac{M_H}{\lambda}(h + h^\dagger) \right) (g_i^2 W_\mu^+ W^{\mu -} + (g_h^2 + g_i^2) Z_\mu Z^\mu) \\
M_H &= v\sqrt{2\lambda} \\
M_W &= \frac{g_i v}{2} \\
M_Z &= \frac{v}{2}\sqrt{g_2^2 + g_1^2} = \frac{M_W}{\cos\theta_W}.
\end{aligned}$$

Here, the electromagnetic interaction has been dropped for clarity. This Lagrangian also exhibits the coupling of the Higgs h field to itself and to the W and Z fields. It implies that the Higgs mass is just a rewriting of the four-Higgs coupling, and either has to be measured to fix the other.

The matter fields couple with maximal parity violation to the weak gauge fields, i. e. their covariant derivatives have the form, for, e. g. the left-handed weak isospin doublet of bottom and top quark $\Psi_L = (t,\ b)_L$

$$\begin{aligned}
\bar{\Psi}_L i\gamma^\mu D_\mu \Psi_L &= \bar{\Psi}_L i\gamma^\mu \partial_\mu \Psi_L - \frac{1}{\sqrt{2}}\bar{t}\gamma_\mu \frac{1-\gamma_5}{2} b W^{\mu +} - \frac{1}{\sqrt{2}}\bar{b}\gamma_\mu \frac{1-\gamma_5}{2} t W^{\mu -} \\
&\quad - \frac{2e}{3}\bar{t}\gamma^\mu \frac{1-\gamma_5}{2} t A_\mu + \frac{e}{3}\bar{b}\gamma^\mu \frac{1-\gamma_5}{2} b A_\mu - \bar{\Psi}_L e \tan\theta \gamma^\mu \Psi_L Z_\mu,
\end{aligned}$$

The problem with a conventional mass term would be that it contains the combination $\bar{\Psi}_L \Psi_R$, with Ψ_R being the sum of the right-handed bottom and top, which is not a singlet under weak isospin transformation, and thus would make the Lagrangian gauge-dependent, yielding a theory which is not physical.

This can be remedied by the addition of an interaction between the fermions and the Higgs of the Yukawa form

$$g_t \bar{\Psi}_L \cdot H t_R + g_b \bar{\Psi}_L \cdot H^\dagger b_R, \tag{2.6}$$

where $\cdot$ indicates a scalar product in isospin space, and which couples the left-handed fermions and right-handed fermions to the Higgs field. This combination is gauge-invariant and physically sensible for arbitrary Yukawa couplings g_b and g_t. When the Higgs develops its vacuum expectation value, masses $m_t = g_t v$ and $m_b = g_b v$ arise for the top and bottom quarks, respectively. This mechanism is replicated for both the other quarks and all leptons, though one of the neutrinos may remain massless without contradiction.

It should be noted that (2.6) can, in general, contain also off-diagonal terms, i. e. terms mixing different flavors. In the standard representation, the quark and lepton fields have been rotated such that they do not appear. The price to be payed is the appearance of the CKM and PMNS matrices in the weak interaction. However, otherwise a fixed flavor would not have a fixed mass. The consequence of this are oscillation phenomena. Note also that thus intrageneration effects, including CP violation, thereby originate from the Higgs-Yukawa interaction, and not from the weak interaction.

Another interesting feature of the Higgs sector is the fact that (2.3) has actually more than the minimum necessary number of degrees of freedom for a sensible theory. In principle, two degrees of freedom would be sufficient for a consistent theory. However, then there would be not enough degrees of freedom to make all three gauge bosons, $W^\pm$ and Z, massive simulatenously, and thus three or more are required by phenomenology. Theoretical consistency then requires at least four, and thus twice as many. Since these two sets of degrees of freedom are not distinguished by the weak interaction this gives rise to an additional SU(2) symmetry, the so-called custodial symmetry. This symmetry implies that the $W^\pm$ and Z would be mass-degenerate in absence of QED. QED, and also the Yukawa interactions (2.6), break this symmetry explicitly. In fact, QED is nothing but gauging the U(1) subgroup of the SU(2) custodial symmetry. While the symmetry is thus not manifest in the standard model, its original structure is still imprinted as an explicitly broken symmetry, which has to be replicated in one way or the other by any extension of the standard model.

2.1.3 Running coupling

There is a further important concept in the standard model, and actually in all quantum field theories, which will be a recurring theme in the search for beyond-the-standard model physics. This is the running of a coupling, or, more generically, the running of a quantity, known from the renormalization program in qunatum field theory. The derivative for a coupling with respect to the renormalization scale defines the β function as

$$\frac{dg}{d\ln\mu} = \beta(g) = -\beta_0\frac{g^3}{16\pi^2} + \mathcal{O}(g^5), \tag{2.7}$$

where the last equality defines the perturbative expansion with the β function coefficients β_i at order i of perturbation theory. Integrating this equation in leading order perturbation theory yields

$$\alpha(q^2) = \frac{g(q^2)^2}{4\pi} = \frac{\alpha(\mu^2)}{1+\frac{\alpha(\mu^2)}{4\pi}\beta_0\ln\frac{q^2}{\mu^2}} \equiv \frac{4\pi}{\beta_0\ln\frac{q^2}{\Lambda^2}}, \tag{2.8}$$

which introduces the scale Λ of a theory as a boundary condition of the ordinary differential equation (2.7). The value of Λ can be determined, e. g., by evaluating $g(\mu)$ perturbatively to this order. It plays the role of a characteristic scale of the theory in question. The value of β_0 depends on the theory under scrutiny, as well as the type and representation of the matter fields which couple to the interaction in question, e. g. for a gauge theory with gauge group G including fermions and Higgs fields in the fundamental representation it is

$$\beta_0 = \frac{11}{3}C_A - \frac{2}{3}N_f - \frac{1}{6}N_H \tag{2.9}$$

where C_A is the adjoint Casimir of the group, and N_f and N_H counts the number of fermion and Higgs flavors, respectively, which are charged in the fundamental representation of the gauge group. Plugging this in for the standard model, the values of β_0 for the strong interactions, the weak isospin, and the hypercharge are 7, 19/6, and -41/6, respectively, if the Higgs effect and all masses are neglected, i. e., at very high energies, $q^2 \gg 250$ GeV. Remapping this to the weak interactions and electromagnetism is only shifting the respective value for the weak interactions and the hypercharge weakly.

Plugging these values in (2.8) implies that for a positive β_0 the coupling decreases with increasing energy, while it increases for a negative value of β_0. The former behavior is known as asymptotic freedom. The latter, in contrast, yields eventually a singularity at high energies, called a Landau pole. This may indicate the breakdown of the theory, or merely the inadequacy of perturbation theory at high energies. If $g(\mu \to \infty)$ becomes a non-zero constants, this is referred to as asymptotic safety and to be discussed in more detail in section 7.6.

Similar equations like (2.9) actually hold also for all other parameters in a quantum field theory, in particular masses. Rather generically, the masses of the particles all decrease when increasing the measured momenta. Thus, the masses of particles become less and less relevant the higher the energy. That is a very important feature in many scenarios which rely on symmetries only broken by such mass terms, called a soft breaking.

2.1.4 Anomalies

There is one particular important property of the standard model, which is very much restricting its structure, and which is recurring in extensions of the standard model. That is the absence of anomalies. An anomaly is that some symmetry, which is present on the classical level, is not present when considering the quantum theory. The symmetry is said to be broken by quantum effects. Generically, this occurs if the action of a theory is invariant under a symmetry, but the measure of the path integral is not. Constructing a theory which is at the same time anomaly-free and consistent with the standard model is actually already quite restricting, and therefore anomalies are an important tool to check the consistency of new proposals for physics beyond the standard model. This will be therefore discussed here in some detail.

Anomalies fall into two classes, global and local anomalies. Global anomalies refer to the breaking of global symmetries by quantum effects. The most important one of these global anomalies is the breaking of dilatation symmetry. This symmetry corresponds to rescaling all dimensionful quantities, e. g., $x \to \lambda x$. Maxwell theory, massless QED, Yang-Mills theory, and massless QCD are all invariant under such a rescaling, at the classical level, though not the Higgs sector of the standard model. This is no longer the case at the quantum level. By a process called dimensional transmutation, surfacing in the renormalization process, an explicit scale is introduced into the theory, and thereby the quantum theory is no longer scale-invariant. Such global anomalies have very direct consequences. E. g., this dilatation anomaly leads to the fact that the photon is massless in massless QED. Another example is the so-called axial anomaly, which occurs due to the breaking of the global axial symmetry of baryons. A consequence of it is the anomalously large η' mass.

In contrast to the global anomalies, the local anomalies are a more severe problem. A local anomaly occurs, when a quantum effect breaks a local gauge symmetry. The consequence of this would be that observable quantities depend on the gauge, and therefore the theory makes no sense. Thus, such anomalies may not occur. There are two possibilities how such anomalies can be avoided. One is that no such anomalies occurs, i. e., the path integral measure must be invariant under the symmetry. The second is by anomaly cancellation, i. e., some parts of the measure are not invariant under the symmetry, but the sum of all such anomalous terms cancel. It is the latter mechanism which makes the standard model anomaly-free. However, the price to pay for this is that the matter content of the standard model has to follow certain rules. It is thus rather important to understand how this comes about. Furthermore, any chiral gauge theory beyond the standard model faces similar, or even more severe, problems.

Eventually, see the lecture on quantum field theory II, the requirement for the absence of gauge anomalies boils down to the condition

$$\mathrm{tr}\left\{\tau_L^a, \tau_L^b\right\}\tau_L^c - \mathrm{tr}\left\{\tau_R^a, \tau_R^b\right\}\tau_R^c = 0,$$

where τ are the generators of the theorie's total gauge group, and L and R indicate the representation of the left-handed fermions and right-handed fermions, respectively. There are now two possibilities how to obtain an anomaly-free theory. Either, the theory is anomaly-free, if each of the terms is individually zero, or they cancel. Indeed, the expression $\mathrm{tr}\{\tau^a, \tau^b\}\tau^c$, the so-called symmetric structure constant, is zero for all (semi-)simple Lie groups, except for $\mathrm{SU}(N \geq 3)$ and $\mathrm{U}(1)$. Unfortunately, these are precisely those appearing in the standard model, except for the SU(2) of weak isospin. For the group SU(3) of QCD, this is actually not a problem, since QCD is vectorial, and thus[6] $\tau_L = \tau_R$, and the terms cancel for each flavor individually. Thus remains only the part induced by the hypercharge.

In this case, each generation represents an identical contribution to the total result, as the generations are just identical copies concerning the generators. It is thus sufficient to consider one generation. The right-handed contributions are all singlets under the weak isospin, and thus they only couple vectorially to electromagnetism, and therefore yield zero. The contributions from the left-handed doublet contain then the generators of the weak isospin, τ^a, and the electric charge $Q = \tau^3 + \mathbf{1}y/2$. The possible combinations contributing are

$$\mathrm{tr}\tau^a\{\tau^b, \tau^c\} \tag{2.10}$$

$$\mathrm{tr}Q\{\tau^a, \tau^b\} \tag{2.11}$$

$$\mathrm{tr}\tau^a Q^2 \tag{2.12}$$

$$\mathrm{tr}Q^3. \tag{2.13}$$

The contribution (2.10) vanishes, as this is a pure SU(2) expression. The term (2.13) is not making a difference between left and right, and is therefore also vanishing. It turns out that (2.11) and (2.12) lead to the same result, so it is sufficient to investigate (2.12). Since the isospin group is SU(2), the anti-commutator of two Pauli matrices just gives a Kronecker-δ times a constant, yielding in total

$$\mathrm{tr}Q\{\tau^a, \tau^b\} = \frac{1}{2}\delta^{ab}\sum_f Q_f,$$

where Q_f is the electric charge of the member f of the generation in units of the electric charge. It has to vanish to prevent any gauge anomaly in the standard model, which is fulfilled:

$$\sum_f Q_f = (0-1) + N_c\left(\frac{2}{3} - \frac{1}{3}\right) = -1 + \frac{N_c}{3} = 0.$$

Therefore, there is no gauge anomaly in the standard model. However, this is only possible, because the electric charges have certain ratios, and the number of colors N_c is three. This implies that the different sectors of the standard model, the weak isospin, the strong interactions, and electromagnetism, very carefully balance each other, to provide a well-defined theory. Such a perfect combination is one of the reasons to believe that the

[6] Actually, unitarily equivalent is sufficient.

standard model is part of a larger theory, which imposes this structure. This leads to the concept of grand-unified theories in chapter 7.

There is actually a further possible anomaly, the so-called Witten anomaly, which comes from the parity violation. This is a problem exclusively applying to the Sp(N) gauge groups, and to SU(2) of the weak interactions because SU(2)≈Sp(1), as well as O($N < 6$) groups, except for SO(2), if the number of chiral (Weyl) fermions is not even. It will not be detailed here, see again the lecture on quantum field theory II. In the standard model, it is canceled because the number of weak fermion states is even. This would not be the case, if, e. g., there would be a single triplet of fermions charged under the weak isospin. In technicolor theories, to be discussed in chapter 5, this is a constraint, as in such theories multiplets with an odd number of fermions may appear.

2.2 General relativity

As will be discussed later, one of the objectives of many proposals for physics beyond the standard model is to include a quantized version of gravity. Therefore, here quickly the basics of gravity necessary in the following will be repeated[7]. The basic ingredient will be the local metric $g_{\mu\nu}(x)$, which will later often be split as

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x)$$

where $\eta_{\mu\nu}$ is the constant Minkowski metric around which the quantum corrections to the metric $h_{\mu\nu}$ fluctuate. Both classically and quantum, the metric describes the invariant length-element ds by

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu.$$

The inverse of the metric is required to exist and is given by the contravariant tensor $g^{\mu\nu}$,

$$g^{\mu\nu} g_{\nu\lambda} = \delta^\mu_\lambda.$$

As a consequence, for any derivative δ of $g_{\mu\nu}$

$$\delta g^{\mu\nu} = -g^{\mu\lambda} g^{\nu\rho} \delta g_{\lambda\rho} \tag{2.14}$$

holds. The metric is assumed to be non-vanishing and symmetric and has a signature such that its determinant is negative,

$$g = \det g_{\mu\nu} < 0.$$

[7] A more detailed introduction is given in the lecture on advanced general relativity and quantum gravity.

The covariant volume element dV is therefore given by

$$\begin{aligned} dV &= \omega d^4x \\ \omega &= \sqrt{-g} = \sqrt{-\det g_{\mu\nu}} > 0, \end{aligned}$$

implying that ω is real (hermitian), and has derivative

$$\delta\omega = \frac{1}{2}\omega g^{\mu\nu}\delta g_{\mu\nu} = -\frac{1}{2}\omega g_{\mu\nu}\delta g^{\mu\nu} \tag{2.15}$$

as a consequence of (2.14).

The most important concept of general relativity is the covariance (or invariance) under a general coordinate transformation $x_\mu \to x'_\mu$ (diffeomorphism) having

$$\begin{aligned} dx'^\mu &= \frac{\partial x'^\mu}{\partial x^\nu}dx^\nu = J^\mu_\nu dx^\nu \\ \det(J) &\neq 0, \end{aligned}$$

where the condition on the Jacobian J follows directly from the requirement to have an invertible coordinate transformation everywhere. Scalars $\phi(x)$ are invariant under such coordinate transformations, i. e., $\phi(x) \to \phi(x')$. Covariant and contravariant tensors of n-th order transform as

$$\begin{aligned} T'_{\mu...\nu}(x') &= \frac{\partial x_\mu}{\partial x'_\alpha}...\frac{\partial x_\nu}{\partial x'_\beta}T_{\alpha...\beta}(x) \\ T'^{\mu...\nu}(x') &= \frac{\partial x'^\mu}{\partial x^\alpha}...\frac{\partial x'^\nu}{\partial x^\beta}T^{\alpha...\beta}(x) \end{aligned} \tag{2.16}$$

respectively, and contravariant and covariant indices can be exchanged with a metric factor, as in special relativity. As a consequence, the ordinary derivative ∂_μ of a tensor A_ν of rank one or higher is not a tensor. To obtain a tensor from a differentiation the covariant derivative must be used

$$\begin{aligned} D_\mu A_\nu &= \partial_\mu A_\nu - \Gamma^\lambda_{\mu\nu}A_\lambda \\ \Gamma^\lambda_{\mu\nu} &= \frac{1}{2}g^{\lambda\sigma}(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}), \end{aligned} \tag{2.17}$$

where Γ are the Christoffelsymbols. Only the combination ωA_ν, yielding a tensor density, obeys

$$D_\mu(\omega A_\nu) = \partial_\mu(\omega A_\nu).$$

As a consequence, covariant derivatives no longer commute, and their commutator is given by the Riemann tensor $R_{\lambda\rho\mu\nu}$ as

$$\begin{aligned} [D_\mu, D_\nu] A^\lambda &= R^\lambda_{\rho\mu\nu} A^\rho \\ R^\lambda_{\rho\mu\nu} &= \partial_\mu \Gamma^\lambda_{\nu\rho} - \partial_\nu \Gamma^\lambda_{\nu\rho} + \Gamma^\lambda_{\mu\sigma}\Gamma^\sigma_{\nu\rho} - \Gamma^\lambda_{\nu\rho}\Gamma^\sigma_{\mu\rho}, \end{aligned}$$

which also determines the Ricci tensor and the curvature scalar

$$\begin{aligned} R_{\mu\nu} &= R^\lambda_{\nu\mu\lambda} \\ R &= R^\mu_\mu, \end{aligned}$$

respectively.

These definitions are sufficient to write down the basic dynamical equation of general relativity, the Einstein equation

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + g_{\mu\nu}\Lambda = -\kappa T_{\mu\nu}, \tag{2.18}$$

which can be derived as the Euler-Lagrange equation from the Lagrangian[8]

$$\mathcal{L} = \omega \left(\frac{1}{2\kappa} R - \frac{1}{\kappa}\Lambda + \mathcal{L}_M \right), \tag{2.19}$$

where the first two terms are the Einstein-Hilbert Lagrangian $\mathcal{L}_{EH}$. The quantity Λ is the cosmological constant, a parameter of the theory, which is measured to be small but non-zero. $\mathcal{L}_M$ is the matter Lagrangian yielding the covariantly conserved energy momentum tensor $T_{\mu\nu}$, which is the variation of $\mathcal{L}_M$ with respect to the metric. In flat-space time it becomes the usual one,

$$T_{\mu\nu} = \left(-\eta_{\mu\nu}\mathcal{L}_M + 2\frac{\delta \mathcal{L}_M}{\delta g^{\mu\nu}} (g_{\mu\nu} = \eta_{\mu\nu}) \right). \tag{2.20}$$

The second constant $\kappa = 16\pi G_N$ is Newton's constant. It describes the strength of the coupling of matter to gravity.

There is an important remark to be made about classical general relativity. The possibility of making a general coordinate transformation leaving physics invariant has the consequence that coordinates, and thus also both energy and three momentum as their canonical conjugate momenta, loose their meaning as physically meaningful concepts, just like charge in a non-Abelian gauge theory. Indeed it is possible to alter the energy of a system by performing a space-time coordinate transformation. Only the concept of total energy (or momentum) of a localized distribution of particles when regarded from far

[8] In the following, usually, the cosmological constant term $g_{\mu\nu}\Lambda$ will be absorbed in the matter part.

away in an otherwise flat space-time can be given an (approximate) physically meaning, similarly to charges. Therefore, many concepts which are usually taken to be physical loose their meaning when general relativity is involved. This carries over to any quantum version.

Chapter 3

Why physics beyond the standard model?

Before discussing actual BSM scenarios, it is useful to understand why they appear necessary and how they could be discovered.

There are a number of reasons to believe that there exists physics beyond the standard model. These reasons can be categorized as being from within the standard model, by the existence of gravity, and by observations which do not fit into the standard model. Essentially all of the latter category are from astronomical observations, and there are currently only very few observations in terrestrial experiments which are reproducible and do not perfectly agree with the standard model, and none which disagree with any reasonable statistical and systematical accuracy.

Of course, it should always be kept in mind that the standard model has never been completely solved. Though what has been solved, in particular using perturbation theory, agrees excellently with measurements, it is a highly non-linear theory. It cannot a-priori be excluded that some of the reasons to be listed here are actually completely within the standard model, once it is possible to solve it exactly.

Many of the observations to be listed can be explained easily, but not necessarily, by new physics at a scale of 1 TeV. However, it cannot be exclude that there is no new phenomena necessary for any of them up to a scale of 10^{15} GeV, called the GUT scale for reasons to become clear in chapter 7, or possibly up to the Planck scale of 10^{19} GeV, in which case the energy domain between the standard model and this scale is known as the great desert.

3.1 Inconsistencies of the standard model

There are a number of factual and perceived flaws of the standard model, which make it likely that it cannot be the ultimate theory.

The one most striking reason is the need of renormalization. It is not possible to determine within the standard model processes at arbitrary high energies. The corresponding calculations break down eventually, and yield infinities. Though we have learned how to absorb this lack of knowledge in a few parameters, the renormalization constants, it is clear that there are things the theory cannot describe. Thus it seems plausible that at some energy scale these infinities are resolved by new processes, which are unknown so far. In this sense, the standard model is often referred to as an effective low-energy theory of the corresponding high-energy theory, or sometimes also called ultraviolet completion.

This sought-for high-energy theory is very likely not a (conventional) quantum field theory, as this flaw is a characteristic of such theories. Though theories exist which reduce the severity of the problem, supersymmetry at the forefront of them, it appears that it is not possible to completely resolve it for any theory compatible with observations[1], though this cannot be excluded. Thus, it is commonly believed that the high-energy theory is structurally different from the standard model, like string theory to be discussed in chapter 9.1.

In a similar vain, there is also a very fundamental question concerning the Higgs sector. At the current time, it is not yet clear whether there can exist, even in the limited sense of a renormalizable quantum field theory, a meaningful theory of an interacting scalar field. This is the so-called triviality problem. So far, it is essentially only clear that the only consistent four-dimensional theory describing a spin zero boson alone is one without any interactions. Whether this can be changed by adding additional fields, as in the standard model, is an open question. However, since this problem can be postponed to energy scales as high as 10^{15} GeV, or possibly even higher, this question is not necessarily of practical relevance.

There are a number of aesthetic flaws of the standard model as well. First, there are about thirty five different free parameters of the theory, varying by at least twelve orders of magnitude. There is no possibility to understand their size or nature within the standard model, and this is unsatisfactory. Even if their origin alone could be understood, their relative size is a mystery as well. This is particularly true in case of the Higgs and the electroweak sector in general. There is no reason for the Higgs to have a mass which is small compared to the scale of the theory from which the standard model emerges. In

[1]Some conformal theories or lower-dimensional theories do not need renormalization, they are intrinsically finite.

particular, no symmetry protects the Higgs mass from the underlying theory, which could make it much more massive, and therefore inconsistent with experimental data, than all the other standard model particles. Why this is not so is called the hierarchy problem, despite the fact that it could just be accidentally so, and not a flaw of the theory. Even if this scale should be of the order of a few tens of TeV, there is still a factor of possibly 100 involved, which is not as dramatic as if the scale would be, say, 10^{15} GeV. Therefore, this case is also called the little hierarchy problem.

There is another strikingly odd thing with these parameters. The charges of the leptons and quarks need not to be the same just because of QED - in contrast to the weak or strong charge, actually. They could differ, and in particular do not need to have the ratio of small integer numbers as they do: -1 to $2/3$ or $1/3$. This is due to the fact that the gauge group of QED is Abelian. However, if they would not match within experimental precision, which is more than ten orders of magnitude, then actually the standard model would not work, and neither would physics with neutral atoms. This is due to the development of a quantum anomaly, i. e., an effect solely related to quantizing the theory which would make it inconsistent, as discussed in section 2.1.4. Only with the generation structure of quarks and leptons with the assigned charges of the standard model this can be avoided. This is ultimately mysterious, and no explanation exists for this in the standard model, except that it is necessary for it to work, which is unsatisfactory.

There is also absolutely no reason inside the standard model why there should be more than one family, since the aforementioned cancellation works within each family independently and is complete. However, at least three families are necessary to have inside the standard model CP violating processes, i. e. processes which favor matter over anti-matter. As will be discussed in section 3.4.4, such processes are necessary for the observation that the world around us is made from matter. But there is no reason, why there should only be three families, and not four, five, or more. And if there should be a fourth family around, why its neutrino is so heavy compared to the other ones, as can already be inferred from existing experimental data.

In this context it is also important that it is not yet clear whether the ground state of the standard model is actually the world we are living in, or whether this is just a metastable state which could collapse at some point in the future to the true ground state, with potentially catastrophic consequences. Finding the answer to this question is currently primarily a question of computability, but it is already clear that the answer is sensitive to the mass ratio of the Higgs to the top quark, two not related quantities in the standard model. This makes it again suspicious why these two numbers should be so close. This is known as the electroweak stability problem.

Finally, when extrapolating the running gauge couplings to an energy scale of about 10^{15} GeV their values almost meet, suggesting that at this scale a kind of unification could be possible. However, they do not meet exactly, and this is somewhat puzzling as well. Why should this seem to appear almost, but not perfectly so?

Though often referred to as beyond the standard model, the conventional realization of neutrino oscillations can be accommodated in the standard model just by making them Dirac fermions and the introduction of parameters for their masses, and a second CKM-matrix in the lepton sector, the PMNS matrix. This will therefore not be considered beyond the standard model for the scope of this lecture. This does not explain why their masses are several orders of magnitude smaller than all the other fermions masses nor why the PMNS matrix is so different from the CKM matrix, and this will be a subject of this lecture.

Also, questions of computability, in particular within perturbation theory, are deemed here to be completely irrelevant in this lecture, since its nature and not our ability to compute something which decides about physics. Thus, especially the concept of perturbativity, i. e. the demand that the theory is readily accessible to perturbative calculations, will not be considered as a valid constrain for anything.

3.2 Gravity

3.2.1 Problems with quantization

One obviously, and suspiciously, lacking element of the standard model is gravity. Up to now no consistent quantization of gravity has been obtained beyond reasonable doubt. Usually the problem is that a canonical quantized theory of gravity is not renormalizable perturbatively. This is visible when writing down the Lagrangian of gravity (2.19): The coupling constant involved, κ or equivalently Newton's constant, is dimensionful. Superficial (perturbative) power counting immediately implies that the theory is perturbatively non-renormalizable. As a consequence, an infinite hierarchy of counter terms, all to be fixed by experiment, would be necessary to perform perturbative calculations, spoiling any predictivity of the theory. In pure gravity, these problems occur at two-loop order, for matter coupled to gravity already at the leading order of radiative corrections.

In particular, this implies that the theory is not reliable beyond the scale $\sqrt{\kappa}$. Though this may be an artifact of perturbation theory, this has led to developments like super quantum gravity based on local supersymmetry or loop quantum gravity.

Irrespective of the details, the lack of gravity is an obvious flaw of the standard model.

Along with this lack comes also a riddle. The natural scale of quantum gravity is given by the Planck scale

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

This is 17 orders of magnitude larger than the natural scale of the electroweak interactions, and 19 orders of magnitude larger than the one of QCD. The origin of this mismatch is yet unsolved, and also known as (a) hierarchy problem

One of the most popular explanations, discussed in detail in chapter 8, is that this is only an apparent mismatch: The scales of gravity and the standard model are the same, but gravity is able to propagate also in additional dimensions not accessible by the remainder of the standard model. The mismatch comes from the ratio of the total volumes, the bulk, an the apparent four dimensional volume, which is thus only a boundary, a so-called brane.

3.2.2 Asymptotic safety

Reiterating, the problem with the renormalizability of quantum gravity is a purely perturbative statement, since only perturbative arguments have been used to establish it. Thus, the possibility remains that the theory is not having such a problem, it is said to be asymptotically safe, and the problem is a mere artifact of perturbation theory. In this case, when performing a proper, non-perturbative calculation, no such problems would arise. In fact, this includes the possibility that κ imposes just an intrinsic cutoff of physics, and that this is simply the highest attainable energy, similarly as the speed of light is the maximum velocity. As a consequence, the divergences encountered in particle physics then only results from taking the improper limit energy$\to \infty \gg \kappa$.

This concept of asymptotic safety can be illustrated by the use of a running coupling, this time the one of quantum gravity. The naive perturbative picture implies that the running gravitational coupling increases without bounds if the energy is increased, similarly to the case of QCD if the energy is decreased: The theory hits a Landau pole. Since the theory is non-linearly coupled, an increasing coupling will back-couple to itself, and therefore may limit its own growth, leading to a saturation at large energies, and thus becomes finite. This makes the theory then perfectly stable and well-behaved. However, such a non-linear back-coupling cannot be captured naturally by perturbation theory, which is a small-field expansion, and thus linear in nature. It thus fails in the same way as it fails at small energies for QCD. Non-perturbative methods, like renormalization-group methods or numerical simulations, have provided indication that indeed such a thing may happen in quantum gravity, though this requires further confirmation.

As an aside, it has also been proposed that a similar solution may resolve both the hierarchy problem and the triviality problem of the Higgs sector of the standard model, when applied to the combination of Higgs self-coupling and the Yukawa couplings, and possibly the gauge couplings.

This scenario will be discussed in more detail in section 7.6.

3.3 Observations from particle physics experiments

There are two generic types of particle physics experiments to search for physics beyond the standard model, both based on the direct interaction of elementary particles. One are those at very high energies, where the sheer violence of the interactions are expected to produce new particles, which can then be measured. The others are very precise low-energy measurements, where very small deviations from the standard model are attempted to be detected. Neither of these methods has provided so far any statistically and systematically robust observation of a deviation from the standard model. Indeed, it has happened quite often that a promising effect vanishes when the statistical accuracy is increased. Also, it has happened that certain effects have only been observed in some, but not all, of conceptually similar experiments. In these cases, it can again be a statistical effect, or there is always the possibilities that some, at first glance, minor difference between both experiments can fake such an effect at one experiment, or can shadow it at the other. So far, the experience was mostly that in such situation a signal was faked, but this then usually involves are very tedious and long search for the cause.

At the time of writing, while almost daily new results are coming in, there are few remarkable results, and which await further scrutiny. The two most prominent are the muon $g-2$ and lepton flavor universality violation. The first originates from the fact that the measured value of anomalous magnetic moment of the muon differs from the one expected in the standard model in experiments. The other refers to the fact that the decays of bottom quarks to leptons does not happen, up to trivial mass effects, at the same rate into different types of leptons. Both effects have been seen at a less than fully convincing statistical accuracy in experiments, and currently larger efforts are undertaken to increase the statistics.

On the other hand, both effects are dominated by hadronic, and thus theoretically hard to control, uncertainties. Once from hadronic vacuum fluctuations, and once from the structure of the meson into which the bottom quark is embedded. It is thus entirely possible that both will, as so often in the past, turn out to be just deficiencies in our ability to estimate the systematic errors of calculations.

But then, maybe not.

3.4 Astronomical observations

During the recent decades a number of cosmological observations have been made, which cannot be reconciled with the standard model. These will be discussed here.

3.4.1 Dark matter

One of the most striking observation is that the movement of galaxies, in particular how matter rotates around the center of galaxies, cannot be described just by the luminous matter seen in them and general relativity. That is actually a quite old problem, and known since the early 1930s. Also gravitational lensing, the properties of hot intergalactic clouds in galaxy clusters, the evolution of galaxy clusters and the properties of the large-scale structures in the universe all support this finding. In fact, most of the mass must be in the form of invisible dark matter. This matter is concentrated in the halo of galaxies, as analyses of the rotation curves show. This matter cannot just be due to non-self-luminous objects like planets, brown dwarfs, cold matter clouds, or black holes, as the necessary density of such objects would turn up in a cloaking of extragalactic light and of light from globular clusters. This matter is therefore not made out of any conventional objects, in particular, it is non-baryonic. Furthermore, it is gravitational but not electromagnetically active. It also shows different fluid dynamics (observed in the case of colliding galaxies) as ordinary luminous matter. Also, the dark matter cannot be strongly interacting, as it otherwise would turn up as bound in nuclei.

Thus this matter has to have particular properties. The only particle in the standard model which could have provided it would have been a massive neutrino. However, though the neutrinos do have mass, the upper limits on their mass is so low, and the flux of cosmic neutrinos too small, to make up even a considerable fraction of the dark matter. This can be seen by a simple estimate. If the neutrinos have mass and would fill the galaxy up to the maximum possible by Fermi-statistics, their density would be

$$n_\nu = \frac{p_F^3}{\pi^2}$$

with the Fermi momentum p_F in the non-relativistic case given by $m_\nu v_\nu$. Since neutrinos have to be bound gravitationally to the galaxy, their speed is linked via the Virial theorem to their potential energy

$$v_\nu^2 = \frac{G_N M_{\text{galaxy}}}{R},$$

with Newton's constant G_N and R the radius of the galaxy. Putting in the known numbers, and using furthermore that the observational results imply that n_ν, the total number of neutrinos approximated to be inside a sphere size the galaxy, must give a total mass larger than the one of the galaxy leads to the bound

$$m_\nu > 100\text{ eV}\left(\frac{0.001c}{3v_\nu}\right)^{\frac{1}{4}}\left(\frac{1\text{ kpc}}{R}\right)^{\frac{1}{2}},$$

yielding even for a neutrino at the speed of light a lower bound for the mass of about 3 eV, which is excluded by direct measurements in tritium decays.

Therefore, a different type of particles is necessary to fill this gap. In fact, many theories offer candidates for such particles, in particular supersymmetry. But so far none has been detected, despite several dedicated experimental searches for dark matter. These proceed either by trying to produce them in high-energetic collisions or by searching them from astronomical sources using highly sensitive detectors of a wide variety of techniques, including underground detectors and satellites.

These yielded only very few candidates for an observation of dark matter particles, and those are hard to distinguish from background, in particular natural radioactivity and cosmic rays. Though, every once in a while, satellites find excesses in cosmic rays which seem to hint for signals of dark matter annihilation, but so far none of these has survived further scrutiny. The origin of dark matter stays therefore mysterious.

But not only the existence of dark matter, also its properties are surprising. The observations are best explained by dark matter which is in thermal equilibrium. But how this should be achieved if it is really so weakly interacting is unclear. The best guess so far is that it is more strongly interacting with itself than with ordinary matter and/or consists out of more than a single particle type.

On the other hand, the fact that dark matters needs to interact gravitationally is also posing problems, not only a solution. In particular, there is no reason why it should neither form celestial dark bodies, which should be observable by passing in front of luminous matter, or why it should not be bound partly in the planets of our solar system, or other celestial bodies. Only if it is temperature is so high that binding is prohibited this would be in agreement, but then the question remains why it is so hot, and what is the origin of the enormous amount of energy stored in the dark matter.

It should be noted that there are also attempts to explain these observations by a departure of gravity from its classical behavior also at long distances. Though parametrizations exist of such a modification, often called modified Newtonian dynamics, or MOND, which are compatible with observational data, no clean explanation or necessity for such a modification in classical general relativity has been established. This proposal is also challenged by observations of colliding galaxies which show that the center-of-mass of the total matter and the center of luminous matter move differently, which precludes any simple modification of the laws of gravity, and is much more in-line with the existence of dark matter. In the same vain, some dwarf galaxies have become candidates to have vastly different amounts, in either direction, of dark matter and ordinary matter than other galaxies. Still, this cannot be excluded yet. In this class of solutions falls also the possibility that asymptotic safety of quantum gravity may be related to the apparent existence of dark matter.

3.4.2 Inflation

A second problem is the apparent smoothness of the universe around us, while having at the same time small highly non-smooth patches, like galaxies, clusters, super clusters, walls and voids. In the standard model of cosmological evolution this can only be obtained by a rapid phase of expansion (by a factor $\sim e^{60}$) of the early universe, at temperatures much larger than the standard model scale, but much less than the gravity scale. This is called inflation. During the inflationary period, space-time itself expanded at superluminal velocities, which is not in contradiction to general relativity. Therefore, large parts of matter, which equilibrated beforehand, were no longer causally connected, but still maintained their common equilibrium. Only afterwards they started to develop differently, leading to the small regions of inhomogeneities.

Also the standard model can create such periods of inflation, especially the elctroweak and strong crossovers/phase transitions. But they occurred far too late in the evolution of the universe, and could not sustain more than a factor of perhaps $e^4 - e^5$ expansion. Thus, none of the standard model physics can explain inflation, nor act as an agitator for it. In particular, it is also very complicated to find a model which at the same time explains the appearance of inflation and also its end after just the right amount.

However, the predictions of inflation have been very well confirmed by the investigation of the cosmic microwave background radiation, including non-trivial features and up to rather high precision. They also are important for the curvature of the universe to be discussed next.

3.4.3 Curvature and cosmic expansion

Another problem is the apparent flatness of the universe. Over large scales, the angle sum of a triangle is observed to be indeed π. This is obtained from the cosmic microwave background radiation, in particular the position of the quadrupole moment[2], but also that the large-scale structure in the universe could not have been formed in the observed way otherwise. For a universe, which is governed by Einstein's equation of general relativity, this can only occur if there is a certain amount of energy inside it. Even including the unknown dark matter, the amount of registered mass can provide at best about 30% of the required amount to be in agreement with this observation. The other contribution, amounting to about 70%, of what origin it may ever be, is called dark energy. Even then, the extreme flatness of the universe also requires an inflationary period to be possible.

A second part of the puzzle is that the cosmic expansion is found to be accelerating. This is found from distant supernova data, which are only consistent if the universe expands accelerated today. In particular, other explanations are very hard to reconcile with the data, as it behaves non-monotonous with distance, in contrast to any kind of light-screening from any known physical process. Furthermore, the large-scale structures of the universe indicate this expansion, but also that the universe would be too young (about 10.000.000.000 years) for its oldest stars (about 12-13.000.000.000 years) if this would not be the case. For such a flat universe such an acceleration within the framework of general relativity requires a non-zero cosmological constant Λ, which appears in the Einstein equations (2.18). This constant could also provide the remaining 70% of the mass to close the universe, and is in fact a (dark) vacuum energy. Such a constant is covariantly conserved, since both $T_{\mu\nu}$ and the first two terms in (2.18) together are independently in general relativity, and thus indeed constant. However, the known (quantum) effects contributing to such a constant provide a much too large value for Λ, about 10^{40} times too large. These include quantities like the chiral condensate and gluon condensates. These are of order GeV, and in addition would have the wrong sign. What leads to the necessary enormous suppression is unclear. Also, it is not clear whether this is a valid comparison, as this is a quantum effect. Thus, this kind of hierarchy problem may also be just a deficiency of the calculational tools.

Alternatively, weakly broken supersymmetry could remove this contribution, when a gluino and a squark condensate cancel essentially quark and gluon condensates. Unfortunately, supersymmetry broken sufficiently weakly to be in agreement with the observed value of the condensates generates in general super partners with masses too close to those of ordinary matter as that they could have escaped experimental detection. Only enormous fine-tuning, leading to another hierarchy problem, could prevent this.

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

3.4.4 Matter-antimatter asymmetry

In the standard model, matter and antimatter are not perfectly symmetric. Due to the CP violations of the electroweak forces, matter is preferred above antimatter, i. e., decays produce more matter than antimatter, and also baryon and lepton number are not independently conserved quantities, only their sum is. However, this process is dominantly non-perturbative. The most striking fact that this is a very weak effect is the half-life of the proton, which is (experimentally and theoretically) larger than 10^{34} years. Indeed, only at very high-temperature can the effect become relevant.

After the big-bang, the produced very hot and dense matter was formed essentially from a system of rapidly decaying and recombining particles. When the system cooled down, the stable bound states remained in this process, leading first to stable nucleons and leptons in the baryogenesis, and afterwards to stable nuclei and atoms in the nucleosynthesis. Only over this time matter could have become dominant over antimatter, leading to the stable universe observed today. But the electroweak effects would not have been strong enough for the available time to produce the almost perfect asymmetry of matter vs. antimatter observed today, by a factor of about 10^{19}. Thus, a further mechanism must exist which provides matter dominance today.

There is a profound connection to inflation. It can be shown that inflation would not have been efficient enough, if the number of baryons would have been conserved in the process. In particular, the almost-baryon-number conserving electroweak interactions would have permitted only an inflationary growth of e^{4-5} instead of e^{60}.

The possibility that this violation is sufficient to create pockets of matter at least as large as our horizon, but not on larger scales, has been tested, and found to yield only pockets of matter much smaller than our horizon.

A further obstacle to a standard-model conform breaking of matter-antimatter symmetry is the necessity for a first order phase transition. This is required since in a equilibrium (or almost equilibrium like at a higher-order transition), the equilibration of matter vs. anti-matter counters the necessary breaking. However, the mass of the standard-model Higgs is too high for this.

3.5 Why one TeV?

There is the common expectation that something of these new theories will show up at an energy scale of one TeV or slightly above. That the Tevatron has not seen anything of this is actually not surprising. Since it collides protons and anti-protons, the actually interacting partons, quarks and gluons, have almost always significantly less energy than the maximum energy. So, it is up to the LHC to explore this energy range.

The gateway to this kind of new physics is likely the Higgs. The reason is that the Higgs is instrumental for the balancing in the standard model. If there is something just slightly different, it will most likely surface first in the Higgs sector. And the balancing becomes already quite sensitive to new effects at 1 TeV.

The simplest explanation why 1 TeV is such a crucial scale can be seen, e. g. by the scattering cross-section of two longitudinally polarized W bosons to two longitudinally polarized W bosons, a process which in the standard model will occur a-plenty at these energies. At tree-level, the scattering amplitude without the Higgs is given by

$$M_{WW} = M_2(\cos\theta)\frac{s}{m_W^2} + M_1(\cos\theta)\ln\frac{s}{m_W} + M_0(\cos\theta), \tag{3.1}$$

where s is the center-of-mass energy, θ the angle between the scattered W bosons, and the amplitudes M_i describe the processes of scattering different polarizations of the W bosons. Unitarity, and thus preservation of causality, requires that this amplitude is bounded for $s \to \infty$, which is obviously not the case for the terms containing M_2 and M_1. Thus, for a center-of-mass energy significantly larger than the W boson mass m_W (in fact, about a TeV), unitarity is violated. In the standard model, interference with diagrams containing the Higgs removes this problem. This is known as the Goldstone boson equivalence theorem. But if there are additional contributions, this will be slightly different. And then the linear dependency in s will magnify this effect. Of course, at sufficiently large s again unitarity has to be restored, but for some time there would be a quick and apparent deviation, which should be detectable in experiments.

Hence, all in all, though there is no guarantee that something interesting beyond a rather light Higgs has to happen at the TeV scale, there is quite some evidence in favor of it. Time will tell. And what this something may be, this lecture will try to give a glimpse of.

3.6 How can new physics be discovered?

A task at least as complicated as the theoretical description of new physics is its experimental observation. One of the objectives of theoretical studies is therefore to provide signals on which the experiments can focus on. Here, some of the more popular experimental signatures, which can be calculated theoretically, will be introduced.

3.6.1 Lessons from the past on reliability

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

To quantify the amount of statistics available, usually the number of events is quoted in inverse barn, i. e., as an inverse cross section. Consequently, if there is 10 fb^{-1}, a typical amount of data collected at hadronic colliders like the Tevatron or the LHC, implies that a process with a cross section of 0.1 fb will be observed in this data set once. The current aim for the LHC is, however, much larger, at about 3000 fb^{-1} until 2030, and more than 100 fb^{-1} delivered to date.

An important effect in this is the so-called 'look-elsewhere' effect. The amount of experimental measurements, especially using modern machine-learning techniques, has grown immensely into the thousands. Thus, it is statistically likely that a single measurement will show deviations at the evidence level, just due to the amount of statistical fluctuations in such a large set of measurements. Thus, the relevance of a statistical fluctuation is reduced by the fact that in some measurement a large statistical fluctuation is statistically expected. Therefore, an actual statement about reliability needs to take this into account. Hence, the statistical uncertainty of a single measurement is also called a local significance, while one which is taking into account the likelihood of finding a deviation in a large set of measurements is called global significance. Thus, new physics will require either a single very large local significance and a discovery-level global significance, or many local discovery-level significant measurements, as in the presence of many anomalies local and global significance approach each other.

An alternative way to present data is the so-called p-value, which recasts the significance into the probability to being a statistical fluctuation, essentially the total probability minus two times the tail amount at the 3/5 sigma level. Thus, the lower the p value the more probably something new has been found.

3.6.2 New particles

When in an interaction two particles exchange another particle, the cross-section of this process will in the s-channel in the lowest order be proportional to the square of the propagator D of the exchanged particle, i. e.

$$D(p) = \frac{i}{p^2 - M^2 + i\epsilon},$$

where p is the energy transfer, and M is the mass of the new particle. Therefore, the cross-section will exhibit a peak when the transferred energy equals the mass of the particle. Such resonances can be identified when the cross section is measured. If the mass does not belong to any known particle, this signals the observation of a new particle.

In practice, however, this simple picture is complicated by interference, other channels, a finite decay width of the exchanged particles, and higher order effects, and very often more than just the two original particles will appear in the final state. Identifying the peak in any particular channel of the interaction is therefore very complicated, and there are several instances of ghost peaks known, created by constructive interference. Still, this is one of the major ways of discovering a new particle directly.

This discovery mode has the advantage that this is a counting experiment, i. e. the number of particles in the final state are counted and plotted as a (binned) function of the invariant mass, and then peaks are searched. Thus, no modeling is needed to identify the new resonance, making it rather robust discoveries. E. g., the Higgs has been discovered in this way. Of course, theory enters by selecting of which particles invariant mass plots should be made, as with about 20 particle in the final states at the LHC it becomes even with modern computers combinatorially challenging, especially when taking three or more particle decay channels into account, to check every possibility.

3.6.3 Missing energy

In principle akin to the concept of a resonance, the signature of missing energy is also associated directly with the (non-)observation of a particle. When two particles interact, the resulting particle may not be virtual, but real and stable, or at least sufficiently long-lived to escape the detector, in particular when its interactions with the standard model particles is small. Such a particle would surface in experiments as missing energy, i. e., the total observed energy would be smaller than before the collision, the remainder energy being carried away by the new particle. Looking for the smallest amount of missing energy would identify, using appropriate kinematics, the mass of the new particle. Dark matter, e. g., is assumed to produce precisely such a signature in collider experiments.

Again, in practice this concept is highly non-trivial, in particular due to muons, which can be compensated to some extent, and especially due to neutrinos. Thus, it is actually searched for a difference of missing energy compared to the standard model, and it is thus not a simple counting experiment, and precise values for the amount of missing energy in the standard model are needed.

Thus, it is a highly complicated theoretically problem to identify further properties of missing energy events to identify cases where the missing energy can be unambiguously associated with the production of a new particle, even if this is as simple as an abundance of missing energy events.

3.6.4 Precision observables

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

But it is also possible to conduct such investigations at collider. As an example, consider the rather popular oblique (electroweak) radiative corrections. Start with a generalization of the formulas for the W and Z bosons masses as

$$\begin{aligned} M_W^2 &= \frac{v_W^2}{2} g_i^2 \\ M_Z^2 &= \frac{1}{2} v_Z^2 (g_h^2 + g_i^2), \end{aligned}$$

thus permitting that the W and the Z perceive the vacuum expectation value of the Higgs differently. At tree-level, v_W and v_Z coincide in the standard model with the tree-level condensate v. Radiative corrections make all these quantities running, i. e., evolving with the momentum scale q^2 as $g_h^2(q^2)$, $g_i^2(q^2)$, $v_W^2(q^2)$, and $v_Z^2(q^2)$.

Now, rescale the weak isospin gauge field and the hypercharge gauge field as $\tilde{W}_\mu^a = g_i W_\mu^a$ and $\tilde{B}_\mu = g_i B_\mu$. The propagator $D_{\mu\nu}^{ij}$, with $i = 1..3$ or B, or $i = +, -, 3$ and A, of these gauge bosons can then be written as

$$D_{\mu\nu}^{ij}(q) = g_{\mu\nu}\Pi_{ij}(q) - q_\mu q_\nu \Pi_{ij}^T(q),$$

with the longitudinal and transverse self-energies Π and Π^T, respectively. Define now v_W and v_Z as

$$
\begin{aligned}
1/2v_Z^2 &= \frac{v^2}{2} - \Pi_{3B} = \frac{v^2}{2} - \Pi_{3A} + \Pi_{33} \\
1/2v_W^2 &= \frac{v^2}{2} + \Pi_{+-} - \Pi_{3A}
\end{aligned}
$$

and the couplings

$$
\begin{aligned}
\frac{1}{g_i^2} &= \frac{1}{g_{iu}^2} - \Pi_{33}^T - \Pi_{3B}^T = \frac{1}{g_{iu}^2} - \Pi_{3A}^T \\
\frac{1}{g_h^2} &= \frac{1}{g_{hu}^2} - \Pi_{BB}^T - \Pi_{3B}^T = \frac{1}{g_{hu}^2} + \Pi_{3A}^T - \Pi_{AA}^T
\end{aligned}
$$

where g_{iu} and g_{hu} are the unrenormalized coupling constants.

In the standard model, the dominant contributions at $q^2 = 0$ come from the third generation fermions and the Higgs. Computing the difference $v_W^2 - v_Z^2$ at $q^2 = 0$ in leading order perturbation theory yields

$$
\begin{aligned}
v_W^2 - v_Z^2 &= \frac{N_c}{32\pi^2}\left(m_t^2 + m_b^2 - \frac{2m_t^2 m_b^2}{(m_t^2 - m_b^2)} \ln\left(\frac{m_t^2}{m_b^2}\right)\right. \\
&\left. + \frac{M_W^2 m_H^2}{m_H^2 - M_W^2} \ln\left(\frac{m_H^2}{M_W^2}\right) - \frac{M_Z^2 m_H^2}{m_H^2 - M_Z^2} \ln\left(\frac{m_H^2}{M_Z^2}\right)\right).
\end{aligned}
$$

This is often also expressed as the Veltman ρ parameter as

$$
\begin{aligned}
\rho = \frac{v_W^2}{v_Z^2} &= 1 + \frac{N_c}{32v^2\pi^2}\left((m_t^2 + m_b^2) - \frac{2m_t^2 m_b^2}{(m_t^2 - m_b^2)} \log\left(\frac{m_t^2}{m_b^2}\right)\right. \\
&\left. + \frac{M_W^2 m_H^2}{m_H^2 - M_W^2} \ln\left(\frac{m_H^2}{M_W^2}\right) - \frac{M_Z^2 m_H^2}{m_H^2 - M_Z^2} \ln\left(\frac{m_H^2}{M_Z^2}\right)\right).
\end{aligned}
$$

New physics will modify these results. In particular, additional heavy fermion generations will add further terms, which could be detectable.

The deviation from the standard model can be parametrized by three parameters, the Peskin-Takeuchi parameters S, T, and U,

$$
\begin{aligned}
S &= 16\pi\left(\frac{\partial}{\partial q^2}\Pi_{33}\Big|_{q^2=0} - \frac{\partial}{\partial q^2}\Pi_{3Q}\Big|_{q^2=0}\right) \\
T &= \frac{4\pi}{\sin^2\theta_W \cos^2\theta_W M_Z^2}\left(\Pi_{WW}\big|_{q^2=0} - \Pi_{33}\big|_{q^2=0}\right) \\
U &= 16\pi\left(\frac{\partial}{\partial q^2}\Pi_{WW}\Big|_{q^2=0} - \frac{\partial}{\partial q^2}\Pi_{33}\Big|_{q^2=0}\right).
\end{aligned}
$$

At the current level of approximation, these parameters take the values

$$\begin{aligned}
S &= \frac{N_c}{6\pi}\left(1-y_Q\ln\frac{m_b^2}{m_t^2}\right)\\
T &= \frac{N_c}{4\pi\sin^2\theta_W\cos^2\theta_W M_Z^2}\left(m_t^2+m_b^2-\frac{2m_t^2m_b^2}{(m_t^2-m_b^2)}\ln\frac{m_t^2}{m_b^2}\right.\\
&\quad\left.+\frac{M_W^2m_H^2}{m_H^2-M_W^2}\ln\frac{m_H^2}{M_W^2}-\frac{M_Z^2m_H^2}{m_H^2-M_Z^2}\ln\frac{m_H^2}{M_Z^2}\right)\\
U &= \frac{N_c}{6\pi}\left(-\frac{5m_t^4-22m_t^2m_b^2+5m_b^4}{3(m_t^2-m_b^2)^2}\right.\\
&\quad\left.+\frac{m_t^6-3m_t^4m_b^2-3m_t^2m_b^4+m_b^6}{(m_t^2-m_b^2)^3}\ln\frac{m_t^2}{m_b^2}\right).
\end{aligned}$$

Since all input quantities can be determined independently directly from experiments, it is straightforward to compute the values of S, T, and U in the standard model. On the other hand, these three quantities can be also (indirectly) measured experimentally. Thus, comparing both ways of determining them should yield coinciding results inside the standard model. Thus, not only deviations from the theoretical value but also any discrepancies between both ways of determination should indicate new physics.

3.6.5 Anomalous couplings

Another interesting observable are the couplings of the standard model, which are essentially determined by cross-sections. In particular, at tree-level all scattering processes with a single interaction are directly proportional to the square of the coupling constants, and perturbatively higher orders can be computed. They can therefore be measured precisely. This allows for two different types of tests.

One is a comparison of some coupling measured in different processes. Since every interaction affects multiple particles in the standard model, this is possible. The second is that certain coupling constants are related due to the coupling universality in non-Abelian gauge theories as well as the Goldstone boson equivalence theorem. This affects especially the electroweak three-point and fout-point couplings. Thus, measuring anomalous values of these couplings would directly hint at new effects.

This is usually described in the κ framework. This defines κ =actual coupling/coupling in the standard model. Thus, if the standard model is the accurate theory, all $\kappa = 1$. Theoretically, this is modelled by modifiyng tree-level interactions as $g \to \kappa g$, while experimentally cross sections are measured and then divided by the theoretical expectations. This corresponds to an energy-independent modification of the couplings, and thus is especially sensitive to the high-energy tail of interactions, where single tree-level processes dominate. Current measurements yield essentially always that either κ is consistent with one or, if not yet statistically significantly detected, an upper bound.

3.6.6 Low-energy effective theories

Especially observations of the type discussed in section 3.6.4 and 3.6.5 are rather ambiguous, and can arise from very many types of new physics. Conversely, any deviation can usually be accommodated by many models. Thus systematically searching for such effects is at the same time highly model-dependent and not very constraining.

To avoid having to scan all possible models for all possible kinds of deviations has led to the use of low-energy effective theories. This approach, well developed for use in hadronic physics, is based on the following recipe: Start with the standard-model Lagrangian. Then add all possible higher-dimensional operators, up to some canonical dimension, which can be build from the standard-model fields and are compatible with the desired symmetries, usually, but not always, the symmetries of the standard model. Concerning the latter point e. g. explicit violations of C, P, CP or the custodial symmetry are often admitted, as they are not generically conserved in many BSM models. Finally, perturbation theory is done. Of course, this yields unitarity violation as such a theory is generically non-renormalizable. This introduces additional counter-terms and an explicit cutoff, which become parameters of the theory, and need to be fixed experimentally.

These deviations are usually encoded in terms of dimensionless Wilson coefficients c , which multiply these additional terms as c/Λ^n, with n a suitable power to make the terms in the Lagrangian having the correct canonical dimensions. Usually approaches limit n to a maximum value, yielding a low-energy effective theory up to a certain order in Λ. The scale Λ is given then as scale of new physics. Experiments will only be able to give expressions for the combinations, but theoretially often a common scale is assumed for all terms. Thus experimental limits are usually upper bounds to the combinations c/Λ^n, which can be satisfied by either making Λ large or the c small. Note that the various options of which terms are admitted to the effective theory can yield different limits for the Wilson coefficients. Conversely, for any given extension of the standard model at high energies, such a low-energy effective theory can be derived, providing predictions for the Wilson coefficients. However, this approximation will break down once the energies probed experimentally are of the same size as Λ, in which case the framework will no longer be reliable. It is thus particularily suited for extensions with a large Λ.

While this point is disadvantageous, the setup is still desirable as it allows to systematically parametrize all deviations in precision measurements of the types done in sections 3.6.4 and 3.6.5. Especially, it identifies the sectors of the standard model relevant to a deviation, and the dimensionful couplings give an estimate of the energy scale where new physics becomes relevant, if either a deviation is measured or a lower bound if none is measured and the couplings have therefore to be reduced.

While conceptually rather clean, and tested in hadron physics, this is mostly useful at tree-level, as the standard-model allows for many possible operators and thus requires many additional inputs at loop-level. Also, because field redefinitions in the standard-model and Fierz transformations allow many equivalent writings of the low-energy effective theory, but with differing values of coupling constants, it is mandatory to make sure that any conventions are strictly observed.

Chapter 4

Supersmmetry

Supersymmetry is much more than a particular theory. It is a conceptual idea, on which a multitude of theories rest. Supersymmetric quantum field theories have furthermore unique features, not shared by any other type of quantum field theories. They are therefore particular interesting candidates for our understanding of nature. Also, theories more complex than quantum field theories, like string theories, very often induce as the low-energy effective quantum field theories supersymmetric theories.

Supersymmetry offers not always a compelling solution to the issues of the standard model, but often an attractive one. Although, there are many technical details unsolved of how these solutions should be implemented. It is therefore worthwhile to understand the basics of it.

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

4.1 The conceptual importance of supersymmetry

To contemplate what supersymmetry implies it is worthwhile to have a look at a, somewhat hand waving, version of the Coleman-Mandula theorem.

All previous symmetries in particle physics are of either of two kinds. One are the external symmetries, like translational and rotational ones. These are created by the momentum and angular momentum operators. The other one are internal ones, like electric charges. The difference between both is that external charge operators carry a Lorentz index, while internal ones are Lorentz scalars. The natural question from a systematic point of view is, whether there are other conserved quantities besides momentum and angular momentum, which have a Lorentz index.

The Coleman-Mandula theorem essentially states that this is impossible in a quantum field theory. Since the most general vector and anti-symmetric tensors are already assigned to the momentum and angular momentum operator, the simplest one would be a symmetric tensor operator $Q_{\mu\nu}$. Acting with it on a single particle state would yield

$$Q_{\mu\nu}|p\rangle = (\alpha p_\mu p_\nu + \beta\eta_{\mu\nu})|p\rangle,$$

where the eigenvalue is the most general one compatible with Poincare symmetry, with eigenvalues α and β. Since a symmetry is looked for, Q must be diagonalizable simultaneously with the Hamiltonian, and therefore these must be momentum-independent. One could ask what if the eigenvalues themselves would have a direction, and the single-particle state would thus be characterized by two vectors. In this case, the scalar product of these two vectors would single out a direction, and therefore break the isotropy of space-time, and thus the Poincare group. Lacking any experimental evidence for this so far, this possibility is excluded, and would anyhow alter the complete setting.

So far, there is no contradiction. Acting with $Q_{\mu\nu}$ on a two-particle state of two identical particles of the type would yield

$$Q_{\mu\nu}|p,q\rangle = (\alpha(p_\mu p_\nu + q_\mu q_\nu) + 2\beta\eta_{\mu\nu})|p,q\rangle.$$

In this case, it was assumed that Q is a one-particle operator, i. e. its charge is localized on a particle, and the total charge is obtained by the sum of the individual charges. Though operators of other types can be considered, even in lack of physical evidence, this would only complicate the argument in the following, without changing the outcome. This will therefore be ignored.

Now consider elastic scattering of these two particles. Since Q should describe a symmetry, the total charge before and afterwards must be the same. Furthermore, 4-momentum conservation must hold. This implies that

$$\begin{aligned} p_\mu p_\nu + q_\mu q_\nu &= p'_\mu p'_\nu + q'_\mu q'_\nu \\ p_\mu + q_\mu &= p'_\mu + q'_\mu \end{aligned}$$

The only solution to these equations is $p = p'$ and $q = q'$ or $p = q'$ and $q = p'$. Hence no interaction occurs, since the second possibility is indistinguishable for two identical particles. Hence, any theory with such a conserved symmetric tensor charge would be non-interacting, and therefore not interesting. The generalized version of this statement is the Coleman-Mandula theorem, which includes all the subtleties and possible extensions glossed over here.

How does supersymmetry change the situation? For this simple example of elastic scattering, this is rather trivial. Supersymmetry will be defined by allowing to change the spin of the particles. Thus, it does not affect the example, as spin plays no role in this process, as in an elastic scattering the particle identities are not changed. In the more general case, it can be be shown that for the more general Coleman-Mandula theorem it is actually an assumption that this never happens. Hence, introducing such a symmetry violates the assumption, and therefore invalidates the argument. Again, the complete proof is rather subtle. This leaves the question open, whether any interesting, consistent, non-trivial, let alone experimentally relevant, theories actually harbor supersymmetry. The second question is still open, and no experimental evidence in strong favor of supersymmetry has so far been found.

The first question is, whether such theories can be formulated. Indeed, it will be shown that there are interesting, consistent, and non-trivial supersymmetric theories. Furthermore, it will be shown that supersymmetry is, from a conceptual point, a fundamental change. This can already be inferred from the following simple argumentation.

Since Poincare symmetry remains conserved in a supersymmetric theory, any operator which changes a boson into a fermion must carry itself a half-integer spin, and thus be a spinor Q_a. Otherwise, the spin on the left-hand side and the right-hand side would not be conserved. Furthermore, if it is a symmetry, it must commute with the Hamiltonian $[Q_a, H] = 0$. In addition, so must its anti-commutator

$$[\{Q_a, Q_b\}, H] = 0$$

In the simplest case, a spinor has two independent components, and therefore the anti-commutator can have up to four independent components. Since two spinors form together an object of integer spin, it must therefore be (at least) a vector. The only-vector-valued operator, however, is the momentum operator P_μ. Therefore, one would expect that

$$\{Q_a, Q_b\} \sim P_\mu. \tag{4.1}$$

It will be shown that this is indeed the correct structure. However, this means that supersymmetry enlarges indeed the Poincare symmetry non-trivially, since otherwise all (anti)commutator relation would be closed within the supersymmetry operators Q_a, therefore fulfilling the original goal. Moreover, the operators Q_a therefore behave, in a certain sense to be made precise later, like a square-root of the momentum operator. Similar like the introduction of i as the square-root of -1, this concept will require to enlarge the concept of space-time by adding additional, fermionic dimensions, giving birth to the concept of superspace. This already shows how conceptually interesting supersymmetry is, and that it is therefore worthwhile to pursue it for its own sake, even if no experimental evidence in favor of it exists.

4.2 Non-interacting supersymmetric quantum field theories

Supersymmetry is a theory, which will relate bosons and fermions, as will be seen. Thus, it requires to have both of them. To show, what supersymmetry is and how it comes about, it is useful to start outh with a non-interacting theory.

4.2.1 Fermions

While bosons can be incorporated in supersymmetric theories rather straightforwardly, a little more is needed in case of fermions. It will be found that supersymmetry requires the same number of bosonic and fermionic degrees of freedom to appear in a theory. Fermions in particle physics are encountered, e. g., in the form of electrons, which are described by Dirac spinors. These spinors include not only the electron, but also its antiparticle. As both have the possibility to have spin up or down, these are four degrees of freedom. This would require at least four bosons to build a supersymmetric theory. This is already quite a number of particles. However, it is also possible to construct fermions which are their own antiparticles. Therefore the number of degrees of freedom is halved. These are called Majorana fermions. Since these work quite a little differently than ordinary fermions, these will be introduced in this section. However, as yet this is a purely theoretical concept. No Majorana fermions have been observed in nature so far, although there are speculations that neutrinos, which are usually described by ordinary fermions, may be Majorana fermions, but there is no clear experimental evidence for this. These Majorana fermions, with identical particle and anti-particle, can mathematically also described as only one particle. This the so-called Weyl-fermion formulation[1]. It is this formulation,

[1]Note that in the more general case of non-supersymmetric theories there are subtle differences between Weyl fermions and Majorana fermions, especially if the number of dimensions differ from 4.

which will be used predominantly here. However, also the Majorana formulation is useful, and will be introduced briefly.

Note that fermions always have to have at least spin 1/2 as a consequence of the so-called CPT-theorem (or, equivalently, Lorentz invariance). These are two degrees of freedom. Hence, it is not possible to construct a supersymmetric theory with less than two fermionic and two bosonic degrees of freedom, at least in four dimensions.

As the spinors describing fermions are actually complex, and only by virtue of the equations of motion are reduced to effectively two degrees of freedom, in principle also four bosonic degrees of freedom are needed off-shell, that is without imposing the equations of motions. This will be ignored for now, and will only be taken up later, when it becomes necessary to take this distinction into account when quantizing the theory.

It is useful to introduce a compact index notation to treat Weyl spinors. This will be done, similarly to the case of special relativity, by the position of the indices. This notation is essentially based on the structure of the Lorentz group.

The Lorentz group consists out of rotations J and boosts K. In general, commutators of J and K do not vanish. However, defining skew versions of these operators

$$\begin{aligned} A &= \frac{1}{2}(J+iK) \\ B &= \frac{1}{2}(J-iK) \end{aligned}$$

this is the case. The Lorentz algebra becomes then a direct product of two SU(2) algebras

$$\begin{aligned} [A_i, A_j] &= \epsilon_{ijk}A_k \\ [B_i, B_j] &= \epsilon_{ijk}B_k \\ [A_i, B_j] &= 0. \end{aligned} \tag{4.2}$$

Hence, any representation of the Lorentz group can be assigned two independent quantum numbers, which are either integer or half-integer. E. g. scalars are then just twice the trivial case. Right-handed and left-handed fermions, however, belong to the $(1/2, 0)$ and $(0, 1/2)$ representations, vectors like the momentum belong to the $(1/2, 1/2)$ representation, and antisymmetric tensors like the generators of angular momentum to the $(1, 0) + (0, 1)$ representation. The simplification will now be used on distinguishing indices of the two different representations.

For this purpose, define the meaning of the index position for a left-handed spinor χ by

$$\begin{pmatrix} \chi^1 \\ \chi^2 \end{pmatrix} = i\sigma_2\chi = \begin{pmatrix} \chi_2 \\ -\chi_1 \end{pmatrix}. \tag{4.3}$$

Hence, given an ordinary left-handed spinor χ with components χ_1 and χ_2, the corresponding right-handed spinor ψ has components χ^1 and χ^2.

Since scalars are obtained by multiplying left-handed spinors with right-handed spinors, these can be obtained as

$$\alpha^T\beta = (\alpha^1\alpha^2)\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} = \alpha^1\beta_1 + \alpha^2\beta_2 = \alpha^a\beta_a.$$

This is very similar to the case of special relativity. Note that spinors are usually Grassmann-valued. Hence the order is relevant. The common convention is that the indices appear from top left to bottom right. Otherwise a minus-sign appears in the case of Grassmann-spinors,

$$\alpha^a\beta_a = -\beta_a\alpha^a,$$

and correspondingly for more elements

$$\alpha^a\beta^b\gamma_a\delta_b = -\alpha^a\gamma_a\beta^b\delta_b = -\gamma_a\alpha^a\delta_b\beta^b.$$

From the definition (4.3), it is also possible to read-off a 'metric' tensor, which can be used to raise and lower an index, the totally anti-symmetric rank two tensor ϵ^{ab}, yielding

$$\chi^a = \epsilon^{ab}\chi_b.$$

where $\epsilon^{12} = 1$ and $\epsilon_{12} = -1$.

This fixes the notation for left-handed spinors. Since there are also right-handed spinors, it is necessary to introduce a corresponding notation for them. However, in general the same notation could quickly lead to ambiguities. Therefore, a different convention is used: Left-handed spinors receive also upper and lower indices, but these in addition have a dot,

$$\begin{pmatrix}\psi_{\dot{1}}\\ \psi_{\dot{2}}\end{pmatrix} = -i\sigma_2\psi = \begin{pmatrix}-\psi^{\dot{2}}\\ \psi^{\dot{1}}\end{pmatrix}.$$

It is then possible to contract these two indices analogously to obtain a scalar, but this time the ordering will be defined to be from bottom left to top right

$$\alpha^T\beta = \alpha_{\dot{a}}\beta^{\dot{a}}.$$

Given this index notation, there are no ambiguities left in case of expressions with explicit indices. To be able to separate these also without using the indices explicitly, usually right-handed spinors are written as $\bar{\psi}$. This is not the same as the conventional Dirac-bar, and the equalities

$$\psi^{\dot{1}} = \bar{\psi}^{\dot{1}*}$$
$$\psi^{\dot{2}} = \bar{\psi}^{\dot{2}*}$$

hold. However, since complex conjugation is involved when it comes to treating left-handed spinors, here the definition is

$$\bar{\chi}_{\dot{a}} = \chi^*_{\dot{a}}$$

Therefore, a scalar out of left-handed spinors can now be written as

$$\chi^\dagger \chi = \bar{\chi}\chi$$

and similarly

$$\psi^\dagger \psi = \psi\bar{\psi}.$$

Another scalar combination, which will appear often is

$$\bar{\psi} \cdot \bar{\chi} = \bar{\psi}^\dagger \bar{\chi} = \epsilon_{ab}\psi^*_a \chi^{b*} = -\psi^*_1\chi^*_2 - \psi^*_2\chi^*_1.$$

Having now available a transformation which transforms a left-handed spinor into a right-handed spinor, it is natural to investigate what happens if both are combined into one single 4-component spinor. To obtain the correct transformation properties under Lorentz transformation, this object is

$$\Psi = \begin{pmatrix} \psi^{\dot{1}} \\ \psi^{\dot{2}} \\ -\psi^{\dot{2}*} \\ \psi^{\dot{1}*} \end{pmatrix}.$$

Since there are only two independent degrees of freedom, the spinor Ψ cannot describe, e. g., an electron. Its physical content is made manifest by performing a charge conjugation

$$C\Psi = \begin{pmatrix} 0 & i\sigma_2 \\ -i\sigma_2 & 0 \end{pmatrix} \begin{pmatrix} \psi^* \\ -i\sigma_2\psi \end{pmatrix} = \begin{pmatrix} \psi \\ -i\sigma_2\psi^* \end{pmatrix} = \Psi,$$

i. e., it is invariant under charge conjugation and thus describes a particle which is its own antiparticle, like the photon. Spin 1/2-particles with this property are called Majorana fermions, and thus this is a Majorana spinor. Note that this combination is not possible for arbitrary dimensions (and arbitrary space-time manifolds), but is correct in four dimensional Minkowski space-time. In this case, which covers almost all of this lecture, Weyl and Majorana fermions can be used synonymously.

4.2.2 The simplest supersymmetric theory

This is sufficient to set the scene for a first supersymmetric quantum field theory.

As discussed previously, it will be necessary to have the same number of fermionic and bosonic degrees of freedom. This requires at least two degrees of freedom, since it is not possible to construct a fermion with only one. Consequently, two scalar degrees of freedom are necessary. The simplest system with this number of degrees of freedom is a non-interacting system of a complex scalar field ϕ and a free Weyl fermion χ, which will be described by the undotted spinor. The corresponding Lagrangian is given by

$$\mathcal{L} = \partial^\mu \phi^\dagger \partial_\mu \phi + i\chi^\dagger \bar{\sigma}^\mu \partial_\mu \chi. \tag{4.4}$$

Note that here already with the fully quantized theory will be dealt. The corresponding physics will be invariant under a supersymmetry transformation if the action is invariant, up to anomalies. Since it is assumed that the fields vanish at infinity this requires invariance of the Lagrangian under the supersymmetry transformation up to a total derivative.

The supersymmetry transformations can be constructed by trial and error. Here, they will be introduced with hindsight of the results, and afterwards their properties will be analyzed. The transformation

$$A' = A + \delta A$$

takes for the scalar field the form

$$\delta\phi = \xi^a \chi_a = (-i\sigma_2 \xi)^T \chi. \tag{4.5}$$

Herein, ξ is a constant, Grassmann-valued spinor. By dimensional analysis, ξ has units of $1/\sqrt{\text{mass}}$. The corresponding transformation law for the spinor is

$$\delta\chi = -i\sigma^\mu \bar{\xi} \partial_\mu \phi = \sigma^\mu \sigma_2 \xi^* \partial_\mu \phi. \tag{4.6}$$

The pre-factor is fixed by the requirement that the Lagrangian is invariant under the transformation. The combination of ξ with σ_μ guarantees the correct transformation behavior of the expression under Lorentz transformation in spinor space. The derivative, which appears, is necessary to construct a scalar under Lorentz transformation in space-time, and to obtain the correct mass-dimension. It is the only object which can be used for this purpose, as it is the only one which appears in the Lagrangian (4.4), besides the scalar field. The general structure is therefore fixed by the transformation properties under Lorentz transformation. That the pre-factors are in fact also correct can be shown by explicit calculation,

$$\begin{aligned} \delta\mathcal{L} &= \partial_\mu((\delta\phi)^\dagger)\partial^\mu\phi + \partial_\mu\phi^\dagger\partial^\mu(\delta\phi) + (\delta\chi)^\dagger i\bar{\sigma}^\mu\partial_\mu\chi + \chi^\dagger i\bar{\sigma}^\mu\partial_\mu(\delta\chi) \qquad (4.7)\\ &= i\partial_\mu\chi^\dagger\sigma_2\xi^*\partial^\mu\phi + i\chi^\dagger\bar{\sigma}^\nu\sigma^\mu\sigma_2\xi^*\partial_\nu\partial_\mu\phi - i\partial_\mu\phi^\dagger\partial^\mu(\xi^T\sigma_2\chi) - i\xi^T\sigma_2\sigma^\mu\bar{\sigma^\nu}\chi\partial_\nu\partial_\mu\phi^\dagger \end{aligned}$$

Herein partial integrations have been performed, as necessary to obtain this form. There are two linearly independent terms, one proportional to ξ^* and one to ξ in this expression.

Both have therefore to either individually vanish or be total derivatives. To show this, it is helpful to note that

$$\bar{\sigma}^\nu\partial_\nu\sigma^\mu\partial_\mu = (\partial_0 - \sigma^j\partial_j)(\partial^0 + \sigma_i\partial^i) = \partial_0\partial^0 - \partial_i\partial^i = \partial^\mu\partial_\mu, \tag{4.8}$$

where it has been used that $\sigma_i^2 = 1$. Taking now only the terms proportional to ξ^* yields

$$i\partial_\mu\chi^\dagger\sigma_2\xi^*\partial^\mu\phi + i\chi^\dagger\sigma_2\xi^*\partial_\mu\partial^\mu\phi = \partial_\mu(\chi^\dagger i\sigma_2\xi^*\partial^\mu\phi). \tag{4.9}$$

This term is therefore indeed a total derivative. Likewise, also the term proportional to ξ^T can be manipulated to yield a pure total derivative. However, this is somewhat more complicated, as the combination (4.8) is not appearing. The last term can be rewritten as

$$-i\xi^T\sigma_2\sigma^\mu\bar{\sigma}^\nu\chi\partial_\nu\partial_\mu\phi^\dagger = \partial_\mu(\phi^\dagger i\xi^T\sigma_2\sigma^\mu\bar{\sigma}^\nu\partial_\nu\chi) + \phi^\dagger i\xi^T\sigma_2\sigma^\mu\bar{\sigma}^\nu\partial_\mu\partial_\nu\chi$$

It is then possible to use (4.8) on the last term to obtain

$$\partial_\mu(\phi^\dagger i\xi^T\sigma_2\sigma^\nu\bar{\sigma}^\mu\partial_\nu\chi) + \phi^\dagger i\xi^T\sigma_2\partial_\mu\partial_\mu\chi.$$

The first term is already a total derivative. The second term combines with the second-to-last term of (4.7) to a total derivative. Hence, the total transformation of the Lagrangian reads

$$\delta\mathcal{L} = \partial_\mu(\chi^\dagger i\sigma_2\xi^*\partial^\mu\phi + \phi^\dagger i\xi^T\sigma_2\sigma^\nu\bar{\sigma}^\mu\partial_\nu\chi + \phi^\dagger i\xi^T\sigma_2\partial^\mu\chi)$$

which is a total derivative.

Therefore, this theory is indeed supersymmetric. The set of fields ϕ and χ is called a supermultiplet. To be more precise, it is a left-chiral supermultiplet, because the spinor is left-handed. Replacing it with a right-handed spinor yields a right-chiral supermultiplet, without changing the supersymmetry of the theory, although, of course, the transformation is modified.

There should be a note of caution here. Unfortunately, it will turn out that this demonstration is insufficient to show supersymmetry of the quantized theory, and it will be necessary to modify the Lagrangian (4.4). This problem will become apparent when discussing the supersymmetry algebra. However, most of the calculations performed so far can be used unchanged.

4.3 Supersymmetry algebra

It turns out that the supersymmetry transformations (4.5) and (4.6) will form an algebra. This algebra can be used to systematically construct supermultiplets, and is useful for many other purposes. Therefore, this algebra will be constructed here, based first on the simplest examples of supersymmetry transformations (4.5) and (4.6) and will be generalized thereafter.

4.3.1 The superalgebra

The conserved Noether-supercurrent j^μ is

$$\begin{aligned} j^\mu &= -K^\mu + \frac{\partial\mathcal{L}}{\partial\partial_\mu\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial\partial_\mu\phi^\dagger}\delta\phi^\dagger + \frac{\partial\mathcal{L}}{\partial\partial_\mu\chi}\delta\chi \\ &= -\chi^\dagger i\sigma_2\xi^*\partial^\mu\phi + \partial_\nu\phi^\dagger i\xi^T\sigma_2\sigma^\nu\bar\sigma^\mu\chi + \phi^\dagger i\xi^T\sigma_2\partial_\mu\chi \\ &\quad -\partial^\mu\phi^\dagger\xi^T i\sigma_2\chi + \chi^\dagger i\sigma_2\xi^*\partial^\mu\phi + \chi^\dagger\bar\sigma^\mu\sigma_\nu i\sigma_2\xi^*\partial^\nu\phi \end{aligned}$$

Here, and in the following, the necessary contribution from the hermitian conjugate contribution are not marked explicitly. This result can be directly reduced, since some terms cancel, to

$$\begin{aligned} j^\mu &= \chi^\dagger\bar\sigma^\mu\sigma_\nu i\sigma_2\xi^*\partial^\nu\phi - \partial_\nu\phi^\dagger i\xi^T\sigma_2\sigma^\nu\bar\sigma^\mu\chi \\ &= \xi^T(-i\sigma_2)J^\mu + \xi^* i\sigma_2 J^{\mu *} \\ J^\mu &= \sigma^\nu\bar\sigma^\mu\chi\partial_\nu\phi^\dagger. \end{aligned}$$

J^μ is the so-called supercurrent, which forms the conserved current by a hermitian combination, similar to the probability current in ordinary quantum mechanics, $i\psi^\dagger\partial_i\psi + i\psi\partial_i\psi^\dagger$. To write the complex-conjugate part of the current, it has been used that

$$\sigma_2\sigma^\nu\bar\sigma^\mu = \sigma^\nu\bar\sigma^\mu\sigma_2$$

which follows by the anti-commutation rules for the Pauli matrices. This permits to construct the supercharge

$$Q = \int d^3x\sigma^\nu\chi\partial_\nu\phi^\dagger.$$

This indeed generates the transformation for the fields ϕ and χ. It is now possible to construct the algebra.

First of all, the (anti-)commutators

$$[\bar{Q},\bar{Q}] = 0 \qquad [Q,Q] = 0 \qquad \{Q,Q\} = 0 \qquad \{\bar{Q},\bar{Q}\} = 0$$

all vanish, since in all cases all appearing fields (anti-)commute. There are thus, at first sight, only one non-trivial commutator and one non-trivial anti-commutator.

For the non-vanishing cases, it is simpler to evaluate two consecutive applications of SUSY transformations. To perform this, note first that

$$[q,[p,f]] + [p,[f,q]] + [f,[q,p]] = 0.$$

This can be shown by direct expansion. It can be rearranged to yield

$$[[q,p],f] = [q,[p,f]] - [p,[q,f]].$$

If q and p are taken to be ξQ and $\bar{\eta}\bar{Q}$, and f taken to be ϕ, this implies that the commutator of two charges can be obtained by determining the result from two consecutive applications of the SUSY transformations. Using (4.5) and (4.6), it is first possible to obtain the result for this double application. It takes the form

$$[\xi Q + \bar{\xi}\bar{Q}, [\eta Q + \bar{\eta}\bar{Q}, \phi]] = -i[\xi Q + \bar{\xi}\bar{Q}, \eta^T(-i\sigma_2)\chi] = i\eta^T(-i\sigma_2)\sigma^\mu(-i\sigma_2\xi^*)\partial_\mu\phi. \quad (4.10)$$

Subtracting both possible orders of application yields then the action of the commutator

$$[[\xi Q + \bar{\xi}\bar{Q}, \eta Q + \bar{\eta}\bar{Q}], \phi] = i(\xi^T(-i\sigma_2)\sigma^\mu(-i\sigma_2\eta^*) - \eta^T(-i\sigma_2)\sigma^\mu(-i\sigma_2\xi^*))\partial_\mu\phi.$$

Here it has been used that $\bar{Q}$ is commuting with ϕ, as it does not depend on $\phi^\dagger$.

Aside from a lengthy expression $f(\eta, \xi)$, which gives the composition rule for the parameters, there is one remarkable result: The appearance of $-i\partial_\mu\phi$, which is the action of P_μ on ϕ, the momentum or generator of translations. Hence, the commutator is given by

$$[\xi Q + \bar{\xi}\bar{Q}, \eta Q + \bar{\eta}\bar{Q}] = f(\eta, \xi)P_\mu. \quad (4.11)$$

In fact, this is not all, due to aforementioned subtlety involving the fermions. This will be postponed to later.

Though anticipated in the introductory section 4.1, the appearance of the momentum operator seems at first surprising. Still, this implies that the supercharges are also something like the squareroot of the momentum operator, which leads to the notion of the supercharge being translation operators in fermionic dimensions. This idea will be taken up later when the superspace formulation will be discussed in section 4.5.

Hence, the algebra for the SUSY-charges will not only contain the charges themselves, but necessarily also the momentum operator. However, the relations are rather simple, as the supercharges do not depend on space-time and thus (anti-)commute with the momentum operator, as does the latter with itself.

Thus, the remaining item is the anti-commutator of Q with $\bar{Q}$. For this, again the commutator is useful, as it can be expanded as

$$\begin{aligned}[\eta Q, \bar{\xi} Q^\dagger] &= \eta_1\xi_1^*(Q_2Q_2^\dagger + Q_2^\dagger Q_2) - \eta_1\xi_2^*(Q_2Q_1^\dagger + Q_1^\dagger Q_2) \\ &\quad -\eta_2\xi_1^*(Q_1Q_2^\dagger + Q_2^\dagger Q_1) + \eta_2\xi_2^*(Q_1Q_1^\dagger + Q_1^\dagger Q_1)\end{aligned}$$

Thus, all possible anticommutators appear in this expression. The explicit expansion of (4.11) is

$$(\eta_2\xi_2^*(\sigma_\mu)_{11} - \eta_2\xi_1^*(\sigma_\mu)_{12} - \eta_1\xi_2^*(\sigma_\mu)_{21} + \eta_1\xi_1^*(\sigma_\mu)_{22})P^\mu.$$

Thus, by coefficient comparison the anti-commutator is directly obtained as

$$\{Q_a, Q_b^\dagger\} = (\sigma^\mu)_{ab}P_\mu. \quad (4.12)$$

This completes the algebra for the supercharges.

4.3.2 General properties of superalgebras

The superalgebra is what is a graded Lie algebra. So far, there had only been a single supercharge. As will be seen, additional supercharges can be introduced. But since even in the case of multiple supercharges there is still only one momentum operator, the corresponding superalgebras are coupled. In general, this requires the introduction of another factor δ^{AB}, where A and B count the supercharges. In total, it is shown in the Haag-Lopuszanski-Sohnius theorem that the most general superalgebra, up to rescaling of the charges, is

$$\{Q_a^A, \bar{Q}_b^B\} = 2\delta_{AB}\sigma^{\mu}_{ab}P_{\mu} \quad (4.13)$$

$$\{Q_a^A, Q_b^B\} = \epsilon_{ab}Z^{AB}, \quad (4.14)$$

and there is a corresponding anti-commutator for $\bar{Q}^A$ and $\bar{Q}^B$. The symbol ϵ_{ab} is anti-symmetric, and thus this anti-commutator couples the algebra of different supercharges. Furthermore, there appears an additional anti-symmetric operator Z^{AB} called the central charge. This additional operator can be shown to commute with all other operators, especially of internal symmetries, and must belong therefore to an Abelian U(1) group. Still, this quantum number characterize states, if a theory contains states with non-trivial representations. However, this can only occur if there is more than one supercharge. Finally, the number of independent supercharges is labeled by $\mathcal{N}$.

The theory with only one supercharge is thus an $\mathcal{N} = 1$ theory. Cases with $\mathcal{N} > 1$ are called $\mathcal{N}$-extended supersymmetries. Since the full algebra also involves the Poincare group generators, it turns out that it is not possible to have an arbitrary number of independent supercharges. This number depends on the size of the Poincare algebra, and thus on the number of dimensions. Furthermore, it also depends on the highest spins of particles involved. Especially, theories above a certain $\mathcal{N}$, 4 in four dimensions, require necessarily gravitons or spin 3/2 particles. Above a certain $\mathcal{N}$, 8 in four dimensions, even objects of still higher spins are required. The latter case is not particularly interesting presently, as it can be shown that a four-dimensional, perturbatively renormalizable, interacting quantum-field theory cannot include non-trivially interacting particles of spin higher than 2. Including gravitons requires to include gravity, a possibility which will for the moment not be considered. There is furthermore no physical evidence (yet) for spin 3/2 particles, so this option will also be ignored. Hence, for any non-gravitational theory in four dimensions, the maximum is $\mathcal{N} = 4$. The odd values 5 and 3 generate in four dimensions only the particle and anti-particle content of theories with larger $\mathcal{N}$, and therefore do not provide different theories: The particle and anti-particle content has to be included to satisfy the CPT theorem, and hence other possibilities are only relevant from a mathematical point of view, but not from a quantum-field-theoretical one.

Hence, in four dimensions there are thus besides $\mathcal{N}=1$ theories $\mathcal{N}=2$ theories and $\mathcal{N}=4$ theories. The additional charges provide more constraints, so theories with more supercharges become easier to handle. However, currently there seems to exist no hint that any other than $\mathcal{N}=1$-theories could be realized at energy scales accessible in the foreseeable future. Thus, here primarily this case will be treated. However, given the importance of such more complicated theories, especially in the context of string theory, it is worthwhile to gather here some more conceptual points about superalgebras.

All of this applies quantitatively to four dimensions. The number of independent supercharges can actually be larger or smaller for different dimensionalities, e. g. $\mathcal{N}=16$ without gravity is possible in two dimensions.

First of all, it should be noted that any theory with $\mathcal{N}>1$ necessarily contains also $\mathcal{N}=1$ supersymmetry. Hence, any theory with $\mathcal{N}>1$ can only be a special case of the most general form of a $\mathcal{N}=1$ theory. The appearance of higher symmetry is then obtained by restrictions on the type of interactions and the type of particles in the theory.

If, for a given theory, all central charges vanish, the algebra is invariant under a U($\mathcal{N}$) rotation of the supercharges, which is true especially for $\mathcal{N}=1$. In fact, in the case of $\mathcal{N}=1$, this R symmetry, or R parity, is just a (global) U(1) group, i. e. an arbitrary phase of the supersymmetry charges, which is part of the full algebra by virtue of

$$[T_R, Q_\alpha] = -i(\gamma_5)_\alpha^\beta Q_\beta.$$

This R symmetry forms an internal symmetry group with generator T_R. However, this symmetry may be explicitly, anomalously, or even spontaneously broken, without breaking the supersymmetry itself[2]. A broken R symmetry indicates merely that the relative orientation (and size) of the supersymmetry charges is (partly) fixed.

From the algebra (4.13-4.14) it can be read off that supercharges must have dimension of mass$^{\frac{1}{2}}$, and hence central charges of mass. This observation is of significance, as it embodies such theories with an inherent mass-scale. In fact, it can be shown that the mass of massive particles which form a supermultiplet in an extended supersymmetry must obey the constraint

$$M \geq \frac{1}{\mathcal{N}}\mathrm{tr}\sqrt{Z^\dagger Z},$$

and thus there is a minimum mass. If the mass satisfies the bound, such particles are

[2]This is not true in theories which satisfy the conformal version of the Poincare group, i. e. conformal theories. In this case, the R symmetry needs necessarily to be intact, and the combination of supersymmetry, conformal symmetry, and R symmetry forms together the superconformal symmetry.

called Bogomol'nyi-Prasad-Sommerfeld (BPS) states[3].

Even if the scale symmetry is broken, the superalgebras imposes constraints. Especially, in theories with R parity, this yields a connection between the R current and the trace of the energy-momentum tensor, which is intimately connected to the scale violation. Especially, various relations between R parity and scale operators remain intact even if both the R parity and the scale symmetry are broken at the quantum level. However, because of the Higgs sector, the standard model of particle physics is classically not scaleless, and therefore these relation do not apply in particle physics, until an extension of the standard model is found and experimentally supported which has both symmetries.

4.3.3 Supermultiplets

As already noted, supersymmetry requires different multiplets of particles to be present in a theory. This is a central concept, and requires further scrutiny. To simplify the details, once more mainly $\mathcal{N} = 1$ superalgebras will be considered.

A supermultiplet is a collection of fields which transform into each other under supersymmetry transformations. The naming convention is that a fermionic superpartner of a field a is called a-ino, and a bosonic super-partner s-a.

One result of the algebra obtained in the previous subsections was that the momentum operator (anti-)commutes with all supercharges. Consequently, also P^2 (anti-)commutes with all supercharges,

$$[Q_a, P^2] = [\bar{Q}, P^2] = \{Q, P^2\} = \{\bar{Q}, P^2\} = 0.$$

Since the application of P^2 just yields the mass of a pure state, the masses of a particle s and its super-partner ss must be degenerate, symbolically

$$\begin{aligned} P^2|s\rangle &= & m^2|s\rangle \\ P^2|ss\rangle &= P^2Q|s\rangle = QP^2|s\rangle = Qm^2|s\rangle = m^2Q|s\rangle = & m^2|ss\rangle. \end{aligned}$$

Further insight can be gained from considering the general pattern of the spin in a supermultiplet. The first step should be to show the often assured statement that the number of bosonic and fermionic degrees of freedom equals. Here, this will be done only for massless states. The procedure can be generalized to massive states, but this only complicates matters without adding anything new.

[3]That this is the same name as for certain topological excitations in gauge theories is not coincidental. In extended supersymmetric gauge theories both quantities are related.

As a starting point consider some set of massless states. Every state is then characterized by its four-momentum p^μ, with $p^2 = 0$, it spin s, and its helicity h, which for a massless particle can take only the two values $h = \pm s$ if $s \neq 0$ and zero otherwise.

Taking the trace of the spinor indices in (4.13) yields

$$Q_{i\alpha}Q^{\dagger j\alpha} + Q^{\dagger j\alpha}Q_{i\alpha} = \delta_{ij}P^0.$$

Applying to this a rotation operator by 2π and taking the trace over states with the same energy but different spins and helicities yields

$$\sum_{sh}\left\langle p,s,h\left|\left(Q_{i\alpha}Q^{\dagger j\alpha} + Q^{\dagger j\alpha}Q_{i\alpha}\right)e^{-2\pi iJ_3}\right|p,s,h\right\rangle = \delta^{ij}\sum_{sh}\left\langle p,s,h\left|P^0e^{-2\pi iJ_3}\right|p,s,h\right\rangle.$$

Since the supercharges are fermionic they anticommute with the rotation operator. Furthermore, the trace is cyclic, as it is a finite set of states. Thus, the expression can be rewritten as

$$\sum_{sh}\left\langle p,s,h\left|Q_{i\alpha}Q^{\dagger j\alpha}e^{-2\pi iJ_3} - Q_{i\alpha}Q^{\dagger j\alpha}e^{-2\pi iJ_3}\right|p,s,h\right\rangle = 0,$$

and thus also the right-hand-side must vanish. However, the right-hand-side just counts the number of states, weighted with 1 or -1, depending on whether the states are bosonic or fermionic. The number of helicity states differ, depending on whether the states are massive or not, yielding

$$\sum_{s\geq 0}(-1)^{2s}(2s+1)n_s = 0 \qquad (4.15)$$

$$n_0 + 2\sum_{s\geq 0}(-1)^{2s}n_s = 0, \qquad (4.16)$$

for the massive and the massless case, respectively. n_s is the number of states with the given spin, and n_0 the number of massless spin-0 particles. The factor 2 actually does not arise from the formula, as the trace is also well-defined when taking only one helicity into account. It is the CPT theorem which requires to include both helicity states for a physical theory.

In case of the massive Wess-Zumino model, the numbers are $n_0 = 2$ and $n_{1/2} = 1$, yielding $n_0 - 1(2)n_{1/2} = 0$. Note that this therefore counts on-shell degrees of freedom. For the massless case, the n_s are the same, but this time the 2 comes from a different place, $2 + 2(-1)1 = 0$.

So far, it is only clear that in a supermultiplet bosons and fermions must be present with the same number of degrees of freedom, but not necessarily their relative spins, as long as the conditions (4.15-4.16) are fulfilled. In the free example, one was a spin-0 particle and one a spin-1/2 particle. This pattern of a difference in actual spin of 1/2 is general.

This can be seen by considering the commutation relations of the supercharges under rotation. Since the supercharges are spinors, they must behave under a rotation δ_ϵ as

$$\delta_\epsilon Q = -\frac{i\epsilon\sigma}{2} Q = i\epsilon[J, Q],$$

where J is the generator of rotations. In particular, for the third component follows

$$[J_3, Q] = -\frac{1}{2}\sigma_3 Q$$

and thus

$$[J_3, Q_1] = -\frac{1}{2}Q_1 \qquad [J_3, Q_2] = \frac{1}{2}Q_2 \tag{4.17}$$

for both components of the spinor Q. It then follows directly that the super-partner of a state with total angular momentum j and third component m has third component $m \pm 1/2$, depending on the transformed spinor component. This can be seen as

$$J_3 Q_1|jm\rangle = (Q_1 J_3 - [Q_1, J_3])|jm\rangle = \left(Q_1 m - \frac{1}{2}Q_1\right)|jm\rangle = Q_1\left(m - \frac{1}{2}\right)|jm\rangle.$$

Likewise, the other spinor component of Q yields the other sign, and thus raises instead of lowers the third component. Of course, this applies vice-versa for the hermitian conjugates.

To also determine the value of j, assume a massless state with momentum $(p, 0, 0, p)$. The massive case is analogous, but more tedious. Start with the lowest state with $m = -j$. Then, of course, the state is annihilated by Q_1 and $Q_2^\dagger$, as they would lower m further.

Also $Q_1^\dagger$ annihilates the state, which is a more subtle result. The anti-commutator yields the result

$$Q_1^\dagger Q_1 + Q_1 Q_1^\dagger = (\sigma_\mu)_{11} P^\mu = p^0 - p^3,$$

where the minus-sign in the second term appears due to the metric. Thus

$$\left\langle pj - j \left| Q_1^\dagger Q_1 + Q_1 Q_1^\dagger \right| pj - j \right\rangle = p^0 - p^3 = p - p = 0 \tag{4.18}$$

but also

$$\langle pj - j|Q_1^\dagger = (Q_1|pj - j\rangle)^\dagger = 0,$$

as discussed above. Hence the first term in (4.18) vanishes, and leaves

$$\left\langle pj - j \left| Q_1 Q_1^\dagger \right| pj - j \right\rangle = 0.$$

But this is just the norm of $Q_1^\dagger|pj-j\rangle$. A zero-norm state is however not appropriate to represent a particle state, and thus $Q_1^\dagger$ has to annihilate the state as well, the only alternative to obtain the same result.

This leaves only $Q_2|pj-j\rangle$ as a non-zero state. This state has to be proportional to a state of type $|pj-j+1/2\rangle$. Since Q_2 is Grassmann-valued, and therefore nil-potent, a second application of Q_2 yields again zero. Furthermore, since Q_1 and Q_2 anticommute, its application also yields zero,

$$Q_1Q_2|pj-j\rangle = -Q_2Q_1|pj-j\rangle = 0.$$

The application of $Q_1^\dagger$ can be calculated as in the case of (4.18). But the appearing momentum combination is p_1+p_2, being zero for the state. This leaves only $Q_2^\dagger$. Applying it yields

$$Q_2^\dagger Q_2|pj-j\rangle = ((\sigma_\mu)_{22}P^\mu - Q_2Q_2^\dagger)|pj-j\rangle = (p^0+p^3+0)|pj-j\rangle = 2p|pj-j\rangle.$$

Hence, this returns the original state. Thus, the value of j in a supermultiplet can differ only by one half, and there are only two (times the number of internal quantum number) states in each supermultiplet. It does not specify the value of j, so it would be possible to have a supermultiplet with $j=0$ and $j=1/2$, as in the example above.

Note that only the states $m=0$ and $m=-1/2$, but not $m=+1/2$, the anti-particle state, are contained in the supermultiplet. This is called a chiral supermultiplet. Alternatively, it would be possible to have $j=1/2$ and $j=1$, the vector supermultiplet, or $j=2$ and $j=3/2$, the gravity supermultiplet appearing in supergravity.

While algebraically the anti-state is not necessary, any reasonable quantum field theory is required to have CPT-symmetry. Thus for any supermultiplet also the corresponding antiparticles, the antimultiplet, have to appear in the theory as well. By this, the missing $m=1/2$ state above is introduced into the theory.

To have supermultiplets which include more states requires to work with an $\mathcal{N}>1$ algebra, where the additional independent supercharges permit further rising and lowering. It is then possible to have supermultiplets, which include more different spin states. Also, the absence of central charges was important, since the anti-commutator of the fields has been used. E.g. in $\mathcal{N}=2$ SUSY, the supermultiplet contains two states with $j=0$ and two with $m=\pm 1/2$. This is also the reason why theories with $\mathcal{N}\neq 1$ seem to be no good candidates for an extension of the standard model: SUSY, as will be seen below, requires that all superpartners transform the same under other transformations, like e. g. gauge transformations. In an $\mathcal{N}=2$ theory, there would be a left-handed electron and a right-handed electron, both transforming under all symmetries in the same way. But in the standard model, the weak interactions couple differently to left-handed electrons and right-handed electrons, and thus it is not compatible with $\mathcal{N} > 1$ SUSY. For $\mathcal{N} = 1$ SUSY, however, independent chiral multiplets (and, thus, more particles) can be introduced to solve this problem. Since also the quarks are coupled differently in the weak interactions, this applies to all matter-fields, and thus it is not possible to enhance some particles with $\mathcal{N} = 1$ and others with $\mathcal{N} > 1$ SUSY.

One of the key quantities in the above discussion has been spin. However, spin is considered usually a good quantum number because it is a well-defined observable, as it commutes with the Hamiltonian. This is no longer precisely true when supersymmetry is involved. Rather, spin has a similar standing a sthe megnatic quantum number. Hence, a more general concept is needed.

A particle species is specified as a representation of the Poincare group or of one of its subgroup, i. e. fields are orbits in the Poincare group. Such orbits can be classified using Casimir invariants, as group theory shows.

One of the Casimirs is P^2, the square of the momentum operator, yielding the rest mass of a particle The spin appears as a second Casimir of the Poincare groupbuild from the Pauli-Lubanski vector

$$W_\mu = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}P^\nu M^{\rho\sigma}, \tag{4.19}$$

being, due to the Levi-Civita tensor, orthogonal to the momentum vector, and thus linearly independent. Its square W^2 is the searched-for second Casimir operator. In the rest frame of a massive particle

$$[W_i, W_j] = im\epsilon_{ijk}W_k \tag{4.20}$$

holds. This is, up to a normalization, just the spin algebra. This especially implies that its eigenvalues behave, up to a factor of m, like the ones of a spin, and indeed the eigenvalues of W^2 are thus spin eigenvalues.

In a supersymmetric theory, this is no longer true. The commutator of W^μ with a supercharge Q yields

$$\begin{aligned} [W^\mu, Q] &= \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}P_\nu[M_{\rho\sigma}, Q] = i\sigma^{\mu\nu}QP_\nu \\ \sigma_{\mu\nu} &= \frac{i}{4}(\sigma^\mu\bar{\sigma}^\nu - \sigma^\nu\bar{\sigma}^\mu) \end{aligned} \tag{4.21}$$

and thus

$$[W^2, Q] = W^\mu[W_\mu, Q] + [W^\mu, Q]W_\mu = 2iP^\mu\sigma_{\mu\nu}P^\nu Q$$

and therefore W^2 is no longer a suitable operator to characterize a particle, as it no longer belongs to the maximal set of commuting operators.

The problem arises, because of the connection of the superalgebra and the momentum operators. A suitable solution is therefore to generalize the Pauli-Lubanski-vector to a quantity also involving the supercharges. As it will turn out a suitable choice is the superspin, defined as

$$S^\mu = W^\mu - \frac{1}{4}Q\sigma^\mu\bar{Q}.$$

To check this, consider the commutation relation of the second part with a supercharge

$$\left[-\frac{1}{4}Q\sigma^\mu\bar{Q}, Q\right] = 2Q\sigma^\mu\bar{\sigma}^\nu P_\nu = 2\sigma^\nu\bar{\sigma}^\mu Q P_\nu,$$

which directly follows from the superalgebra. This implies, together with (4.21),

$$[S^\mu, Q] = -\frac{1}{2}QP^\mu, \tag{4.22}$$

where it has been used that $(i\sigma^{\mu\nu} + \sigma^\nu\bar{\sigma}^\mu/2) = g^{\mu\nu}/2$, as can be shown by explicit calculation. Since S^μ is Hermitian, the commutation relation with $\bar{Q}$ follows directly. Furthermore, S^μ commutes with P_μ as the Pauli-Lubanski vector does and so does the supercharge. Combining all of this together yields

$$[S^\mu, S^\nu] = i\epsilon_{\mu\nu\rho\sigma}P^\rho S^\sigma,$$

which reduces in the rest system in the same way as in (4.20) to

$$[S^i, S^j] = im\epsilon^{ijk}S^k.$$

This is again the same type of spin-algebra, characterized by eigenvalues s, as before. Thus, the superspin acts indeed as a spin.

It then only remains to construct an adequate Casimir operator. Define for this the antisymmetric matrix

$$C^{\mu\nu} = S^\mu P^\nu - S^\nu P^\mu,$$

which commutes with the supercharges

$$[C^{\mu\nu}, Q] = [S^\mu, Q]P^\nu - [S^\nu, Q]P^\mu = \frac{1}{2}\left(-QP^\mu P^\nu + QP^\nu P^\mu\right) = 0.$$

The square of this operator, $C^2 = C^{\mu\nu}C_{\mu\nu}$, commutes with the supercharges and also with P^μ, as it is a scalar. It remains to show that this Casimir is indeed different from P^2. An explicit calculation shows

$$C^2 = 2m^2S^2 - 2(SP)^2$$

which in the rest frame for a massive particle reduces to

$$C^2 = -2m^4s(s+1)$$

where S is the superspin in the rest frame and zero, integer, or half-integer. The situation for massless particles is as before, yielding for every spin only the corresponding helicities. Therefore, any particle can be assigned to representations of the Poincare group with the continuous parameters mass $m^2 \geq 0$ and the discrete superspin S, where the latter coincides in the rest frame, but only there, with the usual spin. Due to this coincidence, usually no difference is made between superspin and spin in name, and both expressions are used synonymously. Note that

$$[S^3, W^3] = 0,$$

and therefore the eigenvalues of both operators can be used to characterize the magnetic quantum number, especially in the rest frame. Thus, the above discussed counting in the rest frame is indeed legit.

It should be noted that the eigenvalue of the superspin can be used to characterize the supermultiplet, as the supercharges do not change it, but change between states inside the supermultiplet. Since the spin operator and supercharges do not commute, their application then changes the spin of particles belonging to the same supermultiplet.

E. g. for a supersin 0, the supercharges create the four states of the Wess-Zumino model. A superspin of 1/2 creates the vector multiplet, and of 1 the gravity multiplet. For different numbers of supercharges then also the number of particles in a multiplet at fixed superspin changes.

4.3.4 Off-shell supersymmetry

Now, it turns out that the results so far are not complete. If in (4.10) instead of ϕ the superpartner χ is used, it turns out that complications arise. Thus, something has to be modified.

To see this inconsistency, start by first performing two supersymmetry transformations on χ

$$\begin{aligned}
\delta_\eta \delta_\xi \chi_a = [\eta Q + \bar{\eta}\bar{Q}, [\xi Q + \bar{\xi}\bar{Q}, \chi_a]] &= -i[\eta Q + \bar{\eta}\bar{Q}, \sigma^\mu (i\sigma_2 \xi^*)_a \partial_\mu \phi] \\
&= -i(\sigma^\mu (i\sigma_2 \xi^*))_a \partial_\mu [\eta Q + \bar{\eta}\bar{Q}, \phi] \\
&= -i(\sigma^\mu (i\sigma_2 \xi^*))_a (\eta^T (-i\sigma_2)\partial_\mu \chi). \qquad (4.23)
\end{aligned}$$

To simplify this further, note that for any three spinors η, ρ, λ

$$\lambda_a \eta^b \rho_b + \eta_a \rho^b \lambda_b + \rho_a \lambda^b \eta_b = 0 \qquad (4.24)$$

holds. This follows by explicit calculation. E.g., for $a = 1$

$$\lambda_1 \eta_1 \rho_2 - \lambda_1 \eta_2 \rho_1 + \eta_1 \rho_1 \lambda_2 - \eta_1 \rho_2 \lambda_1 + \rho_1 \lambda_1 \eta_2 - \rho_1 \lambda_2 \eta_1 = 0.$$

To rearrange the terms such that they cancel always an even number of transpositions are necessary, and thus the Grassmann nature is not changing the signs. The case $a = 2$ can be shown analogously. Now, identify $\lambda = \sigma_\mu(-i\sigma_2)\xi^*$, $\eta = \eta$, and $\rho = \partial_\mu\chi$. Thus, (4.23) can be rewritten as

$$\delta_\eta\delta_\xi\chi = -i\left(\eta_a\partial_\mu\chi^T(-i\sigma_2)\sigma^\mu(-i\sigma_2)\xi^* + \partial_\mu\chi_a(\sigma^\mu(-i\sigma_2\xi^*))^T(-i\sigma_2)\eta\right).$$

Using

$$(-i\sigma_2)\sigma^\mu(-i\sigma_2) = (-\bar\sigma^\mu)^T, \tag{4.25}$$

which can be checked by explicit calculation, shortens the expressions significantly. Performing also a transposition, it then takes the form

$$\begin{aligned}\delta_\eta\delta_\xi\chi &= -i\eta_a(\xi^+\bar\sigma^\mu\partial_\mu\chi) + \partial_\mu\chi_a(\sigma^\mu(-i\sigma_2\xi^*)^T(-i\sigma_2)\eta)\\ &= -i\eta_a(\xi^T\bar\sigma^\mu\partial_\mu\chi) - i\eta^T(-i\sigma_2)\sigma^\mu(-i\sigma_2)\xi^*\partial_\mu\chi_a,\end{aligned}$$

where also the second term became transposed. To construct the transformation in reverse order is achieved by exchanging η and ξ, thus yielding

$$\begin{aligned}(\delta_\eta\delta_\xi - \delta_\xi\delta_\eta)\chi_a &= i(\xi^T(-i\sigma_2)\sigma^\mu(-i\sigma_2)\eta)^* - \eta^T(-i\sigma_2)\sigma^\mu(-i\sigma_2)\xi^*)\partial_\mu\chi_a\\ &\quad + i\xi_a(\eta^T\bar\sigma^\mu\partial_\mu\chi) - i\eta_a(\xi^T\bar\sigma^\mu\partial_\mu\chi).\end{aligned} \tag{4.26}$$

The first term is exactly the same as in (4.10), but with η and ξ exchanged. Hence, if this term would be the only one, the commutator of two SUSY transformations would be, in fact, the same irrespective of whether it acts on ϕ or χ, as it should. But it is not. The two remaining terms seem to make this impossible.

However, on closer inspection it becomes apparent that in both terms the expression $\bar\sigma^\mu\partial_\mu\chi$ exists. This is precisely the equation of motion for the field χ, the Weyl equation. Thus the two terms vanish, if the field satisfies its equation of motion. In a classical theory, this would be sufficient. However, in a quantum theory exist virtual particles, i. e., particles which not only not fulfill energy conservation, but also not their equations of motions. Hence such particles, which are called off (mass-)shell, are necessary. Thus the algebra so far is said to close only on-shell. Hence, although the theory described by the Lagrangian (4.4) is classically supersymmetric, it is not so quantum-mechanically. Quantum effects break the supersymmetry of this model.

Therefore, it is necessary to modify (4.4), to change the theory, to obtain one which is also supersymmetric on the quantum level. Actually, this result is already an indication of how this can be done. Off-shell, the number of degrees of freedom for a Weyl-fermion is four, and not two, as there are two complex functions, one for each spinor component.

Thus, the theory cannot be supersymmetric off-shell, as the scalar field has only two degrees of freedom. To make the theory supersymmetric off-shell, more (scalar) degrees of freedom are necessary, which, however, do not contribute at the classical level.

This can be done by the introduction of an auxiliary scalar field F, which has to be complex to provide two degrees of freedom. It is called auxiliary, as it has no consequence for the classical theory. The later can be most simply achieved by giving no kinetic term to this field. Thus, the modified Lagrangian takes the form

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial^\mu \phi + \chi^\dagger i\bar{\sigma}^\mu \partial_\mu \chi + F^\dagger F. \tag{4.27}$$

It should be noted that this field has mass-dimension two, instead of one as the other scalar field ϕ. The equation of motion for this additional field is

$$\partial_\mu \frac{\delta \mathcal{L}}{\delta \partial_\mu F} - \frac{\delta \mathcal{L}}{\delta F} = F^\dagger = 0 (= F).$$

Thus, indeed, at the classical level it does not contribute.

Of course, if it should contribute at the quantum level, it cannot be invariant under a SUSY transformation. The simplest (and correct) guess is that this transformation should only be relevant off-shell. As it must make a connection to χ, the ansatz is

$$\begin{aligned}
\delta_\xi F &= -i\xi^\dagger \bar{\sigma}^\mu \partial_\mu \chi \\
\delta_\xi F^\dagger &= i\partial_\mu \chi^\dagger \bar{\sigma}^\mu \xi,
\end{aligned}$$

where ξ has been inserted as its transpose to obtain a scalar. The appearance of the derivative is also enforced to obtain a dimensionally consistent equation.

This induces a change in the Lagrangian under a SUSY transformation as

$$\delta \mathcal{L}_F = F i\partial_\mu \chi^\dagger \bar{\sigma}^\mu \xi - F^\dagger i\xi^\dagger \bar{\sigma}^\mu \partial_\mu \chi.$$

This expression is not a total derivative. Hence, to obtain a supersymmetric theory additional modifications for the transformation laws of the other fields are necessary. However, since part of the fermion term already appears, the modifications

$$\begin{aligned}
\delta\chi &= \sigma_\mu \sigma_2 \xi^* \partial_\mu \phi + \xi F \\
\delta\chi^\dagger &= i\partial_\mu \phi^\dagger \xi^T (-i\sigma_2) \sigma^\mu + F^\dagger \xi^\dagger
\end{aligned}$$

immediately lead to cancellation of the newly appearing terms, and one additional total derivative,

$$\begin{aligned}
\delta \mathcal{L}_F &= F^\dagger \xi^\dagger i\bar{\sigma}^\mu \partial_\mu \chi + \chi^\dagger i\bar{\sigma}^\mu \xi \partial_\mu F + F i\partial_\mu \chi^\dagger \bar{\sigma}^\mu \xi - F^\dagger \xi^\dagger i\bar{\sigma}^\mu \partial_\mu \chi \\
&= \partial_\mu (i\chi^\dagger \bar{\sigma}^\mu \xi F) + F^\dagger \xi^\dagger i\bar{\sigma}^\mu \partial_\mu \chi - F^\dagger \xi^\dagger i\bar{\sigma}^\mu \partial_\mu \chi \\
&= \partial_\mu (i\chi^\dagger \bar{\sigma}^\mu \xi F)
\end{aligned}$$

Thus, without modifying the transformation law for the ϕ-field, the new Lagrangian (4.27) indeed describes a theory which is supersymmetric on-shell. To check this also off-shell, the commutator of two SUSY transformations has to be recalculated, which will be skipped here, as a full proof requires the recalculation also of all (anti-)commutators. This is a tedious work, but finally it turns out that the theory indeed has the same commutation relations, which hold also off-shell. In particular, the supercurrent is not modified at all, as no kinetic term for F appears, and the surface term and the one coming from the transformation of χ exactly cancel.

4.4 Interacting supersymmetric quantum field theories

4.4.1 The Wess-Zumino model

The theory treated so far was non-interacting and, after integrating out F using its equation of motion, had a very simple particle content. Of course, any relevant theory should be interacting. The simplest case will be constructed in this section. It is an extension of the free theory.

The starting point is the Lagrangian (4.27), supplemented with a yet unspecified interaction

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial^\mu \phi + \chi^\dagger i \bar{\sigma}^\mu \partial_\mu \chi + F^\dagger F + \mathcal{L}_i(\phi, \phi^\dagger, \chi, \chi^\dagger, F, F^\dagger) + \mathcal{L}_i^\dagger.$$

The last term is just the hermitian conjugate of the second-to-last term, necessary to make the Lagrangian hermitian. It is now necessary to find an interaction Lagrangian $\mathcal{L}_i$ such that supersymmetry is preserved.

Since the theory should be perturbatively renormalizable, the maximum dimension (for the relevant case of 3+1 dimensions) of the interaction terms is 4. The highest dimensional fields are F and χ. Furthermore, the interaction terms should be scalars. Thus the possible form is restricted to

$$\mathcal{L}_i = U(\phi, \phi^\dagger) F - \frac{1}{2} V(\phi, \phi^\dagger) \chi^a \chi_a.$$

The $-1/2$ is introduced for later convenience. On dimensional grounds, with F having dimension mass^2 and χ mass$^{\frac{3}{2}}$, no other terms involving these fields are possible. In principle it would appear possible to also have a term $Z(\phi, \phi^\dagger)$, depending only on the scalar fields. However, such a term could not include derivative terms. Under a SUSY transformation ϕ is changed into χ. Hence, the SUSY transformation will yield a term with three ϕ fields and one χ field for dimensional reasons, and no derivatives or F fields.

But all kinetic terms will include at least one derivative, as well as any transformations of the F field. The only possible term would be the one proportional to χ^2. But its SUSY transformation includes either an F field or the derivative of a ϕ-field. Hence none could cancel the transformed field. Since no derivative is involved, it can also not be changed into a total derivative, and thus such a term is forbidden.

Further, on dimensional grounds, the interaction term U can be at most quadratic in ϕ and V can be at most linear.

The free part of the action is invariant under a SUSY transformation. It thus suffices to only investigate the interacting part. Furthermore, if $\mathcal{L}_i$ is invariant under a SUSY transformation, so will be $\mathcal{L}_i^\dagger$. Thus, to start, consider only the second term. Since a renormalizable action requires the potentials U or V to be polynomial in the fields, the transformation rule is for either term, called Z here,

$$\delta_\xi Z = \frac{\delta Z}{\delta\phi}\delta_\xi\phi + \frac{\delta Z}{\delta\phi^\dagger}\delta_\xi\phi^\dagger.$$

The contribution from the SUSY transformation acting on the potential V yields

$$\frac{\delta V}{\delta\phi}(\xi^a\chi_a)(\chi^b\chi_b) + \frac{\delta V}{\delta\phi^\dagger}(\xi^{\dot{a}}\chi_{\dot{a}})(\chi^a\chi_a).$$

Again, none of these terms can be canceled by any other contribution appearing, since none of these can involve three times the χ field. Also, both cannot cancel each other, since the contributions are independent. As noted, the first term can be at maximum of the form $a + b\phi$, and thus only a common factor. By virtue of the identity (4.24) setting $\lambda = \rho = \eta = \chi$ it follows that this term is zero: In the case of all three spinors equal, the terms are all identical. This, however, does not apply to the second term, as no identity exist if one of the indices is dotted. There is no alternative other than to require that V is not depending on $\phi^\dagger$. Hence, the function V can be, and in four dimensions for renormalizable theories is, a holomorphic function[4] of the field ϕ. Being holomorphic is a quite strong constraint on a function, and hence it is quite useful in practice to obtain various general results for supersymmetric theories.

As an aside, this fact is of great relevance, as it turns out that it is not possible to construct the standard model Yukawa interactions due to this limitation with only one Higgs doublet as it is done in the non-supersymmetric standard model, but instead requires at least two doublets. The more detailed reason is once more the necessity to have two independent supermultiplets to represent the left-right asymmetry of the electroweak

[4]In theories with both left chiral supermultiplets and right chiral supermultiplets, e. g. $\mathcal{N} = 2$ theories, it is possible to also form a further potential which depends on both, the so-called Kähler potential. Kähler potentials are, in a sense, generalizations of holomorphic potentials, and hence also quite constraining.

interaction of the standard model, and this requires ultimately the doubling of Higgs particles in the minimal supersymmetric standard model. However, it is not required that both have the same mass.

This is already sufficient to restrict the function V to the form

$$V = M + y\phi.$$

The first term gives a mass to the fermionic fields, and the second term provides a Yukawa interaction. It is convenient, as will be shown latter, to write a generating functional for this term V as

$$V = \frac{\delta W}{\delta\phi\delta\phi}$$

with

$$W = B + A\phi + \frac{1}{2}M\phi^2 + \frac{1}{6}y\phi^3. \tag{4.28}$$

This function W is called the superpotential for historically reasons, and will play a central role, as will be seen later. The linear term is not playing a role here, but can be important for the breaking of SUSY, as will be discussed later. The constant is essentially always irrelevant.

The next step is to consider all terms which produce a derivative term upon a SUSY transformation. These are

$$\begin{aligned}
& -iU\xi^\dagger\bar{\sigma}^\mu\partial_\mu\chi + i\frac{1}{2}V\chi^T\sigma_2\sigma^\mu\sigma_2\xi^*\partial_\mu\phi - i\frac{1}{2}V\xi^\dagger\sigma_2\sigma^{\mu T}\sigma_2\chi\partial_\mu\phi \\
= \; & -iU\xi^\dagger\bar{\sigma}^\mu\partial_\mu\chi - i\frac{1}{2}V\xi^\dagger\sigma_2\sigma^{\mu T}\sigma_2\chi\partial_\mu\phi - i\frac{1}{2}V\xi^\dagger\sigma_2\sigma^{\mu T}\sigma_2\chi\partial_\mu\phi \\
= \; & -iU\xi^\dagger\bar{\sigma}^\mu\partial_\mu\chi - i\frac{1}{2}V\xi^\dagger\bar{\sigma}^\mu\chi\partial_\mu\phi - i\frac{1}{2}V\xi^\dagger\bar{\sigma}^\mu\chi\partial_\mu\phi \\
= \; & -iU\xi^\dagger\bar{\sigma}^\mu\partial_\mu\chi - iV\xi^\dagger\bar{\sigma}^\mu\chi\partial_\mu\phi,
\end{aligned}$$

where (4.25) was used twice. The combination of derivatives of ϕ and χ cannot be produced by any other contribution, since all other terms will not yield derivatives. Since neither U nor V are vectors, there are also no possibilities for a direct cancellation of both terms. The only alternative is thus that both terms could be combined to a total derivative. This is possible, if

$$U = \frac{\delta W}{\delta\phi}$$

as can be seen as follows

$$\begin{aligned}
& -i\xi^\dagger\bar\sigma^\mu\left(U\partial_\mu\chi+\chi V\partial_\mu\phi\right)\\
= & -i\xi^\dagger\bar\sigma^\mu\left(U\partial_\mu\chi+\chi\partial_\mu\frac{\delta W}{\delta\phi}\right)\\
= & -i\xi^\dagger\bar\sigma^\mu\left(\frac{\delta W}{\delta\phi}\partial_\mu\chi+\chi\partial_\mu\frac{\delta W}{\delta\phi}\right)\\
= & -i\xi^\dagger\bar\sigma^\mu\partial_\mu\left(\frac{\delta W}{\delta\phi}\chi\right).
\end{aligned}$$

This is indeed a total derivative. Hence, by virtue of the form of the superpotential (4.28)

$$U=\frac{\delta W}{\delta\phi}=A+M\phi+\frac{1}{2}y\phi^2.$$

This fixes the interaction completely. It remains to check that also all the remaining terms of the SUSY transformation cancel or form total derivatives. These terms are

$$\begin{aligned}
& \frac{\delta U}{\delta\phi}F\xi^a\chi_a-\frac{1}{2}V(\xi^a\chi_aF+\chi^a\xi_aF)\\
= & \frac{\delta^2 W}{\delta\phi\delta\phi}F\xi^a\chi_a-\frac{1}{2}\frac{\delta^2 W}{\delta\phi\delta\phi}(\xi^a\chi_aF+\xi^a\chi_aF)=0,
\end{aligned}$$

and hence the theory is, in fact, supersymmetric. Here it has been used how spinor scalar-products of Grassmann numbers can be interchanged, giving twice a minus-sign in the second term.

This permits to write down the full, supersymmetric Lagrangian of the Wess-Zumino model. It takes the form

$$\begin{aligned}
\mathcal{L}=\partial_\mu\phi^\dagger\partial^\mu\phi+\chi^\dagger i\bar\sigma^\mu\partial_\mu\chi+F^\dagger F & + M\phi F-\frac{1}{2}M\chi\chi+\frac{1}{2}y\phi^2F-\frac{1}{2}y\phi\chi\chi \\
& + M^*\phi^\dagger F^\dagger-\frac{1}{2}M^*(\chi\chi)^\dagger+\frac{1}{2}y^*\phi^{\dagger 2}F^\dagger-\frac{1}{2}y^*\phi^\dagger(\chi\chi)^\dagger,
\end{aligned}\tag{4.29}$$

where linear and constant terms have been dropped, and which is now the full, supersymmetric theory.

The appearance of three fields makes this theory already somewhat involved. However, the field F appears only quadratically, and without derivatives in the Lagrangian remains an auxiliary field, just as

$$\mathcal{L}_F=F^\dagger F+\frac{\delta W}{\delta\phi}F+\frac{\delta W^\dagger}{\delta\phi^\dagger}F^\dagger.$$

Thus, it is directly possible to integrate out F. This yields

$$\begin{aligned}
\mathcal{L} = & \partial_\mu\phi^\dagger\partial^\mu\phi+\chi^\dagger i\bar\sigma^\mu\partial_\mu\chi-MM^*\phi^\dagger\phi-\frac{1}{2}M\chi^T(-i\sigma_2)\chi-\frac{1}{2}M^*\chi^\dagger(i\sigma_2)\chi^{\dagger T}\\
& -\frac{y}{2}(M\phi\phi^{\dagger 2}+M^*\phi^\dagger\phi^2)-yy^*\phi^2\phi^{\dagger 2}-\frac{1}{2}(y\phi\chi^a\chi_a+y^*\phi^\dagger\chi^{a\dagger}\chi_a^\dagger)
\end{aligned}\tag{4.30}$$

There are a number of interesting observations to be made in this Lagrangian. First of all, at tree-level it is explicit that the fermionic and the bosonic field have the same mass. Of course, SUSY guarantees this also beyond tree-level. Secondly, the interaction-structure is now surprisingly the one which was originally claimed to be inconsistent with SUSY. The reason for this is that of course also in the SUSY transformations the equations of motions for F and $F^\dagger$ have to be used. As a consequence, these transformations are no longer linear, thus making such an interaction possible. Finally, although two masses and three interactions terms do appear, there are only two independent coupling constants, M and y. That couplings for different interactions are connected in such a non-trivial way is typical for SUSY. It was one of the reasons for hoping that SUSY would unify the more than thirty independent masses and couplings appearing in the standard model. Unfortunately, as will be discussed below, the necessity to break SUSY jeopardizes this, leading, in fact, for the least complex theories to many more independent couplings and masses, about three-four times as much.

4.4.2 Majorana form

For many actual calculations the form (4.30) of the Wess-Zumino model is actually somewhat inconvenient. It is often more useful to reexpress it in terms of Majorana fermions

$$\Psi = \begin{pmatrix} (i\sigma_2)\chi^* \\ \chi \end{pmatrix},$$

and the bosonic fields

$$\begin{aligned} A &= \frac{1}{\sqrt{2}}(\phi + \phi^\dagger) \\ B &= \frac{1}{\sqrt{2}}(\phi - \phi^\dagger). \end{aligned}$$

It can then be verified by direct expansion that the new Lagrangian in terms of these fields for a single flavor becomes

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}\bar{\Psi}(i\gamma^\mu\partial_\mu - M)\Psi + \frac{1}{2}\partial^\mu A\partial_\mu A - \frac{1}{2}M^2A^2 + \frac{1}{2}\partial^\mu B\partial_\mu B - \frac{1}{2}M^2B^2 \\ &\quad -MgA(A^2+B^2) + \frac{1}{2}g^2(A^2+B^2)^2 - g(A\bar{\Psi}\Psi + iB\bar{\Psi}\gamma_5\Psi) \end{aligned}$$

where M and $g = y$ have been chosen real for simplicity. In this representation, A and B do no longer appear on equal footing: A is a scalar field while B is necessarily pseudoscalar, due to its coupling to the fermions. Therefore, it can also only appear quadratic and not linear in the three-scalar term, explaining the absence of a BA^2 term. Still, despite these differences, this is a standard Lagrangian for which the Feynman rules are known. It will now be used to demonstrate the convenient behavior of a supersymmetric theory when it comes to renormalization.

4.4.3 The scalar self-energy to one loop

A benefit of supersymmetric theories is that they solve the so-called naturalness problem. What this explicitly means, and how it is solved, will be discussed here.

The naturalness problem is simply the observation that the Higgs particle is rather light, although the theory would easily permit it to be much heavier, of the order of almost the Planck scale, without loosing its internal consistency. The reason for this is the unconstrained nature of quantum fluctuations. Supersymmetric theories make it much harder for the Higgs to be very heavy, in fact, its mass becomes exponentially reduced compared to a non-supersymmetric theory. To see this explicitly, it is simplest to perform a perturbative one-loop calculation of the scalar self-energy in a theory with a very similar structure as the Wess-Zumino model, but with a fermion-boson coupling which is instead chosen to be h for the moment. This can be regarded as a simple mock-up of the electroweak sector of the standard model, dropping the gauge fields.

At leading order, there are three classes of diagrams appearing in the one-particle irreducible set of Feynman diagrams. The first is a set of tadpole diagrams, the second a set of one-loop graphs with internal bosonic particles, and the third the same, but with fermionic particles. These will be calculated in turn here. The calculation will be performed for the case of the A boson. It is similar, but a little more tedious, for the B boson.

The mathematically most simple ones are the tadpole diagrams. There are two of them, one with an A boson attached, and one with a B boson attached. Their contribution Π_t to the self-energy is

$$\Pi_t = -\frac{12}{2}g^2 \int \frac{d^4p}{(2\pi)^4} \frac{1}{p^2 - M^2 + i\epsilon} - \frac{4}{2}g^2 \int \frac{d^4p}{(2\pi)^4} \frac{1}{p^2 - M^2 + i\epsilon}, \tag{4.31}$$

where the first term stems from the A-tadpole and the second one from the B-tadpole, and the factors $1/2$ are symmetry factors. These integrals are divergent. Regularizing them by a cut-off Λ^2 turns it finite. This expression can then be calculated explicitly to yield

$$\begin{aligned}\Pi_t &= \frac{ig^2}{\pi^2}\left(\Lambda^2\sqrt{1+\frac{M^2}{\Lambda^2}}-M^2\ln\left(\frac{\Lambda+\Lambda\sqrt{1+\frac{M^2}{\Lambda^2}}}{M}\right)\right)\\ &\approx \frac{ig^2}{\pi^2}\left(\Lambda^2-M^2\ln\left(\frac{\Lambda}{M}\right)+\mathcal{O}(1)\right).\end{aligned} \tag{4.32}$$

This result already shows all the structures which will also appear in the more complicated diagrams below. First of all, the result is not finite as the cutoff is removed, i.e., by sending Λ to infinity. In fact, it is quadratically divergent. In general, thus, this expression would need to be renormalized to make it meaningful. Since the leading term is momentum-independent, this will require a renormalization of the mass, which is thus quadratically divergent. This is then the origin of the naturalness problem: In the process of renormalization, the first term will be subtracted by a term $-\Lambda^2+\delta m^2$, where the first term will cancel the infinity, and the second term will shift the mass to its physical value. However, even a slight change in Λ or g^2 would cover even a large change in δm, if the final physical mass is small. There is no reason why it should be small therefore, and thus the mass is not protected. If, e.g., the first would not be present, but only the logarithmic second one, the cancellation would be of type $M^2\ln\left(\frac{\Lambda}{M}\right)+\delta m^2$. Now, even large changes in Λ will have only little effect, and thus there is no fine-tuning involved to obtain a small physical mass. This will be exactly what will happen in a supersymmetric theory: The quadratic term will drop out in contrast to a non-supersymmetric one, and thus will provide a possibility to obtain a small physical mass without fine tuning of g, Λ and M.

The next contribution stems from the loop graphs involving a boson splitting in two. With an incoming A boson, it can split in either two As, two Bs, or two χs. The contribution from the bosonic loops are once more identical, up to a different prefactor due to the different coupling. Their contribution is

$$-\frac{1}{2}((6Mg)^2+(2Mg)^2)\int\frac{d^4p}{(2\pi)^4}\frac{1}{p^2-M^2+i\epsilon}\frac{1}{(p-q)^2-M^2+i\epsilon},$$

where the factor $1/2$ is a symmetry factor and q is the external momentum of the A particle. An explicit evaluation of this expression is possible, and discussed in many texts on perturbation theory. This is, in particular when using a cutoff-regularization, a rather lengthy exercise. However, to explicitly show how the naturalness problem is solved, it is only interesting to keep the quadratically divergent piece of the contribution. However, the integrand scales as $1/p^4$ for large momenta. Thus, the integral is only logarithmically divergent, and will thus only contribute at order $M^2\ln(\Lambda/M)$, instead of at Λ^2. For the purpose at hand, this contribution may therefore be dropped.

This leaves the contribution with a fermion loop. It reads

$$-2h^2\int\frac{d^4p}{(2\pi)^4}\frac{\mathrm{tr}((\gamma_\mu p^\mu+M)(\gamma_\nu(p^\nu-q^\nu)+M))}{(p^2-M^2+i\epsilon)((p-q)^2-M^2+i\epsilon)}.$$

The factor of 2 in front stems from the fact that for a Majorana fermion particle and anti-particle are the same. Thus, compared to an ordinary fermion, which can only split into particle and anti-particle to conserve fermion number, the Majorana fermion can split into two particles, two anti-particles, or in two ways in one particle and an anti-particle, in total providing a factor four. This cancels the symmetry factor 1/2 and lets even a factor of 2 standing. Using the trace identities $\mathrm{tr}1=4$, $\mathrm{tr}\gamma_\mu$=0, and $\mathrm{tr}\gamma_\mu\gamma_\nu=4g_{\mu\nu}$ this simplifies to

$$-8h^2\int\frac{d^4p}{(2\pi)^4}\frac{p(p-q)+M^2}{(p^2-M^2+i\epsilon)((p-q)^2-M^2+i\epsilon)}.$$

Since the numerator scales with p^2, the integral is quadratically divergent. Again, it suffices to isolate this quadratic piece, yielding

$$-\frac{ih^2}{\pi^2}\Lambda^2+\mathcal{O}\left(M^2\ln\frac{\Lambda^2}{M^2}\right).$$

This cannot cancel the previous contribution, unless $g=h$. However, for a supersymmetric theory, supersymmetry dictates $g=h$. But then this is just the negative of (4.32), and thus cancels exactly this contribution. Thus, all quadratic divergences appearing have canceled exactly, and only the logarithmic divergence remains. As has been allured to earlier, this implies a solution of the naturalness problem. In fact, it can be shown that this result also holds in higher order perturbation theory, and only logarithmic divergences appear, thus lower than just the superficial degree of divergence.

In fact, in the present case it is possible to reduce the number of divergences even further. For simplicity, the calculation above has been performed with the F field integrated out. Keeping this field explicitly, it is found (after a more tedious calculation) that the mass of the bosons (and fermions) become finite, and the divergences are all pushed into a wave-function renormalization. Hence, the masses of the particles become fixed, making supersymmetric theories much more predictive (and 'natural') than non-supersymmetric ones.

This feature of canceling quadratic divergences is no accident, but is a general feature of supersymmetric theories. Since fermions and bosons contribute with opposite sign, the fact that supersymmetry requires a precise match between both species leads always to cancellations which lower the degree of divergences. This is one of the most striking benefits of supersymmetric theories.

4.5 Superspace formulation

4.5.1 Supertranslations

After having now a first working example of a supersymmetric theory and seen its benefits, it is necessary to understand a bit more of the formal properties of supersymmetry. The possibly most striking feature is the somewhat mysterious relation (4.12). It is still not very clear what the appearance of the momentum operator in the SUSY algebra signifies. It will turn out that this connection is not accidental, and lies at the heart of a very powerful, though somewhat formal, formulation of supersymmetric theories in the form of the superspace formulation. This formulation will permit a more direct understanding of why the supermultiplet is as it is, and will greatly aid in the construction of more supersymmetric theories. For that reason, it has become the preferred formulation used throughout the literature.

To start with the construction of the super-space formulation, note that the supercharge Q is a hermitian operator. Thus, it is possible to construct a unitary transformation from it by exponentiating it, taking the form[5]

$$U(\theta,\theta^*) = \exp(i\theta Q)\exp(i\bar{\theta}\bar{Q}).$$

Note that the expressions in the exponents are appropriate scalar products, and that θ and θ^* are independent, constant spinors.

Acting with U on any operator ϕ yields thus a new operator ϕ dependent on θ and θ^*

$$U(\theta,\theta^*)\phi U(\theta,\theta^*)^{-1} = \phi(\theta,\theta^*)$$

by definition. This is reminiscent of ordinary translations. Given the momentum operator P_μ, and the translation operator $V(x) = \exp(ixP)$, a field ϕ at some point, say 0, acquires a position dependence by the same type of operation,

$$V(x)\phi(0)V(x)^\dagger = \phi(x).$$

Thus, in a sense, the operator U provides a field with an additional fermionic coordinate. Of course, this interpretation is done with hindsight, as any unitary operator is providing a field acted upon an additional degree of freedom.

That this interpretation actually makes sense can be most easily seen by applying the operator U twice. To evaluate the operator

$$U(\xi,\xi^*)U(\theta,\theta^*) = \exp(i\xi Q)\exp(i\bar{\xi}\bar{Q})\exp(i\theta Q)\exp(i\bar{\theta}\bar{Q})$$

[5] The order of Q and $\bar{Q}$ is purely conventional, see below.

it is most convenient to use the Baker-Campbell-Hausdorff formula

$$\exp(A)\exp(B) = \exp\left(A + B + \frac{1}{2}[A,B] + \frac{1}{6}[[A,B],B] + ...\right).$$

The commutator of ξQ and $\bar{\xi}\bar{Q}$ for the first two factors can be reduced to the known commutation relation (4.12) in the following way

$$\begin{aligned}
[i\xi Q, i\bar{\xi}\bar{Q}] &= i^2[\xi^a Q_a, -\xi^{b*}Q_b^\dagger] \\
&= i^2(-\xi^a Q_a \xi^{b*}Q_b^\dagger + \xi^{b*}Q_b^\dagger \xi^a Q_a) \\
&= i^2\xi^a\xi^{b*}(Q_a Q_b^\dagger + Q_b^+ Q_a) \\
&= i^2\xi^a\xi^{b*}\{Q_a, Q_b^\dagger\} \\
&= i^2\xi^a\xi^{b*}(\sigma^\mu)_{ab}P_\mu,
\end{aligned}$$

where (4.12) has already been used in the end. This is an interesting result. First of all, it is practical. Since P_μ commutes with both Q and $Q^\dagger$ all of the higher terms in the Baker-Campbell-Hausdorff formula vanish. Secondly, the appearance of the momentum operator, though not unexpected, lends some support to the idea of interpreting the parameters θ and ξ as fermionic coordinates and U as a translation operator in this fermionic space. But to make this statement more definite, the rest of the product has to be analyzed as well. Note that since all of the formulation has been covariant throughout the expression $i\xi^a\xi^{b*}(\sigma^\mu)_{ab}$, though looking a bit odd at first sight, has actually to be a four vector to form again a Lorentz invariant together with P_μ. Of course, this quantity, as a product of two Grassmann numbers, is an ordinary number, so this is also fine. Note further that since P commutes with Q, this part can be moved freely in the full expression.

Combining the next term is rather straight-forward,

$$\begin{aligned}
U(\xi,\xi^*)U(\theta,\theta^*) &= \exp(i^2\xi^a\xi^{b*}(\sigma^\mu)_{ab}P_\mu)\exp(i(\xi Q + \bar{\xi}\bar{Q}))\exp(i\theta Q)\exp(i\bar{\theta}\bar{Q}) \\
&= \exp(i^2\xi^a\xi^{b*}(\sigma^\mu)_{ab}P_\mu) \times \\
&\quad \times \exp\left(i\left(\xi Q + \bar{\xi}\bar{Q} + \theta Q - \frac{1}{2}[\xi Q + \bar{\xi}\bar{Q}, \theta Q]\right)\right)\exp(i\bar{\theta}\bar{Q}) \\
&= \exp(i^2(\xi^a\xi^{b*} + \xi^a\theta^{b*})(\sigma^\mu)_{ab}P_\mu)\exp(i(\xi Q + \bar{\xi}\bar{Q} + \theta Q))\exp(i\bar{\theta}\bar{Q})
\end{aligned}$$

where it has been used that Q commutes with itself. Also, an additional factor of i^2 has been introduced. The next step is rather indirect,

$$\begin{aligned}
&\exp(i^2(\xi^a\xi^{b*} + \xi^a\theta^{b*})(\sigma^\mu)_{ab}P_\mu)\exp(i((\xi+\theta)Q + \bar{\xi}\bar{Q}))\exp(i\bar{\theta}\bar{Q}) \qquad (4.33) \\
= &\exp(i^2(\xi^a\xi^{b*} + \xi^a\theta^{b*})(\sigma^\mu)_{ab}P_\mu)\exp(i(\xi+\theta)Q) \times \\
&\times \exp(i\bar{\xi}\bar{Q})\exp(-i(\xi^a\xi^{b*})(\sigma^\mu)_{ab}P_\mu)\exp(i\bar{\theta}\bar{Q}) \\
= &\exp(i^2\xi^a\theta^{b*}(\sigma^\mu)_{ab}P_\mu)\exp(i(\xi+\theta)Q)\exp(i(\bar{\xi}+\bar{\theta})\bar{Q}).
\end{aligned}$$

In the second-to-last step, the identity $\exp i(A+B) = \exp iA \exp iB \exp([A,B]/2)$ has been used, which follows from the Baker-Campbell-Hausdorff formula in this particular case by moving the commutator term on the other side. Also, it has been used that $\bar{Q}$ commutes with itself to combine the last two factors.

Hence, in total two consecutive operations amount to

$$U(\xi,\xi^*)U(\theta,\theta^*) = \exp(i^2\xi^a\theta^{b*}(\sigma^\mu)_{ab}P_\mu)\exp(i(\xi+\theta)Q)\exp(i(\bar{\xi}+\bar{\theta})\bar{Q}).$$

This result is not of the form $U(f(\xi,\xi^*),g(\theta,\theta^*))$, and thus the individual supertransformations do not form a group. This is not surprising, as the algebra requires that also the momentum operator must be involved. A better ansatz is thus

$$U(a_\mu,\xi,\xi^*) = \exp(iPa)\exp(i\xi Q)\exp(i\bar{\xi}\bar{Q}).$$

Since the momentum operator commutes with both Q and $\bar{Q}$, it follows directly that

$$\begin{aligned} U(a_\mu,\xi,\xi^*)U(b_\mu,\theta,\theta^*) &= \exp(i(a_\mu+b_\mu+i\xi^a\theta^{b*}(\sigma^\mu)_{ab})P_\mu)\exp(i(\xi+\theta)Q)\exp(i(\bar{\xi}+\bar{\theta})\bar{Q}) \\ &= U(a^\mu+b^\mu+i\xi^a\theta^{b*}(\sigma^\mu)_{ab},\xi+\theta,\xi^*+\theta^*), \end{aligned}$$

and consequently

$$U(x_\mu,\xi,\xi^*)U(a_\mu,\theta,\theta^*)\phi(0)U(x_\mu,\xi,\xi^*)^{-1}U(a_\mu,\theta,\theta^*)^{-1} = \phi(x^\mu+a^\mu+i\xi^a\theta^{b*}(\sigma^\mu)_{ab},\xi+\theta,\xi^*+\theta^*).$$

This then forms a group, as it should be, the group of supertranslations. The group is not Abelian, as a minus sign appears if θ and ξ are exchanged in the parameter for the momentum operator. However, the ordinary translations form an Abelian subgroup of this group of supertranslations. It is now also clear why there is a similarity to ordinary translations, compared to other unitary transformations: The latter only form a simple direct product group with ordinary translations, which is not so in case of the supertranslations. In this case, it is a semidirect product. This also justifies to call the parameters ξ and ξ^* supercoordinates in an abstract superspace.

The construction of these supertranslations, and by this the definition of fermionic supercoordinates and thus an abstract superspace, will now serve as a starting point for the construction of supersymmetric theories using this formalism.

4.5.2 Coordinate representation of supercharges

Ordinarily, infinitesimal translations $U(\epsilon_\mu)$ can be written in terms of a derivative

$$U(\epsilon_\mu) = \exp(-i\epsilon_\mu P^\mu) \approx 1 - i\epsilon_\mu P^\mu = 1 + \epsilon^\mu\partial_\mu. \tag{4.34}$$

If the correspondence of θ and θ^* should be taken seriously, such a differential representation for supertranslations should be possible. Hence for

$$U(\epsilon_\mu,\xi,\xi^*)\phi(x_\mu,\theta,\theta^*)U(\epsilon_\mu,\xi,\xi^*)^{-1}=\phi(x_\mu,\theta,\theta^*)+\delta\phi$$

with ϵ_μ and ξ and ξ^* all infinitesimal it should be possible to write for $\delta\phi$

$$\delta\phi=(\epsilon_\mu-i\theta^a\xi^{b*}(\sigma^\mu)_{ab})\partial_\mu\phi+\xi^a\frac{\partial\phi}{\partial\theta^a}+\xi^*_a\frac{\partial\phi}{\partial\theta^*_a}.$$

In analogy to (4.34), this implies that in the expression

$$\delta\phi=((\epsilon^\mu\partial_\mu-i\theta^a\xi^{b*}(\sigma^\mu)_{ab})\partial_\mu-i\xi^aQ_a-i\xi^*_aQ^{a\dagger})\phi$$

it is necessary to identify

$$Q_a = i\frac{\partial}{\partial\theta^a} \tag{4.35}$$

$$Q^{a\dagger} = i\frac{\partial}{\partial\theta^*_a}+\theta^b(\sigma^\mu)_{ba}\partial_\mu. \tag{4.36}$$

To form the second part of the $Q^\dagger$ part, it is necessary to note that $\xi^*_aQ^{a\dagger}=-\xi^{a*}Q^\dagger_a$, giving the overall sign.

This representation of Q and $Q^\dagger$ is only making sense, if these operators fulfill the corresponding algebra. In particular, Q and $Q^\dagger$ must commute with themselves. That is trivial in case of Q. Since θ and θ^* are independent variables, this is also the case for $Q^\dagger$. Furthermore, the anticommutation relation (4.12) has to be fulfilled[6]. This can be checked explicitly

$$\begin{aligned}\{Q_a,Q_b^\dagger\} &= \left\{i\frac{\partial}{\partial\theta^a},i\frac{\partial}{\partial\theta^*_b}+\theta^c(\sigma^\mu)_{cb}\partial_\mu\right\}\\ &= \left\{i\frac{\partial}{\partial\theta^a},i\frac{\partial}{\partial\theta^*_b}\right\}+\left\{i\frac{\partial}{\partial\theta^a},\theta^c(\sigma^\mu)_{cb}\partial_\mu\right\}\\ &= \left\{i\frac{\partial}{\partial\theta^a},\theta^c\right\}i(\sigma^\mu)_{cb}\partial_\mu\\ &= i(\sigma^\mu)_{ab}\partial_\mu=(\sigma^\mu)_{ab}P_\mu.\end{aligned}$$

where the fact that derivatives with respect to Grassmann variables anticommute has been used in going from the second to the third line and furthermore that

$$\partial_{\theta^a}(\theta^b\phi)+\theta^b\partial_{\theta^a}\phi=(\partial_{\theta^a}\theta^b)\phi+(\partial_{\theta^a}\phi)\theta^b+\theta^b\partial_{\theta^a}\phi=\delta^{ab}\phi+(\partial_{\theta^a}\phi)\theta^b-(\partial_{\theta^a}\phi)\theta^b,$$

[6] This is sufficient, as the commutator can be constructed from the anti-commutator, as has been done backwards in section 4.3.1.

using the anticommutativity of Grassmann numbers. Hence, the operators (4.35) and (4.36) are in fact a possible representation of the supersymmetry algebra, and there indeed exists a derivative formulation for these operators. This again emphasizes the strong similarity of the supersymmetry algebra and the translation algebra, once for fermionic and once for bosonic coordinates in the super space.

4.5.3 Supermultiplets

Now, given this superspace, the first question is what the vectors in this superspace represent. The simplest vector will have only components along one of the coordinates, which will be taken to be θ for now. Furthermore, these vectors are still functions. But their dependence on the Grassmann variables is by virtue of the properties of Grassmann numbers rather simple. Thus such a vector $\Phi(x,\theta)$ can be written as

$$\Phi(x,\theta)=\phi(x)+\theta\chi(x)+\frac{1}{2}\theta\theta F(x).$$

The θ-variables are still spinors, and the appearing products are still scalar products. Due to the antisymmetry of the scalar product, the last term does not vanish, though of course quantities like θ_a^2 vanish. The names for the component fields have been selected suggestively, but at the moment just represent arbitrary (bosonic or fermionic, complex) functions. Not withstanding, the set of the field (ϕ,χ,F) is called a chiral supermultiplet[7].

To justify the notation, it is sufficient to have a look at the infinitesimal transformation properties of Φ under $U(0,\xi,\xi^*)$. Of course, a non-zero translation parameter would just shift the x-arguments of the multiplet, and is thus only a notational complication. The change in the field is thus, as usual, given by

$$\begin{aligned}\delta\Phi &= (-i\xi^a Q_a - i\xi_a^* Q^{a\dagger})\Phi = (-i\xi^a Q_a + i\xi^{a*}Q_a^\dagger)\Phi\\ &= \left(\xi^a\frac{\partial}{\partial\theta^a}+\xi^{a*}\frac{\partial}{\partial\theta^{a*}}+i\xi^{a*}\theta^b(\sigma^\mu)_{ba}\partial_\mu\right)\left(\phi(x)+\theta^c\chi_c+\frac{1}{2}\theta^c\theta_c F\right)\\ &= \delta_\xi\phi+\theta^a\delta_\xi\chi_a+\frac{1}{2}\theta^a\theta_a\delta_\xi F,\end{aligned}$$

where the last line is by definition the change in the individual components of the superfield. Since Φ is not depending on θ^{a*}, the derivative with respect to this variable can be dropped. It thus only remains to order the result by powers in θ.

At order zero, a contribution can only come from the action of the θ-derivative on the χ term. This yields

$$\delta_\xi\phi=\xi\chi$$

[7] Actually a left one, as only the left-handed spinor χ appears.

for the transformation rule of the ϕ-field.

At order one appears

$$\delta_\xi \chi_a = \xi_a F - i\xi^{b*}(\sigma^\mu)_{ab}\partial_\mu\phi = \xi_a F - i(i\sigma_2\xi^*)_b(\sigma^\mu)_{ab}\partial_\mu\phi,$$

giving the transformation for the χ-field.

Finally, because θ^a are Grassmann variables, order two is the highest possible order, yielding

$$\begin{aligned}
\frac{1}{2}\theta\theta\delta_\xi F &= i\xi^{b*}\theta^a(\sigma^\mu)_{ab}\theta^c\partial_\mu\chi_c = -\frac{1}{2}\theta\theta\epsilon^{ac}i\xi^{b*}(\sigma^\mu)_{ab}\partial_\mu\chi_c \\
&= -\frac{1}{2}\theta\theta i\xi^{a*}(\sigma^{T\mu})_{ab}\epsilon^{bc}\partial_\mu\chi_c \\
&= -\frac{1}{2}\theta\theta i(i\sigma_2\xi^*)_a(\sigma^{T\mu})\epsilon^{bc}\partial_\mu\chi_c \\
&= -\frac{1}{2}\theta\theta i\xi^*_d(i\sigma_2)_{da}(\sigma^{T\mu})_{ab}(i\sigma_2)_{bc}\partial_\mu\chi_c \\
&= \frac{1}{2}\theta\theta(-i)\xi^*_d(\bar\sigma^\mu)_{dc}\partial_\mu\chi_c \\
&= \frac{1}{2}\theta\theta(-i)\xi^\dagger(\bar\sigma^\mu)\partial_\mu\chi
\end{aligned}$$

where it has been used that

$$\theta^a\theta^b = -\frac{1}{2}\epsilon^{ab}\theta^c\theta_c. \tag{4.37}$$

This can be shown simply by explicit calculation, keeping in mind that $\theta^{a2} = 0$ and $\epsilon^{12} = 1 = -\epsilon_{12}$, as well as $(i\sigma_2)_{cd} = \epsilon^{cd}$ and $\bar\sigma^\mu = i\sigma_2\sigma^{T\mu}i\sigma_2$.

Thus, the transformation rules for the field components are

$$\begin{aligned}
\delta_\xi\phi &= \xi\chi \\
\delta_\xi\chi_a &= \xi_a F - i(i\sigma_2\xi^*)_b(\sigma^\mu)_{ab}\partial_\mu\phi \\
\delta_\xi F &= -i\xi^\dagger(\bar\sigma^\mu)\partial_\mu\chi.
\end{aligned} \tag{4.38}$$

These are exactly the transformation rules obtained in section 4.3.4. Thus, a chiral supermultiplet under a supertranslation transforms in exactly the same way as the field content of a free (or interacting, in case of the Wess-Zumino model) supersymmetric theory. This result already indicates that supersymmetric theories can possibly be formed by using scalars with respect to supertranslation in much the same way as ordinary theories are built from scalars with respect to Lorentz transformations. The following will show that this is indeed the case. Actually, this is not exactly the way it turns out, but the idea is similar.

4.5.4 Other representations

All of this can be repeated, essentially unchanged, for a right chiral multiplet, where the dependence on θ is replaced by the one on θ^*. Including both, θ and θ^* actually does not provide something new, but just leads to a reducible representation. That is most easily seen by considering the condition

$$\frac{\partial}{\partial\theta_a^*}\Phi(x,\theta,\theta^*)=0.$$

If this condition applies to the field Φ then so does it to the transformed field $\delta\Phi$, since in the latter expression

$$\frac{\partial}{\partial\theta_c^*}\delta\Phi=\frac{\partial}{\partial\theta_c^*}\left(-i\theta^a(\sigma^\mu)_{ab}\xi^{b*}\partial_\mu\Phi+\xi^a\frac{\partial\Phi}{\partial\theta^a}+\xi_a^*\frac{\partial\Phi}{\partial\theta_a^*}\right)$$

at no point an additional dependence on θ^* is introduced. Hence, the fields $\Phi(x,\theta)$ form an invariant subgroup of the SUSY transformation, and likewise do $\Phi(x,\theta^*)$. Any representation including a dependence on θ and θ^* can thus be only a reducible one. Nonetheless, this representation is useful, as it will permit to construct a free supersymmetric theory.

Before investigating this possibility, one question might arise. To introduce the supervectors Φ, the particular supertranslation operator

$$U_I(x,\theta,\theta^*)=\exp(ixP)\exp(i\theta Q)\exp(i\bar\theta\bar Q),$$

called type I, has been used. Would it not also have been possible to use the operators

$$\begin{aligned}U_{II}(x,\theta,\theta^*)&=\exp(ixP)\exp(i\bar\theta\bar Q)\exp(i\theta Q)\\ U_r(x,\theta,\theta^*)&=\exp(ixP)\exp(i\theta Q+i\bar\theta\bar Q)?\end{aligned}$$

The answer to this question is, in fact, yes. Both alternatives could have been used. And these would have generated different translations in the field. In fact, when using the expression $\Phi_i=U_i\Phi(0,0,0)U_i$ to generate alternative superfields the relations

$$\Phi_r(x,\theta,\theta^*)=\Phi_I\left(x^\mu-\frac{1}{2}i\theta^a\sigma^\mu_{ab}\theta^{b*},\theta,\theta^*\right)=\Phi_{II}\left(x^\mu+\frac{1}{2}i\theta^a\sigma^\mu_{ab}\theta^{b*},\theta,\theta^*\right)\qquad(4.39)$$

would have been found. An explicit expression, e. g. in case of the left supermultiplet would be

$$\Phi_r(x,\theta,\theta^*)=\phi+\theta\chi+\frac{1}{2}\theta\theta F-\frac{1}{2}i\theta^a\sigma^\mu_{ab}\theta^{b*}\partial_\mu\phi+\frac{1}{4}i(\theta\theta)(\partial_\mu\chi^a\sigma^\mu_{ab}\theta^{b*})-\frac{1}{16}(\theta\theta)(\bar\theta\bar\theta)\partial^2\phi,\quad(4.40)$$

which is obtained by the Taylor-expansion in θ from the relation (4.39).

I. e., the different supertranslations lead to unitarily equivalent supervectors, which can be transformed into each other by conventional translations, and thus a unitary transformation. This already suggests that these are just unitarily equivalent representations of the same algebra. This can be confirmed: For each possibility it is possible to find a representation in terms of derivatives which always fulfills the SUSY algebra. Hence, all of these representations are equivalent, and one can choose freely the most appropriate one, which in the next section will be the type-I version.

4.5.5 Constructing interactions from supermultiplets

For now, lets return to the (left-)chiral superfield, and the type-I transformations.

Inspecting the transformation rules under a supertranslation of a general superfield one thing is of particular importance: The transformation of the F-component of the superfield, (4.38), corresponds to a total derivative. As a supertranslation is nothing else than a SUSY transformation this implies that any term which is constructed from the F-component of a superfield will leave the action invariant under a SUSY transformation. Hence, the question arises how to isolate this F-term, and whether any action can be constructed out of it.

The first question is already answered: This can be done by twofold integration. Due to the rules for analysis of Grassmann variables, it follows that

$$\int d\theta_1 \int d\theta_2 \Phi = \int d\theta_1 \int d\theta_2 \left(\phi + \theta^a \chi_a + \frac{1}{2}\theta^a \theta_a F\right) = \int d\theta_1 (\chi_2 + \theta_1 F) = F.$$

Thus twofold integration is isolating exactly this part.

The answer to the second question is less obvious. A single superfield will only contribute a field F. That is not producing a non-trivial theory. However, motivated by the construction of Lorentz invariants by building scalar products, the simplest idea is to consider a product of two superfields. This yields a superscalar

$$\Phi\Phi = \left(\phi + \theta\chi + \frac{1}{2}\theta\theta F\right)^2 = \phi^2 + 2\phi\theta\chi + \theta\theta\phi F + (\theta\chi)(\theta\chi),$$

and all other terms vanish, as they would be of higher order in θ. Using (4.37), the last term can be rearranged to yield

$$\theta^a \chi_a \theta^b \chi_b = -\theta^a \theta^b \chi_a \chi_b = \frac{1}{2}\theta\theta \epsilon^{ab} \chi_a \chi_b = -\frac{1}{2}\theta\theta \chi^a \chi_a = \frac{1}{2}(\theta\theta)(\chi\chi), \tag{4.41}$$

and thus

$$\Phi\Phi = \phi^2 + 2\phi\theta\chi + \frac{1}{2}\theta\theta(2F\phi - \chi\chi).$$

Isolating the F-term and multiplying with a constant $M/2$, which is not changing the property of being a total derivative under SUSY transformations, yields

$$W_2 = M\phi F - \frac{M}{2}\chi\chi.$$

However, this is a well-known quantity, it is exactly the terms quadratic in the fields in the Lagrangian (4.29) which are not part of the free Lagrangian, and to which one has to add, of course, also the hermitian conjugate to obtain a real action, even though both are separately SUSY invariant.

The fact that only terms of order two in the fields could have been obtained is clear, as only a square was evaluated. It is surprising at first sight that in fact all terms quadratic to this order have been obtained. However, in the construction of the Wess-Zumino Lagrangian the minimal set has been searched for, and this here is then the minimal set. Note then that this implies that when stopping at this point, the Lagrangian for a massive, but otherwise free, supersymmetric theory has been obtained, as can be checked by integrating out F. Note that obtaining the free part of the Lagrangian, or a massless supersymmetric theory, can also be performed by the present methods, but requires some technical complications to be discussed later.

This suggests that it should be possible to obtain the full interaction part of the Wess-Zumino model by also inspecting the product of three superfields, as this is the highest power in fields appearing in (4.29). And in fact, evaluating the F-component of such a product yields

$$\begin{aligned}\int d^2\theta\Phi^3 &= 3\phi^2 F - \phi\chi\chi + 2\phi\int d^2\theta(\theta\chi)(\theta\chi)\\ &= 3\phi^2 F - \phi\chi\chi + 2\phi\int d^2\theta(-\frac{1}{2}\theta\theta)(\chi\chi) = 3\phi^2 F - 3\phi\chi\chi.\end{aligned}$$

Multiplying this with $y/6$ exactly produces the terms which are cubic in the fields in the Wess-Zumino Lagrangian (4.29). Thus, this Lagrangian could equally well be written as

$$\mathcal{L} = \partial_\mu\phi\partial^\mu\phi^\dagger + \chi^\dagger i\bar{\sigma}^\mu\partial_\mu\chi + F^\dagger F + \int d^2\theta(W + W^\dagger)$$

with

$$W = \frac{M}{2}\Phi\Phi + \frac{y}{6}\Phi\Phi\Phi,$$

being the superpotential. In contrast to the one introduced in section 4.4.1 this superpotential now includes all interactions of the theory, not only the ones involving the ϕ terms. Thus, it was indeed possible to construct the supersymmetric theory just by usage of products of superfields. This construction principle carries over to more complex situations, and can be considered as the construction principle for supersymmetric theories.

The generalization to fields with an internal degree of freedom, like flavor or charge, is straightforward, yielding a superpotential of type

$$W = \frac{M_{ij}}{2}\Phi_i\Phi_j + \frac{y_{ijk}}{6}\Phi_i\Phi_j\Phi_k,$$

just as in the case of the original treatment of the Wess-Zumino model.

In principle, the free part can be constructed in a similar manner. However, it turns out to be surprisingly more complicated in detail. While so far only the field $\Phi_I(x,\theta,0)$ has been used, for the free part it will be necessary to use in addition the quantity $\Phi_r(x,\theta,\theta^*)$. The free part can then be obtained from (4.40) by the expression

$$\begin{aligned}\int d^4\theta\Phi_r(x,\theta,\theta^*)^\dagger\Phi_r(x,\theta,\theta^*) &= -\frac{1}{4}\phi^\dagger\partial^2\phi - \frac{1}{4}\partial^2\phi\phi^\dagger + F^\dagger F \\ &+ \int d^4\theta\frac{1}{4}((i(\bar\chi\bar\theta)(\theta\theta)\partial_\mu\chi^{a\dagger}\sigma^\mu_{ab}\theta^{b*} - i\theta^a\sigma^\mu_{ab}\partial_\mu\chi^b(\bar\theta\bar\theta)(\theta\chi) \\ &+\partial_\mu\phi^\dagger\theta^a\sigma^\mu_{ab}\theta^{b*}\theta^c\sigma^\nu_{cd}\theta^{d*}\partial_\nu\phi)\end{aligned}$$

Since the action is the only quantity of interest, partial integrations are permitted. Thus $\phi^\dagger\partial^2\phi = -\partial\phi^\dagger\partial\phi$. Furthermore, applying (4.41) twice yields the relation

$$(\bar\theta\bar\chi)(\bar\theta\bar\chi) = -\frac{1}{2}(\bar\theta\bar\theta)(\bar\chi\bar\chi),$$

which, of course holds similarly for unbarred quantities. This can be used to reformulate the second line to isolate products of $\theta^2\bar\theta^2$, and also for the third line. These manipulations together yield

$$\begin{aligned}\int d^4\theta\Phi_r(x,\theta,\theta^*)^\dagger\Phi_r(x,\theta,\theta^*) &= \frac{1}{2}\partial_\mu\phi^\dagger\partial^\mu\phi + F^\dagger F + i\bar\chi\bar\sigma^\mu\partial_\mu\chi + \frac{1}{2}\partial_\mu\phi^\dagger\partial^\mu\phi \\ &= \partial_\mu\phi^\dagger\partial^\mu\phi + i\bar\chi\bar\sigma^\mu\partial_\mu\chi + F^\dagger F,\end{aligned}$$

which is exactly the Lagrangian of the free supersymmetric theory, or, more precisely, the integration kernel of the action. Hence, the complete Wess-Zumino model Lagrangian can be written as

$$\mathcal{L} = \int d^4\theta\Phi_r^\dagger\Phi_r + \int d^2\theta\left(\frac{M}{2}\Phi_I\Phi_I + \frac{y}{6}\Phi_I\Phi_I\Phi_I\right) + \int d^2\bar\theta\left(\frac{M}{2}\Phi_{II}\Phi_{II} + \frac{y}{6}\Phi_{II}\Phi_{II}\Phi_{II}\right),$$

which is in fact now an expression where supersymmetry is manifest, and the last term just creates the Hermitiean conjugate of the second-to-last term to have a hermitiean action.

4.6 Supersymmetric gauge theories

All relevant theories in particle physics are actually not of the simple type consisting only out of scalars and fermions, but are gauge theories. Therefore, to construct the standard model it is necessary to work with supersymmetric gauge theories. In the following the superspace formalism is actually less practically useful, so it will not be used.

4.6.1 Supersymmetric Maxwell theory

The simplest gauge theory to construct is a supersymmetric version of Maxwell theory. The photon is a boson with spin 1. Hence its super-partner, the photino, has to be a fermion. The photon has on-shell two degrees of freedom, so the photino has to be a Weyl fermion. Off-shell, however, the photon has three degrees of freedom, corresponding to the three different magnetic quantum numbers possible. So another auxiliary degree of freedom is necessary to cancel all fermionic degrees of freedom. This other off-shell bosonic degree of freedom will be the so-called D field.

Will this be a flavor of quantum electron dynamics then, just with the Dirac electron replaced by a Weyl one and one field added? The answer to this is a strict no. Since the supersymmetry transformation just acts on the statistical nature of particles it cannot change an uncharged photon into a charged photino. Thus, the photino has also to have zero charge, as does the D boson. The simplest supersymmetric gauge theory is then the free supersymmetric Maxwell theory, as there are no interactions possible between uncharged particles. Its form has thus to be of the type

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\lambda^\dagger\bar{\sigma}^\mu\partial_\mu\lambda + \frac{1}{2}D^2. \tag{4.42}$$

Note that the absence of charge also implies that no covariant derivative can appear. Hence, the photino λ has to be invariant under a gauge transformation, as D has to be. Thus the gauge dynamics is completely contained in the photon field.

To construct the SUSY transformation for A_μ, it is necessary that is has to be real, as A_μ is a real field. Furthermore, it has to be a Lorentz vector. Finally, since it is an infinitesimal transformation it has to be at most linear in the transformation parameter ξ. The simplest quantity fulfilling these properties is

$$\delta_\xi A_\mu = \xi^\dagger\bar{\sigma}_\mu\lambda + \lambda^\dagger\bar{\sigma}_\mu\xi.$$

As noted, the photino is a gauge scalar. It can therefore not be directly transformed with the field A_μ, but a gauge-invariant combination of A_μ is necessary. The simplest such quantity is $F_{\mu\nu}$. To absorb the two indices, and at the same time provide the correct transformation properties under Lorentz transformation, the SUSY transformations should be of the form

$$\begin{aligned}\delta_\xi\lambda &= \frac{i}{2}\sigma^\mu\bar{\sigma}^\nu\xi F_{\mu\nu}+\xi D\\ \delta_\xi\lambda^\dagger &= -\frac{i}{2}\xi^\dagger\bar{\sigma}^\nu\sigma^\mu F_{\mu\nu}+\xi^\dagger D,\end{aligned}$$

where the pre-factor has been chosen with hindsight. In principle, it could also be determined a-posterior, provided that otherwise these simplest forms work. The terms containing the D field have been added in analogy with the Wess-Zumino model. Finally, the SUSY transformation for the D-boson has to vanish on-shell, and thus should be proportional to the equations of motion of the other fields. Furthermore, it is a real field, and thus its transformation rule has to respect this, similarly as for the photon. In principle, its transformation could depend on the equations of motions of both the photon and the photino. However, inspired by the properties of the F boson in the Wess-Zumino model the ansatz is one depending only on the equations of motion of the photino, which is indeed sufficient. The transformation rule such constructed is

$$\delta_\xi D=-i(\xi^\dagger\bar{\sigma}^\mu\partial_\mu\lambda-\partial_\mu\lambda^\dagger\bar{\sigma}^\mu\xi),$$

which has all of the required properties.

It remains to demonstrate that these are the correct transformation rules and that the theory is supersymmetric. Since ξ is taken to be infinitesimal and Grassmann, it is sufficient to evaluate the transformations once more only up to an order linear in ξ.

The transformation of the photon term yields

$$\begin{aligned}-\frac{1}{4}\delta_\xi(F_{\mu\nu}F^{\mu\nu}) &= -\frac{1}{4}((\delta_\xi F_{\mu\nu})F^{\mu\nu}+F_{\mu\nu}(\delta_\xi F^{\mu\nu}))=-\frac{1}{2}F_{\mu\nu}\delta_\xi F^{\mu\nu}\\ &= -\frac{1}{2}F^{\mu\nu}(\partial_\mu\delta_\xi A_\nu-\partial_\nu\delta_\xi A_\mu)=-F_{\mu\nu}\partial^\mu\delta_\xi A^\nu=-F_{\mu\nu}\partial^\mu(\xi^\dagger\bar{\sigma}^\nu\lambda+\lambda^\dagger\bar{\sigma}^\nu\xi),\end{aligned}$$

where the antisymmetry of $F_{\mu\nu}$ has been used. The only term which can cancel this is the one part from the transformation of the spinors being proportional to $F_{\mu\nu}$. This contribution is

$$\begin{aligned}\delta_\xi^{F_{\mu\nu}}(i\lambda^\dagger\bar{\sigma}^\mu\partial_\mu\lambda) &= i(\delta_\xi\lambda^\dagger)\bar{\sigma}^\mu\partial_\mu\lambda+i\lambda^\dagger\bar{\sigma}^\mu\partial_\mu(\delta_\xi\lambda)\\ &= \frac{1}{2}\left(\xi^\dagger\bar{\sigma}^\nu\sigma^\mu F_{\mu\nu}\bar{\sigma}^\rho\partial_\rho\lambda+(\partial_\rho\lambda^\dagger)\bar{\sigma}^\rho\sigma^\mu\bar{\sigma}^\nu\xi F_{\mu\nu}\right).\end{aligned}$$

The structure of this term is already quite similar, but the product of three σs is different. However, since the σ are Pauli matrices, it is always possible to rewrite them in terms of single ones,

$$\bar{\sigma}^\mu\sigma^\nu\bar{\sigma}^\rho=g^{\mu\nu}\bar{\sigma}^\rho-g^{\mu\rho}\bar{\sigma}^\nu+g^{\nu\rho}\bar{\sigma}^\mu-i\epsilon^{\mu\nu\rho\sigma}\bar{\sigma}_\sigma, \tag{4.43}$$

where ϵ is totally antisymmetric. This simplifies the expression at lot. The contraction $g^{\mu\nu}F_{\mu\nu}$ vanishes, since g is symmetric and $F_{\mu\nu}$ is antisymmetric. Because

$$F_{\mu\nu}\partial_\rho\lambda = -\lambda(\partial_\rho\partial_\mu A_\nu - \partial_\rho\partial_\nu A_\mu),$$

this expression is symmetric in two indices. Thus any contraction with the ϵ-tensor also vanishes. Thus, the expression reduces to

$$\begin{aligned} & \xi^\dagger F_{\mu\nu}(-g^{\nu\rho}\bar{\sigma}^\mu + g^{\mu\rho}\bar{\sigma}^\nu)\partial_\rho\lambda + (\partial_\rho\lambda^\dagger)(g^{\mu\rho}\bar{\sigma}^\nu - g^{\nu\rho}\bar{\sigma}^\mu)\xi F_{\mu\nu} \\ = \; & -2F_{\mu\nu}\xi^\dagger\bar{\sigma}^\mu\partial^\nu\lambda + 2(\partial^\mu\lambda^\dagger)\bar{\sigma}^\nu F_{\mu\nu} = 2F_{\mu\nu}(\xi^\dagger\bar{\sigma}^\nu\partial^\mu\lambda + \partial^\mu\lambda^\dagger\bar{\sigma}^\nu\xi). \end{aligned}$$

This precisely cancels the contribution from the photon transformation, when combined with the factor 1/2.

The expressions involving D are simpler. The contribution from the photino term is

$$\delta_\xi^D(i\lambda^\dagger\bar{\sigma}^\mu\partial_\mu\lambda) = iD\xi^\dagger\bar{\sigma}^\mu\partial_\mu\lambda - i\partial_\mu\lambda^\dagger\bar{\sigma}^\mu\xi D,$$

which cancels against the contribution from the D term

$$\delta_\xi\frac{1}{2}D^2 = D\delta_\xi D = -iD(\xi^\dagger\bar{\sigma}^\mu\partial_\mu\lambda - \partial_\mu\lambda^\dagger\bar{\sigma}^\mu\xi).$$

Thus the Lagrangian (4.42) describes indeed a supersymmetric theory, consisting of the non-interacting photon, the photino, and the D-boson.

Of course, it would once more be possible to construct the theory just from the D-component of a super-vector. This will not be done here.

4.6.2 Supersymmetric Yang-Mills theory

A much more interesting theory will be the supersymmetric version of the non-Abelian gauge theory, again neglecting the matter part. The first thing to do is to count again degrees of freedom. The gauge field is in the adjoint representation. For SU(N) as a gauge group there are therefore $N^2 - 1$ independent gauge fields. On-shell, this has to be canceled by exactly the same number of fermionic degrees of freedom, so the same number of fermions, called gauginos. Quarks or electrons are in the fundamental representation of the gauge group, and thus there are a different number, e. g. N for SU(N). This cannot match, and the super-partners of the gauge fields, the gauginos, have therefore to be also in the adjoint representation of the gauge group. Therefore, they are completely different from ordinary matter fields like quarks or leptons. Hence, for each field in the standard model below it will be necessary to introduce an independent superpartner.

Of course, to close the SUSY algebra also off-shell, it will again be necessary to introduce additional scalars. However, also these have to be in the adjoint representation of the gauge group. In this case, it is useful to write the Lagrangian explicitly in the index form. This yields the Lagrangian

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a + i\lambda^\dagger_a\bar{\sigma}^\mu D^{ab}_\mu\lambda_b + \frac{1}{2}D^aD_a \\ F^a_{\mu\nu} &= \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - ef^{abc}A^b_\mu A^c_\nu \\ D^{ab}_\mu &= \delta^{ab}\partial_\mu + ef^{abc}A^c_\mu,\end{aligned} \tag{4.44}$$

where f^{abc} are the structure constants of the gauge group and D^{ab}_μ is the covariant derivative in the adjoint representation. The gauge bosons transform in the usual way, but the gaugino and the D-boson transform under gauge transformations in the adjoint representation. I. e., they transform like the gauge field, except without the inhomogeneous term as $g\lambda g^{-1}$, when written as algebra elements. Thus, even the D^2 term is gauge invariant, as no derivatives are involved.

Making the ansatz

$$\begin{aligned}\delta_\xi A^a_\mu &= \xi^\dagger\bar{\sigma}_\mu\lambda^a + \lambda^{a\dagger}\bar{\sigma}_\mu\xi \\ \delta_\xi\lambda^a &= \frac{i}{2}\sigma^\mu\bar{\sigma}^\nu\xi F^a_{\mu\nu} + \xi D^a \\ \delta_\xi\lambda^{a\dagger} &= -\frac{i}{2}\xi^\dagger\bar{\sigma}^\nu\sigma^\mu F^a_{\mu\nu} + \xi^\dagger D^a \\ \delta_\xi D^a &= -i(\xi^\dagger\bar{\sigma}^\mu D^{ab}_\mu\lambda^b - D^{ab}_\mu\lambda^\dagger\bar{\sigma}^\mu\xi)\end{aligned}$$

as the most simple generalization of the Abelian case is actually working. The reason for this is simple: After expressing everything in components, all quantities (anti-)commute. Furthermore, under partial integration

$$\begin{aligned}D^{ab}_\mu\lambda^b F^{\mu\nu a} &= (\partial_\mu\lambda^a + ef^{abc}A^c_\mu\lambda^b)F^{\mu\nu a} = -\lambda^a\partial_\mu F^{\mu\nu a} - ef^{bac}A^c_\mu\lambda^b F^{\mu\nu a} \\ &= -\lambda^a\partial_\mu F^{\mu\nu a} - ef^{abc}A^c_\mu\lambda^a F^{\mu\nu b} = -\lambda^a D^{ab}_\mu F^{\mu\nu b}.\end{aligned} \tag{4.45}$$

Thus, all manipulations performed in section 4.6.1 can be performed equally well in case of the non-Abelian theory. Therefore, only those terms which appear in addition to the Abelian case have to be checked.

The most simple is the modification of the D-term by the appearance of the covariant derivative. These terms trivially cancel with the corresponding gaugino term just as in the Abelian case, as there also a covariant derivative appears

$$\begin{aligned}\delta^D_\xi(i\lambda^\dagger_a\bar{\sigma}^\mu D^{ab}_\mu\lambda_b) &= iD_a\xi^\dagger\bar{\sigma}^\mu D^{ab}_\mu\lambda_b - iD^{ab}_\mu\lambda^\dagger_a\bar{\sigma}^\mu\xi D_b \\ \delta_\xi\frac{1}{2}D^2 = D_a\delta_\xi D^a &= -iD_a(\xi^\dagger\bar{\sigma}^\mu D^{ab}_\mu\lambda_b - D^{ab}_\mu\lambda^\dagger_b\bar{\sigma}^\mu\xi).\end{aligned}$$

The next contribution stems from the appearance of the covariant derivative in the gaugino term. The contribution form the gauge boson-gaugino coupling cancels with the contribution from the self-coupling of the gauge bosons. This contribution alone gives

$$\begin{aligned}
&-\frac{1}{2}F^a_{\mu\nu}\delta^{AA}_\xi F^{\mu\nu}_a+\delta^\lambda_\xi(i\lambda^\dagger_a\bar\sigma_\mu e f^{abc}A^\mu_c\lambda_b)\\
&=\frac{ef^{abc}}{2}F^a_{\mu\nu}((\xi^\dagger\bar\sigma^\mu\lambda_b+\lambda^\dagger_b\bar\sigma^\mu\xi)A^c_\nu+A^b_\mu(\xi^\dagger\bar\sigma^\nu\lambda_c+\lambda^\dagger_c\bar\sigma^\nu\xi))\\
&+\frac{ef^{abc}}{2}F^{\rho\sigma}_a(\xi^\dagger\bar\sigma_\rho\sigma_\sigma\bar\sigma_\mu A^\mu_c\lambda_b-\lambda^\dagger_b\bar\sigma_\mu\sigma_\rho\bar\sigma_\sigma\xi A^\mu_c)
\end{aligned}\tag{4.46}$$

The last expression can be reformulated using (4.43). Since the following proceeds identically for the contributions proportional to λ and $\lambda^\dagger$, only the former will be investigated explicitly. Applying therefore (4.43) to the third contribution yields

$$\frac{ef^{abc}}{2}F^{\rho\sigma}_a\xi^\dagger(g_{\rho\sigma}\bar\sigma_\mu-g_{\rho\mu}\bar\sigma_\sigma+g_{\sigma\mu}\bar\sigma_\rho-i\epsilon_{\rho\sigma\mu\delta}\bar\sigma_\delta)\lambda_b A^\mu_c.$$

The first term yields zero, as the trace of F vanishes. The contribution with the ϵ-tensor vanishes, as $f^{abc}F^{\rho\sigma}_a A^\mu_c$ is symmetric in the three Lorentz indices, and therefore also vanishes upon contraction with the ϵ-tensor, as an explicit calculation shows. The remaining two terms then exactly cancel the two terms from the transformation of $F^a_{\mu\nu}$, just by relabeling the Lorentz indices, and shifting them appropriately up and down. Therefore, also this contribution is not violating supersymmetry.

Then, only the term from the transformation of the gauge boson in the covariant derivative coupling to the gauginos remains. Its transformation is

$$\delta^A_\xi(i\lambda^\dagger_a\bar\sigma_\mu e f^{abc}A^\mu_c\lambda_c)=ief^{abc}\lambda^\dagger_a\bar\sigma^\mu(\xi^\dagger\bar\sigma_\mu\lambda_c+\lambda^\dagger_c\bar\sigma_\mu\xi)\lambda_b.$$

This contribution contains λ cubed, and can therefore not be canceled by any other contribution. However, rewriting the first term in explicit index notation yields

$$ief^{abc}\lambda^*_{ia}\bar\sigma^\mu_{ij}\lambda_{bj}\lambda^*_{ck}\bar\sigma_{\mu kl}\xi_l.$$

The expression $\bar\sigma^\mu_{ij}\bar\sigma_{\mu kl}$ is symmetric in the first and third index for each term individually, and the expression $\lambda^*_{ia}\lambda^*_{ck}$ is antisymmetric in exact these two indices. Therefore, this contribution drops out, and similarly for the second contribution.

Thus, all in all this theory is supersymmetric and therefore the supersymmetric generalization of Yang-Mills theory, called often super-Yang-Mills theory, or SYM, for short.

4.6.3 Supersymmetric QED

The Abelian gauge theory contained only an uncharged fermion, the photino. To obtain a supersymmetric version of QED a U(1)-charged fermion is necessary. Since this cannot be introduced into the vector supermultiplet of the photon, the most direct way to introduce it is by the addition of a chiral supermultiplet which will be coupled covariantly to the vector supermultiplet. This introduces only a Majorana electron, but this will be sufficient for the beginning. Of course, compensating scalar fields to make the theory supersymmetric will be required. Hence, at least a combination of supersymmetric Maxwell theory and a Wess-Zumino theory is required.

The minimally coupled version is

$$\mathcal{L} = (D_\mu\phi)^\dagger D^\mu\phi + i\chi^\dagger\bar{\sigma}^\mu D_\mu\chi + F^\dagger F - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\lambda^\dagger\bar{\sigma}^\mu\partial_\mu\lambda + \frac{1}{2}D^2. \tag{4.47}$$

This theory contains now the photon, its super-partner the photino, the (Majorana) electron, its super-partner the selectron, and the auxiliary bosons F and D. In essence, this Lagrangian is the sum of the non-interacting Wess-Zumino model and the supersymmetric Abelian gauge theory, where in the former the derivatives have been replaced by gauge covariant derivatives. This implies that all newly added fields, ϕ, χ, and F, are charged, and transform under gauge transformations. Only the photino and the D-boson remain uncharged.

Of course, also the derivatives appearing in the supersymmetry transformations of electron, the selectron, and the F-boson have to be replaced by their covariant counterparts. Thus the minimal set of rules for the matter sector reads

$$\begin{aligned}
\delta_\xi\phi &= \xi\chi \\
\delta_\xi\chi &= \sigma^\mu\sigma_2\xi^* D_\mu\phi + \xi F \\
\delta_\xi F &= -i\xi^\dagger\bar{\sigma}^\mu D_\mu\chi
\end{aligned}$$

while those for the gauge sector are left unchanged

$$\begin{aligned}
\delta_\theta A_\mu &= \theta^\dagger\bar{\sigma}_\mu\lambda + \lambda^\dagger\bar{\sigma}_\mu\theta \\
\delta_\theta\lambda &= \frac{i}{2}\sigma^\mu\bar{\sigma}^\nu\theta F_{\mu\nu} + \theta D \\
\delta_\theta D &= -i(\theta^\dagger\bar{\sigma}^\mu\partial_\mu\lambda - \partial_\mu\lambda^\dagger\bar{\sigma}^\mu\theta)
\end{aligned}$$

However, even with (4.45), it can then be shown that this theory is not yet supersymmetric under these transformation. The reason is that supersymmetry requires additional interactions as will be discussed below. The Lagrangian (4.47) is indeed, in contrast to ordinary QED, not the most general[8] one which can be written with this field content and which is supersymmetric, gauge-invariant, and renormalizable. It is possible to add

[8]Here, and in the following, Wess-Zumino-like couplings, which are in fact gauge-invariant and supersymmetric, will be ignored.

additional interactions which do have all of these properties. However, this will require not only a fixed constant of proportionality between ξ and θ, but also a minor change in the transformation of the F field.

Further interactions between the matter and the gauge sector have to be gauge-invariant, and will also be chosen to be perturbatively renormalizable. These two requirements already limit the number of possible terms severely to the interaction terms $\phi^\dagger\chi\lambda$, $\lambda^\dagger\chi^\dagger\phi$, and $\phi^\dagger\phi D$. All other terms are either not gauge-invariant, not Lorentz-invariant or not at the same time (superficially) renormalizable, like terms involving F. Hence the interaction Lagrangian takes the form

$$A(\phi^\dagger\chi\lambda + \lambda^\dagger\chi^\dagger\phi) + B\phi^\dagger\phi D.$$

Checking all terms for their invariance under supersymmetry transformations is a long and tedious exercise, which will not be performed here. Only those elements will be presented, which influence either the values of A and B, or modify the supersymmetry transformations themselves.

The first step is to check all contributions from the transformed part of the interaction Lagrangian which are linear in the D-field. These will occur either from transformations of the photino λ or from the term proportional to B when either of the other fields are transformed. This yields

$$A(\phi^\dagger\chi\theta D + \theta^\dagger\chi^\dagger D\phi) + B(\chi^\dagger\xi^\dagger\phi D + \phi^\dagger\xi\chi D).$$

No other contribution in the Lagrangian will produce such terms which couple the matter fields to the D-boson. Thus, these have to cancel by themselves. This is only possible if

$$A\theta = -B\xi, \tag{4.48}$$

yielding already a constraint for the transformation parameters. Thus, in contrast to the case when two gauge sectors are coupled, coupling two supersymmetric sectors cannot be done independently. The reason is again that both supersymmetry transformations are tied to the momentum transformation, thus not permitting to leave them independent.

The transformation of the ϕ-fields in the A-term will yield terms having only fermionic degrees of freedom,

$$A((\chi^\dagger\xi^\dagger)(\chi\lambda) + (\lambda^\dagger\chi^\dagger)(\xi\chi)).$$

The only other term which can generate such a combination of the electron and the photino field stems from the photon-electron coupling term, which reads

$$-q\chi^\dagger\bar{\sigma}^\mu\chi(\delta_\theta A_\mu) = -q\chi^\dagger\bar{\sigma}^\mu\chi(\theta^\dagger\bar{\sigma}_\mu\lambda + \lambda^\dagger\bar{\sigma}_\mu\theta).$$

This has the same field content as the previous contribution, when the relation between ξ and θ is used, but not the same Lorentz structure. To recast the expression, the identity

$$\begin{aligned} & (\chi^\dagger\bar\sigma^\mu\chi)(\lambda^\dagger\bar\sigma_\mu\theta) = \chi_a^*\bar\sigma^\mu_{ab}\chi_b\lambda_c^*\bar\sigma_{\mu cd}\theta_d \\ = \quad & \chi_a^*\chi_b\lambda_c^*\theta_d\sigma^\mu_{ab}\bar\sigma_{\mu cd} = -2\chi_a^*\chi_b\lambda_c^*\theta_d\delta_{ac}\delta_{bd} = 2(\chi^\dagger\lambda^\dagger)(\chi\theta), \end{aligned}$$

where the involved identity for the σ-matrices follows from direct evaluation, can be used. Evaluating the previous expression then yields

$$-2q\theta((\chi^\dagger\theta^\dagger)(\chi\lambda) + (\chi^\dagger\lambda^\dagger)(\chi\theta)).$$

This implies the relation

$$A\xi = 2q\theta. \tag{4.49}$$

Together with the condition (4.48) this implies that A and B, and θ and ξ have to be proportional to each other, the constant of proportionality involving the charge. However, a relative factor is still permitted, and is required to be fixed. As the covariant derivative already provided one constraint, it is not surprising that the selectron-photon coupling term provides another one. Taking the supersymmetry transformation of the interaction term between two selectrons and one photon yields, when taking only the transformation of the photon field,

$$\begin{aligned} & -iq(\phi^\dagger(\partial_\mu\phi)\delta_\theta A^\mu - (\partial_\mu\phi)^\dagger\phi\delta_\theta A^\mu) \\ = \quad & iq((\partial_\mu\phi^\dagger)\phi(\theta^\dagger\bar\sigma^\mu\lambda + \lambda^\dagger\bar\sigma^\mu\theta) - \phi^\dagger(\partial_\mu\phi)(\theta^\dagger\bar\sigma^\mu\lambda + \lambda^\dagger\bar\sigma^\mu\theta)). \end{aligned} \tag{4.50}$$

Terms with such a contribution can also be generated by both interaction terms, if in the A case the electron and in the B case the D-boson is transformed. Specifically,

$$Ai(\phi^\dagger(\sigma^\mu\sigma_2\xi^*\partial_\mu\phi)\lambda + \lambda^\dagger(\partial_\mu\phi^\dagger)\xi^T\sigma_2\sigma^\mu\phi) - iB(\phi^\dagger\phi(\theta^\dagger\bar\sigma^\mu\partial_\mu\lambda - (\partial_\mu\lambda^\dagger)\bar\sigma^\mu\theta)).$$

To simplify this expression, the relation (4.25) can be used in both A-terms, yielding

$$Ai(\phi^\dagger(\partial_\mu\phi)\xi^\dagger\bar\sigma^\mu\lambda - (\partial_\mu\phi^\dagger)\phi\lambda^\dagger\bar\sigma^\mu\xi) - iB(\phi^\dagger\phi(\theta^\dagger\bar\sigma^\mu\partial_\mu\lambda - (\partial_\mu\lambda^\dagger)\bar\sigma^\mu\theta)).$$

Integrating further in the B-term by parts yields

$$\begin{aligned} & Ai(\phi^\dagger(\partial_\mu\phi)\xi^\dagger\bar\sigma^\mu\lambda - (\partial_\mu\phi^\dagger)\phi\lambda^\dagger\bar\sigma^\mu\xi) - iB(((\partial_\mu\phi^\dagger)\phi + \phi^\dagger(\partial_\mu\phi))\theta^\dagger\bar\sigma^\mu\lambda \\ & -((\partial_\mu\phi^\dagger)\phi + \phi^\dagger(\partial_\mu\phi))\lambda^\dagger\bar\sigma^\mu\theta). \end{aligned} \tag{4.51}$$

Now, in both contributions, (4.50) and (4.51), terms appear proportional to λ and to $\lambda^\dagger$ of the same structure. So both will be vanishing independently, if the pre-factors combine in the same way. The prefactor of λ is

$$iq((\partial_\mu\phi^\dagger)\phi - \phi^\dagger\partial_\mu\phi)\theta_a + Ai\phi^\dagger(\partial_\mu\phi)\xi_a + iB((\partial_\mu\phi^\dagger)\phi + \phi^\dagger\partial_\mu\phi)\theta_a.$$

This will vanish, if the conditions

$$\begin{aligned} q\theta_a + B\theta_a &= 0 \\ -q\theta_a + A\xi_a + B\theta_a &= 0 \end{aligned}$$

are met. Together with the condition (4.48) and (4.49), this yields the result

$$\begin{aligned} A &= -\sqrt{2}q \\ B &= -q \\ \theta &= -\frac{1}{\sqrt{2}}\xi. \end{aligned}$$

Actually, this result is not unique, and it would be possible to replace A by $-A$ and θ by $-\theta$, without problems. So this can be freely chosen, and the conventional choice is the one adopted here.

With these choices, all variations performed will be either total derivatives or will cancel each other. However, one contribution is not working out, which is the one involving the F-contribution from the variation of the electron in the A-term. It yields

$$-\sqrt{2}q(\phi^\dagger\xi\lambda F + \lambda^\dagger\xi^\dagger F^\dagger\phi) \tag{4.52}$$

Since there is no other term available which contains both the selectron and the photino, it is not possible to cancel this contribution. The only possibility is to modify the transformation rule for the F-boson as

$$\delta_\xi F = -i\xi^\dagger\bar{\sigma}^\mu D_\mu\chi + \sqrt{2}q\lambda^\dagger\xi^\dagger\phi.$$

The only consequence of this modification is that the $FF^\dagger$ term transforms, restricted to the photino contributions, as

$$\delta_\xi^\lambda(FF^\dagger) = \sqrt{2}q(\phi^\dagger\xi\lambda F + \lambda^\dagger\xi^\dagger F^\dagger\phi),$$

canceling exactly the offending contribution.

Thus, finally the Lagrangian for supersymmetric QED reads

$$\begin{aligned} \mathcal{L} &= (D_\mu\phi)^\dagger D^\mu\phi + i\chi^\dagger\bar{\sigma}^\mu D_\mu\chi + F^\dagger F - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\lambda^\dagger\bar{\sigma}^\mu\partial_\mu\lambda + \frac{1}{2}D^2 \\ &\quad -\sqrt{2}q((\phi^\dagger\chi)\lambda + \lambda^\dagger(\chi^\dagger\phi)) - q\phi^\dagger\phi D. \end{aligned}$$

A number of remarks are in order. First, though two essentially independent sectors have been coupled, this lead not to a product structure with two independent supersymmetry transformations, but only to one common transformation. The deeper reason for this is the appearance of the momentum operator in both supersymmetry algebras, inevitably coupling both. The second is that the combination of gauge symmetry and supersymmetry required the introduction of interaction terms between both sectors to yield a supersymmetric theory. This interaction is so strongly constrained that its structure is essentially unique, and besides no new coupling constants appear compared to the two original theories. Again, supersymmetry is tightly constraining. Third, the equation of motion for the D-boson yields $D = q\phi^\dagger\phi$. As a consequence, integrating out the D-boson lets a quadratic interaction term $-q^2/2(\phi^\dagger\phi)^2$ appear. This is a necessary ingredient for the possibility of a Brout-Englert-Higgs effect, driven by a selectron condensation. Thus, even the simplest non-trivial supersymmetric gauge-theory provides much more possibilities than ordinary QED.

It should be noted that the supersymmetry transformation derived here is not unique. It is possible to write down a set of transformations which only involve ordinary derivatives instead of covariant derivatives, and again a slightly modified transformation for the F boson. Both formulations yield identically the same physical results, and especially the Lagrangian is the same. However, for the purpose of generalizing to theories like supergravity, the present formalism, called the de Wit-Freedman formalism, is more useful. The difference between both sets of transformations is essentially only how gauge conditions transform under supersymmetry transformations. In the formalism using only ordinary derivatives, the gauge conditions are not transformed covariantly, and therefore any supersymmetry transformation must be accompanied by a gauge transformation to also maintain the gauge condition intact. Since gauge transformation do not change physics, it is thus rather a matter of convenience from a physical perspective.

4.6.4 Supersymmetric QCD

Again, having the standard model in mind, it is necessary to generalize supersymmetric QED to a non-Abelian version, the simplest of which is supersymmetric QCD. However, in the following the gauge group will not be made explicit, and thus the results are valid for any (semi-)simple Lie group as gauge group.

Since in QCD, and in the standard model in general, the fermions are in the fundamental representation, while the gauge fields are in the adjoint representation, it is not possible to promote somehow the matter fields to the super partners of the gauge bosons, despite these being charged in the supersymmetric version of Yang-Mills theory. It is therefore again necessary to introduce the matter fields as independent fields, together with their superpartners, and couple them minimally to the gauge fields. Therefore, besides the gauge-fields, the gluons, their superpartner, the gluinos, the D-bosons, there will be the quarks, their superpartners, the squarks, and the F-bosons.

Fortunately, the results of supersymmetric QED, together with those for supersymmetric Yang-Mills theory, can be generalized. It is thus possible to write down the transformation rules and the Lagrangian immediately. The only item which requires some more investigation are couplings between the fundamental sector and the adjoint sector. This applies in particular to the appearing quark-gluino-squark couplings. In general, uncontracted indices would imply a gauge-variant term, which may not appear in the Lagrangian. To obtain appropriate contractions, it is, e. g., necessary to write instead of $\phi_i^\dagger \phi_i D^\alpha$ the terms

$$\phi^\dagger D\phi = \phi^\dagger \tau^\alpha D^\alpha \phi = \phi_i^\dagger (\tau^\alpha D^\alpha)_{ij}\phi_j,$$

where the index i takes values in the fundamental representation, while the index α takes values in the adjoint representation. Such combinations are gauge-invariant, when traced. Of course, this has also to be applied to the coupling term appearing in the transformation rule for the F-boson. Taking thus the non-Abelian versions of the transformation rules as

$$\begin{aligned}
\delta_\xi \phi^i &= \xi\chi^i \\
\delta_\xi \chi^i &= \sigma^\mu \sigma_2 \xi^* D_\mu^{ij} \phi^j + \xi F^i \\
\delta_\xi F^i &= -i\xi^\dagger \bar{\sigma}^\mu D_\mu^{ij} \chi^j - \sqrt{2} q \phi^i \tau^\alpha \lambda^{\alpha\dagger} \xi^\dagger \\
\delta_\xi A_\mu^\alpha &= -\frac{1}{\sqrt{2}} \left(\xi^\dagger \bar{\sigma}_\mu \lambda^\alpha + \lambda^{\alpha\dagger} \bar{\sigma}_\mu \xi\right) \\
\delta_\xi \lambda^\alpha &= -\frac{i}{2\sqrt{2}} \left(\sigma^\mu \bar{\sigma}^\nu \xi F_{\mu\nu}^\alpha + 2\xi D^\alpha\right) \\
\delta_\xi D^\alpha &= \frac{i}{\sqrt{2}} \left(\xi^\dagger \bar{\sigma}^\mu D_\mu^{\alpha\beta} \lambda^\beta - D_\mu^{\alpha\beta} \lambda^{\beta\dagger} \bar{\sigma}^\mu \xi\right)
\end{aligned}$$

it is possible to show that the non-Abelian generalization can be constructed just by covariantizing all derivatives, and replacing coupling terms by gauge-invariant ones. This yields

$$\begin{aligned}
\mathcal{L} = &-\frac{1}{4} F_{\mu\nu}^\alpha F_\alpha^{\mu\nu} + i\lambda^{\alpha\dagger} \bar{\sigma}^\mu D_\mu^{\alpha\beta} \lambda^\beta + \frac{1}{2} D^\alpha D^\alpha \\
&+ D_\mu^{ij} \phi_j D_{ik}^\mu \phi_k^\dagger + \chi_i^\dagger i\bar{\sigma}^\mu D_\mu^{ij} \chi_j + F_i^\dagger F_i \\
&+ M\phi_i F_i - \frac{1}{2} M\chi_i\chi_i + \frac{1}{2} y_{ijk}\phi_i\phi_j F_k - \frac{1}{2} y_{ijk}\phi_i\chi_j\chi_k + h.c. \\
&- \sqrt{2} e(\phi^\dagger \chi\lambda + \lambda^\dagger \chi^\dagger \phi) - g\phi^\dagger \phi D, \qquad (4.53)
\end{aligned}$$

where it should be noted that inter-representation couplings equal, e. g.,

$$\begin{aligned}
\phi^\dagger \chi\lambda &= \phi^\dagger \tau^\alpha \chi\lambda^\alpha = \phi_i^\dagger \tau_{ij}^\alpha \chi_j^a \lambda_a^\alpha \\
\phi^\dagger \phi D &= \phi^\dagger \tau^\alpha \phi D^\alpha = \phi_i^\dagger \tau_{ij}^\alpha \phi_j D^\alpha.
\end{aligned}$$

Note that the coupling constants y acquired gauge indices in the fundamental representation. Similar to the Higgs-fermion coupling in the standard model, relations between the different elements of y ensure gauge invariance of these couplings, and these could be determined by explicit evaluation. Note further that the interaction between the matter sector in the fundamental representation and the gauge-sector in the adjoint representation is completely fixed by gauge symmetry and supersymmetry, and there is no room left for any other interaction. In particular, despite the fact that six fields interact with altogether 11 interaction vertices, there are only three independent parameters, the mass parameter M, the Yukawa coupling y, which is constrained by gauge symmetry, and the gauge-coupling e. E. g., the mixed term appearing from the supersymmetry transformation of the F coupling in the Yukawa term vanishes due to the antisymmetry of the y-matrix, which is necessary to ensure gauge invariance.

It is worthwhile to evaluate the terms including a D explicitly after using its equation of motion, which reads

$$\frac{\delta\mathcal{L}}{\delta D_\alpha^\dagger} = D^\alpha - e\phi^\dagger\tau^\alpha\phi = 0,$$

and similarly for F. After integrating out both the D field and F field, this yields the total self-interaction (or potential V) of the ϕ field,

$$V = |M|^2\phi^\dagger\phi + \frac{1}{2}e^2(\phi^\dagger\tau^\alpha\phi)^2 - y_{ijk}y^*_{lmk}\phi_i\phi_j\phi_l^\dagger\phi_m^\dagger.$$

This potential is positive definite, and all of the couplings are uniquely defined. Thus, in contrast to the case of the standard model, the Higgs-potential, as this is the role the squark plays, is completely determined due to supersymmetry. This puts, at least perturbatively, strong constraints on the Higgs mass in the supersymmetric version of the standard model. This will be discussed in more detail later.

One remark should be added. It is in principle possible to add to the Lagrangian (4.53) a further term proportional to $\theta\epsilon_{\mu\nu\rho\sigma}F_a^{\rho\sigma}F_a^{\mu\nu}$, with θ a new coupling constant, a so-called topological term. Due to the antisymmetric tensor, any contribution of such a term drops out in perturbation theory, and it can only contribute beyond perturbation theory. It indeed does so, and plays an important role in topics like chiral symmetry breaking and anomalies. This is already true in the non-supersymmetric version. However, in nature it is experimentally known that for any such term in the standard model the parameter θ is very small, and only an upper bound of about 10^{-10} is known.

However, from the point of view of supersymmetry, this term is conceptually interesting. After rescaling the gauge fields with the coupling constant g, it is possible to combine this term with the term $F^a_{\mu\nu}F_a^{\mu\nu}$ in such a way as that the whole theory now depends entirely on the complex combination $G = g + i\theta$, the holomorphic coupling. Unbroken supersymmetry then ensures that the partition function, and thus any quantity, is holomorphic in G, which permits many highly non-trivial statements, also in the non-perturbative domain. However, the details of this are beyond the scope of this lecture.

4.7 Gauge theories with $\mathcal{N} > 1$

Gauge theories with more than one supercharge are only of very limited phenomenological use in the context of particle physics, as for intact supersymmetry parity cannot be broken; left-handed fermions and right-handed fermions are treated on equal footing. Since the weak interactions do break parity, this is at odds with experiment. However, they are relevant for several reasons. One is that in some extensions of the standard model it is possible to start without parity breaking, and such a theory could be supersymmetrized. However, these are rather involved constructions, which do not appear very promising. Second, theories with larger supersymmetries are more constrained, which helps in obtaining results. They therefore can serve as better accessible, simplified models of ordinary theories. Third, gauge theories with extended supersymmetries play an important role in the context of string theory.

The simplest extension is the $\mathcal{N} = 2$ supersymmetric version of Yang-Mills theory. This requires the combination of a $\mathcal{N} = 1$ gauge supermultiplet with a chiral multiplet. However, because now the chiral multiplet and the gauge multiplet is related by the extended supersymmetry, also the members of the chiral multiplet must be in the adjoint representation, in contrast to the case of super QCD. Hence, there are the gauge field, the gauginos and the corresponding D field, as well as a complex adjoint scalar ϕ, a Majorana fermion ψ, and the corresponding complex F fields.

The Lagrangian of this theory for a simple Lie-algebra is[9]

$$\begin{aligned}\mathcal{L} &= -(D_\mu\phi)^\dagger D^\mu\phi - \frac{1}{2}\bar{\psi}\gamma_\mu D^\mu\psi + F^\dagger F - 2\sqrt{2}gf^{abc}\Re(\lambda_a^T\psi_c\phi_b^\dagger) + igf^{abc}\phi_b^\dagger\phi_c D_a \\ &+\frac{1}{2}D^2 - \frac{1}{4}F_\mu^a F_a^{\mu\nu} - \frac{1}{2}\bar{\lambda}\gamma_\mu D^\mu\lambda + \frac{g^2\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}F_a^{\mu\nu}F_{\rho\sigma}^a.\end{aligned}$$

The theory has no free parameters, besides the gauge coupling g and the θ parameter. The two supersymmetry transformations differ by the way on which supermultiplets they act. One acts on the conventional two sets, but the other acts on the mixed sets (ϕ, λ, F) and $(A, -\psi, D)$, though with the same supersymmetry transformation. Furthermore, this theory has a $SU(2)$ R-symmetry which transforms between the two sets.

[9] A term $-D$ has to be added, if the gauge group is U(1), a case which will not be considered here.

This theory supports a Higgs effect, and indeed in this case the masses of the particles turn out to be unaffected by radiative corrections. I. e., the tree-level masses are already exact, and saturate the BPS bound. This will not be detailed further here, but should give an idea of how strongly the dynamics are constrained by supersymmetry.

The only further non-trivial extension possible without adding gravity is the $\mathcal{N}=4$ case. This is the combination of two $\mathcal{N}=2$ theories, which therefore has an SU(4) R-symmetry, permitting the fields to give different supersymmetry transformations. This theory is somewhat involved. It contains besides the gauge supermultiplet a left-chiral supermultiplet with complex fields ψ and ϕ, and two more left-chiral, denoted by primes $'$, multiplets and their complex conjugates. The lengthy Lagrangian, after integrating out the auxiliary fields for brevity, reads

$$\begin{aligned}\mathcal{L} &= -(D_\mu\phi)^\dagger D^\mu\phi-(D_\mu\phi')^\dagger D^\mu\phi'-(D_\mu\phi'')^\dagger D^\mu\phi''-\frac{1}{2}\bar{\psi}\gamma_\mu D^\mu\psi-\frac{1}{2}\bar{\psi}'\gamma_\mu D^\mu\psi'\\&-\frac{1}{2}\bar{\psi}''\gamma_\mu D^\mu\psi''-\frac{1}{2}\bar{\lambda}\gamma_\mu D^\mu\lambda-2\sqrt{2}gf^{abc}\Re(\phi_a\psi_b'\psi_c'')-2\sqrt{2}gf^{abc}\Re(\lambda_a^T\psi_c\phi_b^\dagger)\\&-2\sqrt{2}gf^{abc}\Re(\phi_b'\psi_c''^T\psi_a)-2\sqrt{2}gf^{abc}\Re(\phi_c''\psi_b'^T\psi_a)+2\sqrt{2}gf^{abc}\Re(\psi_b'^T\lambda_a\phi_c'^\dagger)\\&+2\sqrt{2}gf^{abc}\Re(\psi_b''^T\lambda_a\phi_c''^\dagger)-\frac{1}{4}F_{\mu\nu}^aF_a^{\mu\nu}+\frac{g^2\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}F_a^{\mu\nu}F_a^{\rho\sigma}\\&+g^2f^{ade}f^{bce}(\phi_a\phi_b^\dagger+\phi_b\phi_a^\dagger)(\phi_c'\phi_d'+\phi_c''^\dagger\phi_d''^\dagger)+\frac{g^2}{2}\left|f^{abc}(\phi_b'^\dagger\phi_c'-\phi_b''\phi_c''^\dagger)\right|^2\\&-\frac{g^2}{2}f^{abc}f^{ade}\phi_b^\dagger\phi_c\phi_d^\dagger\phi_e+2g^2\left|f^{abc}\phi_b'\phi_c''\right|^2\end{aligned}$$

Though it does not look like it, the potential in the scalars can be symmetrized, albeit of the expense of becoming even more lengthy. However, it is possible to rewrite the Lagrangian in a much shorter form by exploiting the SU(4) R symmetry. Collecting for each color a the right-handed fermions in an SU(4) vector $\Psi=(\psi_R,\lambda_R,\psi_R',\psi_R'')$ and the scalars into an antisymmetric SU(4) tensor

$$\Phi=\begin{pmatrix}0 & \phi^* & \phi'' & -\phi'\\ -\phi^* & 0 & -\phi'^\dagger & -\phi''^\dagger\\ -\phi'' & -\phi'^\dagger & 0 & \phi\\ \phi' & \phi''^\dagger & -\phi & 0\end{pmatrix} \tag{4.54}$$

separately, this yields

$$\begin{aligned}\mathcal{L} &= -\frac{1}{2}(D_\mu\Phi^{ij})_a(D^\mu\Phi^{ij})_a^\dagger-\frac{1}{2}\Psi_{Lia}^T(\gamma_\mu D^\mu\Psi_{Ri})_a+\frac{1}{2}\Psi_{Rai}^T(\gamma_\mu D^\mu\Psi_{Li})_a\\&-\sqrt{2}gf^{abc}\Re(\Phi_{ija}\Psi_{Lib}^T\Psi_{Ljc})-\frac{g^2}{8}\left|f^{abc}\Phi_b^{ij}\Phi_c^{kl}\right|^2-\frac{1}{4}F_{\mu\nu}^aF_a^{\mu\nu}+\frac{g^2\theta}{64\pi^2}\epsilon_{\mu\nu\rho\sigma}F_a^{\mu\nu}F_a^{\rho\sigma},\end{aligned}$$

which is more compact, but treats the gaugino not explicitly different from the matter particles. However, the SU(4) R-parity is manifest. In both cases, $\mathcal{N}=2$ and $\mathcal{N}=4$, the proof of the supersymmetry is rather lengthy, and will be skipped here.

For $\mathcal{N}=4$, the potential is a sum of squares, and thus the vacuum energy is always zero. Hence, supersymmetry remains unbroken in this theory. Furthermore, there exists evidence that there is an exact mapping of the $\mathcal{N}=4$ theory at a given coupling to another $\mathcal{N}=4$ theory with the same structure but with inverse coupling, thus linking a strongly interacting theory and a weakly interacting theory. This is a so-called duality, which are quite useful, if truly existing.

4.8 The β-function of super-Yang-Mills theory

A remarkable fact, stated here without proof, of super-Yang-Mills theories is that for vanishing θ the only appearing infinity is in the one-loop correction to the β-function. As a consequence, the one-loop form is exact, and given by

$$\begin{aligned}
\beta(g) &= -\frac{g^3}{4\pi^2}\left(\frac{11}{12}C_1-\frac{1}{6}C_2^f-\frac{1}{12}C_2^s\right)\\
C_1\delta_{cd} &= f^{abc}f^{abd}\\
C_2^f\delta_{cd} &= \sum_{\text{fermions}}\mathrm{tr}\tau^c\tau^d\\
C_2^s\delta_{cd} &= \sum_{\text{scalars}}\mathrm{tr}\tau^c\tau^d,
\end{aligned}$$

i. e. determined by the representations of the various involved particles. Most notably, for the $\mathcal{N}=4$ case the requirements on the involved fields balance the C_i such that the β-function vanishes. Hence, this theory is finite, i. e. no renormalization is necessary. Moreover, as the β function vanishes, the coupling does not depend on the energy scale. As a consequence, the theory is scale-free, and hence conformal. But this also means that it lacks any kind of observables, and has thus no dynamics. However, such a behavior makes the possibility of a duality also more plausible.

It should be noted that many supersymmetric theories beyond super-Yang-Mills theory exhibit similar features, i. e. perturbative β-functions which contain only a finite number of terms. In some rare cases, supersymmetry provides strong enough constraints to show that this is true even non-perturbatively. However, such theories, which include $\mathcal{N}=4$ super-Yang-Mills theory, usually have such strict constrains that they exhibit little or no dynamics.

4.9 Supersymmetry breaking

All of the theories investigated so far have manifest supersymmetry. As a consequence, as it was shown generally, the masses of the particles and the sparticles have to be degenerate. This is not what is observed in nature: There is no scalar particle with unit electric charge and the mass of the electron in nature. This is an experimentally very well established fact. Nor has any superpartner for any known particle been found so far. Indeed, if they exist, most of them need to be extremely heavy, significantly above 100 GeV in mass. Otherwise the equal coupling strength to the known forces would have made them observable in experiments long ago. Supersymmetry can therefore not be a symmetry of nature. It is therefore necessary to break supersymmetry in some way.

For the breaking of symmetries two prominent mechanisms are available in quantum field theories. A breaking can either be by explicit breaking or by spontaneous breaking. There is also the breaking by quantum anomalies in the quantization process. However, so far no really attractive, consistent, and experimentally relevant mechanisms to break supersymmetry by anomalies has been found. This option will therefore not be followed here.

Explicit breaking refers to the case when some term is added to the Lagrangian which spoils the symmetry of the theory present without this term. A tree-level mass term for quarks in QCD is such a case, where chiral symmetry is broken by this. If the term is superrenormalizable, like a mass-term, this is not affecting the high-energy properties of an asymptotically free theory, and the breaking is said to be soft. However, low-energy properties may be qualitatively different. If the offending term is small compared to all other scales of the theory, its effect is possibly weak, and the symmetry is said to be broken mildly only. Relations due to the original symmetry may therefore be still approximately valid. However, since interacting quantum field theories are non-linear by nature, there is no guarantee for this.

Spontaneous breaking appears when the Lagrangian is invariant under a symmetry transformation, but the ground-state is not[10]. E. g., the magnetization of a ferromagnet with no external magnetic field is an example of such a case. In field theory, QCD with massless fermions is another example. Also there, the chiral symmetry is spontaneously broken, yielding (approximately) the known masses of the hadrons.

Unfortunately, adding an explicit breaking is not always possible. An example is the so-called breaking of electroweak gauge symmetry in the standard model. In this case, any

[10] There are some subtleties involved here what is precisely meant by ground-state in a quantum field theory. This subtleties are often irrelevant, especially in the following discussion. Hence, they will be glossed over.

explicit breaking term will spoil renormalizability[11]. Also, explicit breaking terms have in general less attractive properties, varying from theory to theory. Therefore, spontaneous breaking is more desirable. However, any spontaneous mechanism for supersymmetry breaking known so far is not consistent with the requirements of experiments. Therefore, it is required to introduce explicit breaking of supersymmetry. Unfortunately, no simple possibility is known to obtain acceptable results, and therefore many of the attractive properties of supersymmetry are lost. In particular, almost a hundred additional coupling constants and parameters will be necessary even for the simplest supersymmetric extension of the standard model. This will be detailed in section 4.10. In this section, only the underlying mechanisms will be discussed.

4.9.1 Dynamical breaking

Spontaneous breaking of supersymmetry requires that some quantity ω', which is not invariant under supersymmetry transformations, $\delta\omega' \neq 0$, must develop a vacuum expectation value[12],

$$\langle 0|\omega'|0\rangle \neq 0.$$

Since this implies that ω' belongs to a supermultiplet of some kind, there exists a field ω such that

$$\omega' = i[Q,\omega].$$

This implies

$$\langle 0|\omega'|0\rangle = \langle 0|i[Q,\omega]|0\rangle = \langle 0|iQ\omega - i\omega Q|0\rangle \neq 0.$$

Since Q is hermitian, this implies that $Q|0\rangle \neq 0$, as otherwise this expectation value would vanish. Conversely, this implies that if supersymmetry is unbroken, the vacuum state is uncharged with respect to supersymmetry, $Q|0\rangle = 0$. It can be shown that this exhausts all possibilities.

The implications of this can be obtained when noting that there exist a connection between supersymmetry generators and the Hamiltonian, and thus the energy. The commutation relations for the Q_a yield

$$\begin{aligned}
\{Q_1, Q_1^\dagger\} &= (\sigma^\mu)_{11}P_\mu = & P_0 + P_3\\
\{Q_2, Q_2^\dagger\} &= (\sigma^\mu)_{22}P_\mu = & P_0 - P_3.
\end{aligned}$$

[11]Actually, not superficially, but still.

[12]There are once more field-theoretically subtleties involved with this statement, which will be glossed over here.

Thus the Hamiltonian $H = P_0$ is given by

$$H = \frac{1}{2}\left(Q_1 Q_1^\dagger + Q_1^\dagger Q_1 + Q_2 Q_2^\dagger + Q_2^\dagger Q_2\right).$$

Taking the expectation value of H yields

$$\langle 0|H|0\rangle = \frac{1}{2}\sum_a \left((Q_a|0\rangle)^2 + \left(Q_a^\dagger|0\rangle\right)^2\right).$$

Hence, the ground-state energy for the case of unbroken supersymmetry is zero, as none of the right-hand terms can be different from zero, as announced earlier. Since the right-hand side is a sum of squares it also follows that in case of spontaneous supersymmetry breaking the vacuum energy is not zero, but is larger than zero.

To detect breaking, it is necessary to specify the object ω', which breaks the supersymmetry. In principle, this can also be a composite operator. Such mechanisms are known, e. g., in QCD. Here, it will be restricted to the case where ω' is an elementary field. The situation in the non-gauge case and the gauge case are a little different, and will be treated in turn in the following.

4.9.1.1 The O'Raifeartaigh model

The O'Raifeartaigh model is a non-gauge model, essentially an extension of the Wess-Zumino model, which can exhibit spontaneous supersymmetry breaking. To study the possible elementary fields for developing a vacuum expectation value it is helpful to reconsider the transformations under supersymmetry in the Wess-Zumino model

$$\begin{aligned}
\delta_\xi \phi &= i[\xi Q, \phi] = && \xi\chi \\
\delta_\xi \chi &= i[\xi Q, \chi] = && -i\sigma^\mu i\sigma_2 \xi^* \partial_\mu \phi + \xi F \\
\delta_\xi F &= i[\xi Q, F] = && -i\xi^\dagger \bar{\sigma}^\mu \partial_\mu \chi.
\end{aligned}$$

Phenomenologically, so far Lorentz-symmetry-respecting models are most interesting, and thus any condensates may not break this symmetry. This rules out already χ, $\partial_\mu\chi$ and $\partial_\mu\phi$, as all of these have a definite direction. Hence, the only field remaining is the scalar F field. The value of F is fixed by its equation of motion as

$$F = -\frac{\delta W^\dagger}{\delta\phi^\dagger} = -\left(M\phi + \frac{1}{2}y\phi^2\right)^\dagger. \tag{4.55}$$

The contribution of F to the Hamiltonian, and thus the interaction energy, is given by $FF^\dagger$. This is a positive definite contribution. Its lowest value is achieved, as can be seen from (4.55), exactly when all ϕ-fields vanish. In this case, the contribution of F to the ground-state energy is zero, and is thus not able to break supersymmetry. It is necessary to force F to a value different from zero. For this purpose, it is actually insufficient to just add a constant term to (4.55). Though this formally shifts F to a value different from zero for $\phi = 0$, it is always possible to shift it back if ϕ is replaced by a non-zero, constant value. A non-zero, constant value for all the fields will not produce kinetic energy. So, the only other contribution could come from the Yukawa coupling to the fermions. However, it is still permitted to set these to zero. Thus, the ground state energy becomes once more zero, and supersymmetry is intact, despite the non-vanishing value of F and ϕ. This is in fact not a contradiction: The shift in the ϕ-field can then be taken to be a renormalization of the field, and the resulting theory is in fact supersymmetric.

It thus requires a more complicated approach. However, including a linear term is already a good possibility, but it turns out to be impossible with just one flavor. O'Raifeartaigh showed that it is possible, if there are at least three flavors. The parameters of the superpotential are then chosen as

$$\begin{aligned} A_i &= -gM^2\delta_{i2} \\ M_{ij} &= \frac{m}{2}(\delta_{i1}\delta_{j3} + \delta_{i3}\delta_{j1}) \\ y_{ijk} &= \frac{g}{3}(\delta_{i2}\delta_{j3}\delta_{k3} + \delta_{i3}\delta_{j2}\delta_{k3} + \delta_{i3}\delta_{j3}\delta_{k2}), \end{aligned}$$

with g, m, and M real and positive, yielding a superpotential

$$W = m\phi_1\phi_3 + g\phi_2(\phi_3^2 - M^2).$$

As the Wess-Zumino-Lagrangian is supersymmetric for any form of the superpotential, this choice is not breaking supersymmetry explicitly. However, even at tree-level the minimum energy is non-zero, thus supersymmetry is spontaneously broken. This can be seen as follows. The equations of motion for the three $F_i^\dagger$ fields take the form

$$\begin{aligned} F_1^\dagger &= -m\phi_3 \\ F_2^\dagger &= -g(\phi_3^2 - M^2) \\ F_3^\dagger &= -m\phi_1 - 2g\phi_2\phi_3, \end{aligned}$$

and correspondingly for the F_i. From these equations of motion it follows that, at least at tree-level, either the field F_1 or the field F_2 will have a vacuum expectation value. This cannot be shifted away by a wave-function renormalization of ϕ_3, since anything shifting F_1 to zero will shift F_2 to a non-zero value. Putting it differently, these equations of motions force F_1 and F_2 to have different values. Since the contributions of the F_i to the vacuum energy is a sum of squares of type $F_iF_i^\dagger$, at least the contribution from one flavor is always non-zero.

The contribution of the Yukawa term may at first seem to be a tempting possibility to change the situation. However, this would require that the field χ acquires a vacuum expectation value. This would require that the vacuum has to have a non-zero spinor component, as then $\langle 0|\chi_i|0\rangle$ would be non-zero. This would clearly break Lorentz invariance, and is thus not admissible.

It therefore remains to minimize the potential energy with respect to the fields ϕ_i and F_i. Writing the potential explicitly yields

$$V = \sum_i F_i F_i^\dagger = m^2|\phi_3|^2 + g^2|\phi_3^2 - M^2|^2 + |m\phi_1 + 2g\phi_2\phi_3|^2. \tag{4.56}$$

This has to be minimal for the vacuum state. Since ϕ_1 can always be chosen such that the third term vanishes, it remains to check the first two terms. Rewriting the expression in terms of the real part A and the imaginary part B of ϕ_3 yields the expression

$$V = g^2M^4 + (m^2 - 2g^2M^2)A^2 + (m^2 + 2g^2M^2)B^2 + g^4(A^2 + B^2)^2.$$

If the first expression is positive, i. e. $m^2 \geq 2g^2M^2$, the lowest values of V is achieved for $A = B = 0$. Otherwise, a solution with A and B non-zero exists. This provides little qualitative new results for the present purpose, and so only the first case will be treated. In this case, $\phi_3 = 0$, and consequently thus $\phi_1 = 0$. The value of ϕ_2 is not constrained at all, and ϕ_2 could take, in principle, any value. Therefore, it is called a flat direction of the potential. Under certain circumstances, this may pose a problem in the form of an instability, but this is of no interest here.

With this result, F_2 acquires a vacuum expectation value of size gM^2, and the vacuum energy is the positive value g^2M^4. Since M has the dimension of mass, this vacuum energy at the same time gives the scale of supersymmetry breaking. If g would be of order one, $M = 1$ TeV would, e. g., signal a breaking of supersymmetry at the scale 1 TeV, which would be accessible at the LHC.

It is noteworthy that

$$0 \neq \langle 0|[Q,\chi]|0\rangle = \sum_n (\langle 0|Q|n\rangle\langle n|\chi|0\rangle - \langle 0|\chi|n\rangle\langle n|Q|0\rangle).$$

Since Q and χ are both fermionic, this implies that there exists a state $|n\rangle$, which must also be fermionic, which couples to the vacuum by the generator Q such that $\langle 0|Q|n\rangle$ is non-zero. Since it couples in such a way to the vacuum, it can be shown that it is massless. This is the SUSY version of the Goldstone theorem, which differs by the appearance of a massless fermionic mode from the conventional one. That is, of course, due to the fact that the supercharge is fermionic.

Since the mass-matrix of the superpotential has only entries in the (13)-submatrix, this implies that the flavor 2 fermion is massless, and can take this role. Due to the relation to the Goldstone theorem, it is called goldstino, though there is no Goldstone boson in the theory. Consequently, also the boson ϕ_2 is not having a mass. This correlates to the fact that ϕ_2 is the flat direction in the superpotential: Moving in this direction is not costing any energy, similar to a Goldstone excitation in conventional theories.

That supersymmetry is indeed broken can also be seen explicitly by the masses of the remaining two flavors. By diagonalizing the mass-term for the other two fermion flavors, it is directly obtained that there exists two linear combinations, both with masses m. The mass-terms for the scalars are obtained by taking the quadratic terms of the potential (4.56), yielding

$$m^2\phi_3\phi_3^\dagger + gM^2(\phi_3^2 + \phi_3^{\dagger 2}) + m^2\phi_1\phi_1^\dagger.$$

This confirms the masslessness of the ϕ_2 boson. Furthermore, real and imaginary part of the flavor 1 have mass m^2. The flavor 3 has real and imaginary parts with different masses, $m^2 \pm gM^2$. This already implies that the supermultiplets are no longer mass-degenerate, and supersymmetry is indeed broken.

However, when summing up everything, it turns out that the relation

$$\sum_{\text{scalars}} m_s^2 = 0+0+m^2+m^2+m^2+gM^2+m^2-gM^2 = 2\sum_{\text{fermions}} m_f^2 = 2(0+m^2+m^2) \quad (4.57)$$

holds. Note that the complex scalars correspond to two scalar particles, while each Weyl fermion represents one particle with two different spin orientation. It turns out that this relations holds generally for this (F-type) spontaneous breaking of supersymmetry. This implies that the masses of the particles and their super-partners have to be quite similar. As a consequence, such a mechanism is not suitable for the standard model, since otherwise already super-partners would have been observed[13]. More fundamentally, when constructing the minimal supersymmetric standard model, it will turn out that there is no gauge-invariant scalar field which could play the role of the second flavor in this model. Since the superpotential has to be gauge-invariant term-by-term, it is thus not possible to have a linear term in that case in the superpotential, and this type of spontaneous supersymmetry breaking is not permitted.

[13]Adding an additional heavy fourth generation may seem at first sight a tempting possibility to evade this constraint. However, it can be shown that for the charge structure of the standard model further constraints exist which forces always at least one super-partner to be light enough to be already detected.

4.9.1.2 Spontaneous breaking of supersymmetry in gauge theories

Interactions are necessary for a spontaneous breaking of supersymmetry. The simplest non-trivial case of a gauge theory is that of supersymmetric QED. In the matter sector, the breaking of supersymmetry can only proceed once more due to a Wess-Zumino-like interaction, which is not present in supersymmetric QED. Therefore, the field F cannot develop a vacuum expectation value, and neither the electron nor the selectron are possible candidates, due to the arguments in the preceding section. Inspecting the remaining transformation rules

$$\begin{aligned} i[\xi Q, A_\mu] &= -\frac{1}{\sqrt{2}}(\xi^\dagger\bar{\sigma}_\mu\lambda + \lambda^\dagger\bar{\sigma}_\mu\xi) \\ i[\xi Q, \lambda] &= -\frac{i}{2\sqrt{2}}\sigma^\mu\bar{\sigma}^\nu F_{\mu\nu}\xi + \frac{1}{\sqrt{2}}\xi D \\ i[\xi, D] &= \frac{i}{\sqrt{2}}(\xi^\dagger\bar{\sigma}^\mu\partial_\mu\lambda - (\partial_\mu\lambda)^\dagger\bar{\sigma}^\mu\xi), \end{aligned}$$

then, by the same reasoning as before, suggests that the only field which can provide a scalar condensate is the D-field. However, the contribution of the D-boson to the potential is just $eD\phi^\dagger\phi$, and the equation of motion is $D = e\phi^\dagger\phi$. Thus, there is no sum of terms, as in the Wess-Zumino case, which can be exploited to construct a potential which offers the possibility for supersymmetry breaking.

However, there is another possibility, the addition of the so-called Fayet-Iliopoulos term. In this case, the D-sector of the Lagrangian is replaced by

$$\mathcal{L}_D = M^2 D + \frac{1}{2}D^2 - gD\phi^\dagger\phi.$$

The D-field is gauge-invariant, thus this Lagrangian is also still gauge-invariant. Furthermore, the SUSY transformation of D is a total derivative, and thus any linear term in D is also not spoiling supersymmetry, and this Lagrangian is therefore a perfectly valid extension of supersymmetric QED. The new equation of motion for D is then

$$D = -M^2 + g\phi\phi,$$

yielding the contribution

$$\frac{1}{2}(-M^2 + g\phi^\dagger\phi)^2 \tag{4.58}$$

to the Hamiltonian's potential energy. The sign of g can be selected freely. If g is greater than zero, then a minimum of this potential energy[14] is obtained at $|\phi| = M/\sqrt{g}$. In this

[14]Note that the arguments on which fields can acquire such a vacuum expectation values are no longer valid anymore, since instead of the supersymmetry transformations for this case the gauge transformation have to be investigated.

case, the potential energy contribution is zero, and supersymmetry is unbroken. However, the scalar condensate is equivalent to a Higgs-effect, giving the bosons a mass. But if $g < 0$, then the minimum is obtained for $\langle\phi\rangle = 0$, with a potential energy contribution which is non-zero, $M^4/2$. Thus, supersymmetry is broken in this case, and the D field acquires the expectation value $-M^2$.

As a consequence, the selectron field acquires a mass by its interaction with the D-field, but the other fields remain massless, in particular the photon, the photino, and the electron. Thus, the degeneration in the mass spectrum of the matter fields is indeed broken, signaling consistently the breakdown of supersymmetry. Note that the sum-rule (4.57) cannot be applied here, as this supersymmetry breaking proceeds by a different mechanism.

Unfortunately, this mechanism cannot be extended to non-Abelian gauge groups, as in this case the D field is no longer gauge-invariant, and neither is its supersymmetry transformation anymore a total derivative. Thus, in the non-Abelian case, supersymmetry breaking by the Fayet-Iliopoulos mechanism is not possible.

Also utilizing then just the QED sector of the standard model is not an option: In the standard model the single $\phi^\dagger\phi$ in (4.58) is replaced with a sum over all fields carrying electric charge, and with their respective positive and negative charges. Thus, the corresponding minimum would be obtained by some of the squark fields and slepton fields developing a vacuum expectation value, and some not. As a consequence, in the standard model case, the breaking of supersymmetry with a Fayet-Iliopoulos term would imply the breaking of the electromagnetic symmetry and color gauge symmetry, which is not compatible with experiment. Therefore, also this second mechanism of spontaneous breaking of supersymmetry is not viable for the standard model, and it is necessary to turn to explicit breaking.

4.9.2 Explicit breaking

Thus, at the present time, no satisfactory mechanism exists for the breaking of supersymmetry. Therefore, it is necessary to parametrize this lack of knowledge in the form of explicitly supersymmetry breaking terms[15]. However, if some of the specific properties of supersymmetric theories, in particular the better renormalization properties and the protection of the scalar masses, should be retained, it is not possible to add arbitrary terms to the Lagrangian.

To ensure the survival, or at least only minor modification, of these properties, it is necessary to restrict the explicit supersymmetry breaking contributions to soft contribu-

[15] Of course, it cannot be excluded that supersymmetry in nature is in fact explicitly broken.

tions. Soft contributions are such contributions which become less and less relevant with increasing energy in an asymptotically free theory. This ensures that supersymmetry becomes effectively restored at large energies. To ensure such a property, it is necessary that the terms contain coupling constants which have a positive energy dimensions. In case of the theories discussed so far, these can appear in the form of two types of terms.

One type of such terms are masses which do not emerge from a superpotential. In particular, such terms are allowed even without integrating out the F-bosons. Such masses are possible for both, the bosonic fields and the fermionic fields, but not for the auxiliary fields, as their mass dimensions are not permitting renormalizable mass terms. However, these cannot appear for gauge bosons, as in standard gauge theories. Thus, e. g., in supersymmetric QCD, such terms would be of type

$$-\frac{1}{2}\left(m_\lambda\lambda\lambda + m_\chi\chi\chi + m_\lambda^\dagger\lambda^\dagger\lambda^\dagger + m_\chi^\dagger\chi^\dagger\chi^\dagger + 2m_\phi^2\phi^\dagger\phi\right)$$

Note that all mass terms are gauge-invariant. Also, since this is an effective parametrization, all masses, m_λ, m_χ, and m_ϕ, are independent parameters. Furthermore, these masses may be complex. When the F-field would be integrated out, the additional mass terms for the quarks and the squarks would mix with those introduced above. Not only is such a Lagrangian not explicitly supersymmetric anymore, but, since the super-partners are no longer mass-degenerate, also the spectrum is manifestly no longer supersymmetric.

Furthermore, an additional possibility are three-boson couplings. These couplings would be of type

$$a_{ijk}\phi_i\phi_j\phi_k + b_{ijk}\phi_i^\dagger\phi_j\phi_k + c_{ijk}\phi_i^\dagger\phi_j^\dagger\phi_k + d_{ijk}\phi_i^\dagger\phi_j^\dagger\phi_k^\dagger,$$

where, of course, not all couplings are independent, but are constrained by gauge-invariance. That such couplings break supersymmetry is evident, as they not include the particles and their super-partners equally.

All in all, to the three independent parameters of the supersymmetric QCD, the gauge coupling, and the two Wess-Zumino parameters g and M, four more have emerged, three masses, and at least one three-boson coupling. All of these additions are in fact supersymmetry breaking. Inside such a model, there is no possibility to predict the values of the additional coupling constants. However, in cases where the (broken) supersymmetric theory is just the low-energy limit of a unified theory at some higher scale, this underlying theory can provide such predictions, at least partially. E. g., in the cases above, the expectation values of fields have been obtained in terms of the masses and the coupling constants of the fields, or vice versa.

This already closes the list of possible soft supersymmetry breaking terms, although for specific models much more possibilities may exist.

4.9.3 Breaking by mediation

As has been seen, non-explicit supersymmetry-breaking faces the problem that the relation between particles are inconsistent with experiment. Explicit breaking, however, introduces numerous additional parameters. A compromise is supersymmetry-breaking by mediation. In this case, dynamical breaking of supersymmetry is made possible by additional particles and interactions. Thereby the problem of experimentally unwanted breaking of electromagnetism or the strong interaction as well as relations like (4.57) can be circumvented. Furthermore, this usually requires much less parameters as an explicit breaking, and provides a dynamical origin.

Of course, just adding more generations or variously charged particles will not solve the problem. Though this may circumvent the simple sumrule (4.57), this will lead to other problems, like too strong Yukawa-interactions for the additional generations to be easily compatible with the observed Higgs particle, or conflicts with further sum-rules specific to the standard model. Also, breaking of the electromagnetic interaction and strong interaction can usually not be avoided in this way.

The alternative is then the addition of a complete new sector of particles, including their own interactions. This sector is arranged such that supersymmetry breaking is possible. To communicate this to the standard model requires some kind of interaction. This is performed by so-called messengers. These are further, usually again additional, particles, which are made quite heavy, significantly above the electroweak scale. In this way, even small breakings will be able to introduce substantial effects. This is usually done by resolving the explicit breaking parameters into effective vertices of interactions with this messenger particles. This can be achieved in a similar way as in the Higgs and weak interaction effects at low-energy, where both effects only appear as point-interaction, in the form of the fermion masses and the effective four-fermion interaction of the effective Fermi theory.

This procedure seems still to be an ad hoc resolution of the problem. To embed this in a less ad-hoc framework, two particular possibilities have been pursued.

One is gauge mediation. In this case it is assumed that the three standard-model gauge interactions are just part of one unified gauge-interaction at high scales, see also chapter 7, and so is the supersymmetry-breaking sector. The breaking of this master gauge theory has then separated the standard-model, and fractured it into three interactions, and the breaking sector at some high-scale. The surplus gauge-bosons at this high scale then usually have masses of the breaking scale, but still interact with both sectors. In this way they can act as messengers. While this scenario is in general very attractive, as it solves many problems of the standard model, it also has its own problems. Especially, the larger mass hierarchies emerging in this case are usually accidental, and not well understood.

An alternative is gravity-mediated supersymmetry-breaking, where the gravitino acts as messenger. As this setup requires a full super-gravity theory, it will not be detailed here any further. However, it appears phenomenologically somewhat more appealing, as it is compatible with experimental results with rather little effort, mainly due to the weak interaction of gravity. It usually leads to keV-scale gravitinos, but since they couple gravitationally, and thus very weak, this is not at odds with phenomenology.

4.10 A primer on the minimal supersymmetric standard model

4.10.1 The supersymmetric minimal supersymmetric standard model

From the previous examples of simple theories it is clear that a supersymmetry transformation cannot change any quantum number of a field other than the spin. In the standard model, however, none of the bosons has the same quantum numbers as any of the fermions. Hence, to obtain a supersymmetric version of the standard model, it will be necessary to construct for each particle in the standard model a new super-partner. Of course, additional fields need also be included, like appropriate F-bosons and D-bosons. Here, the supersymmetric version of the standard model with the least number of additional fields will be introduced, the so-called minimal supersymmetric standard model (MSSM). Furthermore, since no viable low-energy supersymmetry breaking mechanism for such a theory is (yet) known, supersymmetry will be broken explicitly. This will require roughly 100 new parameters. This may seem a weakness at first. However, the advantage is that any kind of supersymmetry at high energies can be accommodated by such a theory. Hence, if there are no additional particles or sectors, for which no experimental evidence exists so far, any supersymmetric high-energy theory will look at accessible energies like this minimal-supersymmetric standard model.

The MSSM requires the following new particles

- The photon is uncharged. Therefore, its super-partner has also to be uncharged, and to be a Weyl fermion, to provide two degrees of freedom. It is called the photino

- The eight gluons carry adjoint color charges. This requires eight Weyl fermions carrying adjoint color charges as well, and are called gluinos

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

- Assuming that all neutrinos are massive, no distinction between left-handed and right-handed leptons is necessary, except for the parity-violating weak interactions. However, no new interactions should be introduced due to the superpartners, and hence the superpartners cannot be spin-1 bosons, which would be needed to be gauged. Therefore these superpartners are taken to be scalars, called the sneutrinos, the selectron, the smuon, and the stau, together the sleptons

- The same applies to quarks, requiring the fundamentally charged squarks

- The Higgs requires a fermionic superpartner, the higgsino. However, requiring supersymmetry forbids that the Higgs has the same type of coupling to all the standard model fields. Therefore, a second set of Higgs particles is necessary, with their corresponding super-partners. These are the only new particles required which are not introduced as superpartners of existing particles

- Of course, a plethora of auxiliary D and F bosons will be necessary

It is necessary to discuss the formulation of these fields, and the resulting Lagrangian, in more in detail.

The electroweak interactions act differently on left-handed fermions and right-handed fermions. It thus fits naturally to use independent left-handed chiral multiplets and right-handed chiral multiplets to represent the fermions, together with their bosonic superpartners, the sfermions. However, both Weyl-spinors can be combined into a single Dirac-spinor, as the total number of degrees of freedom match. The electroweak interaction can then couple by means of the usual $1 \pm \gamma_5$ coupling asymmetrically to both components. In this context, it is often useful to use the fact that a charge-conjugated right-handed spinor is equivalent to a left-handed one, as discussed previously, permitting to use exclusively left-handed chiral multiplets, where appropriate charge-conjugations are included.

Furthermore, for each flavor it is necessary to introduced two chiral multiplets, giving a total of 12 supermultiplets for the quarks and the leptons each. The mixing of quark and lepton flavors proceeds in the same way as in the standard model. This requires thus already the same number of parameters for CKM-matrix and the PMNS-matrix as in the standard model.

The gauge-bosons are much simpler to introduce. Again, a gauge-multiplet is needed for the eight gluons. As electroweak symmetry breaking is not yet implemented, and actually cannot without breaking supersymmetry as well as noted below, there is actually a SU(2) gauge multiplet and an U(1) gauge multiplet, which do not yet represent the weak gauge bosons and the photon. Instead, before mixing, these are called $W^\pm$ and W^0, and B. Only after mixing, the W^0 and the B combine to the Z and the conventional photon.

Concerning interactions, there are, of course, the three independent gauge couplings of the strong force, the weak force, and electromagnetic forces, to be denoted by g, g' and e. After electroweak symmetry breaking, g' and e will mix, just as in the standard model. Note that the coupling of the gauge multiplets and the chiral multiplets will induce additional interactions, as in case of supersymmetric QED and suspersymmetric QCD. However, no additional parameters are introduced by this.

It remains to choose the superpotential for the 24 chiral multiplets, and to introduce the Higgs fields. The parity violating weak interactions imply that a mass-term, the component of the superpotential proportional to $\chi\chi$, is not gauge-invariant, since a product of two Weyl-spinors of the type $\chi\chi$ is not, if only the left-hand-type component is transformed under such a gauge transformation. This is just as in the standard model. Therefore, it is not possible, also in the MSSM, to introduce masses for the fermions by a tree-level term. Again, the only possibility will be one generated dynamically by the coupling to the Higgs field[16]. There, however, a problem occurs. In the standard model this is mediated by a Yukawa-type coupling of the Higgs field to the fermions. However, it is necessary, for the sake of gauge-invariance, to couple the two weak charge states of the fermions differently, one to the Higgs field, and one to its complex conjugate. This is not possible in a supersymmetric theory, as the holomorphic superpotential can only depend on the field, and not its complex conjugate. It is therefore necessary to couple both weak charge states to different Higgs fields. This makes it possible to provide a gauge-invariant Yukawa-coupling for both states, but requires that there are two complex Higgs doublets instead of one complex Higgs doublets in the MSSM. Furthermore, also these fields need to be part of chiral supermultiplets, and therefore their partners, the higgssinos, are introduced as Weyl fermions. However, for the Higgs bosons, which have no chirality, the limitations on a mass-term do not apply, and one can therefore be included in the superpotential.

Before writing down the superpotential explicitly, it is necessary to fix the notation. Left-handed Quark fields will be denoted by $u, d, ...$, and squark fields by $\tilde{Q} = \tilde{u}, \tilde{d}, ...$ Note that always a u-type quark and a d-type quark form a doublet with respect to the weak

[16]Actually, a gaugino condensation would also be possible, if a way would be known how to trigger it and reconcile the result with the known phenomenology of the standard model.

interactions, and equivalently the squarks. Right-handed quarks are denoted as $\bar{u}, \bar{d}, ...$ and the corresponding squarks as $\tilde{\bar{q}} = \tilde{\bar{u}}, \tilde{\bar{d}}, ...$. These form singlets with respect to the weak interactions. This notation corresponds to left-handed fields, which are obtained from the original right-handed fields, to be denoted by, e. g. u_R, by charge conjugation. The index L for left-handed will always be suppressed. Likewise, the left-handed leptonic doublets, including an electron-type fermion and a neutrino-type fermion, are denoted as $\nu, e, ...$, and the corresponding sleptons as $L = \tilde{\nu}, \tilde{e}, ...$. Consequently, the right-handed leptons are denoted as $\bar{\nu}, \bar{e}, ...$ and the singlet sleptons as $\tilde{\bar{l}} = \tilde{\bar{\nu}}, \tilde{\bar{e}}, ...$. The two Higgs doublets are denoted by H_u and H_d, denoting to which type of particle they couple. The corresponding higgsinos are denoted by $\tilde{H}_u$ and $\tilde{H}_d$. In contradistinction to the quarks, where the doublet consists out of a component of u and d-quarks, the doublets for the Higgs are formed in the form H_u^+ and H_u^0, and H_d^0 and H_d^-. The reason for this is that after giving a vacuum expectation value to the Higgs-field and expanding all terms of the potentials, effective mass-terms for the u-type quarks will then finally couple only to H_u-type Higgs fields, and so on. The remaining fields are the gluons g with super-partner gluinos $\tilde{g}$, the gauge W-bosons W with super-partner winos $\tilde{W}$, and the gauge field B with its super-partner bino $\tilde{B}$. After mixing, the usual W-bosons $W^\pm$ with superpartner winos $\tilde{W}^\pm$, the Z-boson with superpartner zino $\tilde{Z}$ and the photon γ with the photino $\tilde{\gamma}$ will be obtained.

With this notation, it is possible to write down the superpotential for the MSSM as

$$W = \tilde{y}_u^{ij}\tilde{\bar{u}}_i\tilde{Q}_jH_u - \tilde{y}_d^{ij}\tilde{\bar{d}}_i\tilde{Q}_jH_d + \tilde{y}_{\nu}^{i\tilde{\bar{j}}} e_i\tilde{L}_jH_u - \tilde{y}_e^{ij}\tilde{\bar{e}}_i\tilde{L}_jH_d + \mu H_uH_d, \tag{4.59}$$

where gauge-indices have been left implicit on both, the couplings and the fields, and the auxiliary fields have already been integrated out. The appearing Yukawa-couplings y are the same as in the standard model. In particular, vanishing neutrino masses would be implemented by $y_\nu^{ij} = 0$. If only the masses of the heaviest particles, the top, bottom, and the τ should be retained, this requires that all y-components would be zero, except for $y_u^{33} = y_b$, $y_d^{33} = y_t$ and $y_e^{33} = y_\tau$, leaving gauge indices implicit. The choice of this potential is not unique, but the one for which the MSSM is most similar to the standard model.

In principle, it would also be possible to add contributions like

$$\lambda_e^{ijk}\tilde{L}_i\tilde{L}_j\tilde{\bar{e}}_k + \lambda_L^{ijk}\tilde{L}_i\tilde{Q}_j\tilde{\bar{d}}_k + \mu_L^i\tilde{L}_iH_u \tag{4.60}$$

and

$$\lambda_B^{ijk}\tilde{\bar{u}}_i\tilde{\bar{d}}_j\tilde{\bar{d}}_k. \tag{4.61}$$

In all cases, these are couplings of squarks and/or sleptons. Since these carry the same charges as their counterparts, they will also carry the same lepton and baryon numbers. As a consequence, the squarks $\tilde{Q}$ from the multiplets including the particle-like left-handed quarks carry baryon number $B = 1/3$, while those from the anti-particle-like right-handed chiral multiplets carry baryon number $-1/3$. Similarly, the fields $\tilde{L}$ carry lepton number $L = 1$ and $\bar{e}$ -1. Thus the interaction vertices in (4.60) violate lepton number conservation and the ones of (4.61) baryon number conservation. On principal grounds, there is nothing wrong with this, as both quantities are violated by non-perturbative effects also in the standard model. However, these violations are very tiny, even below nowadays experimental detection limit. Interaction vertices like (4.60) and (4.61), on the other hand, would provide very strong, and experimentally excluded, violations of both numbers, as long as the coupling constants λ_i and μ_L would not be tuned to extremely small values. Such a fine-tuning is undesirable.

However, such direct terms could be excluded, if all particles in the MSSM would carry an additional multiplicatively conserved quantity, called R-parity, which is defined as

$$R = (-1)^{3B+L+2s},$$

with s the spin. This is just the discrete Z_2 subgroup of the U(1) R-symmetry of an $\mathcal{N} = 1$ supersymmetry, which can be retained even after supersymmetry breaking in the MSSM.

Such a quantum number would be violated by interaction terms like (4.60) and (4.61), and these are therefore forbidden[17]. The contribution $2s$ in the definition of R implies that particles and their super-partners always carry opposite R-parity. This has some profound consequences. One is that the lightest superparticle (LSP) cannot decay into ordinary particles. It is thus stable. This is actually an unexpected bonus: If this particle would be electromagnetically uncharged, it is a natural dark matter candidate. However, it has also to be uncharged with respect to strong interactions, as otherwise it would be bound in nuclei. This is not observed, at least as long as its mass is not extremely high, which would be undesirable, as then all superparticles would be very massive, preventing a solution of the naturalness problem by supersymmetry. Therefore, it interacts only very weakly, and is thus hard to detect. Not surprisingly, it has not be detected so far, if it exists.

Summarizing, the assumption of R-parity restricts the superpotential to the form (4.59), and thus fixes the MSSM completely.

Returning then to this remaining possible superpotential (4.59), the parameter μ characterizes the masses of the Higgs and higgsinos. In comparison to the standard model with two independent parameters in the Higgs sector, this is the only new parameter entering the theory when including the Higgs sector and keeping supersymmetry unbroken. The

[17] A non-perturbative violation of R-parity notwithstanding, as for baryon and lepton number in the standard model.

self-interaction of the Higgs field will be entirely determined, due to supersymmetry, just by the already included parameters, thus having potentially less parameters than in case of the standard model.

Unfortunately, there is a catch. Writing all gauge components explicitly, the mass term for the Higgs takes the form

$$\mu H_u H_d = \mu^{ab} H_u^a H_d^b = \mu \epsilon_{ab} H_u^a H_d^b,$$

where the index structure of the parameter μ is dictated by gauge invariance. The corresponding interaction terms in the Lagrangian are then of the form

$$\frac{\delta W}{\delta H_i} F^i + h.c. = \mu(H_u^1 F_d^2 + H_d^2 F_u^1 - H_u^2 F_d^1 - H_d^1 F_u^2) + h.c..$$

Integrating out all F-bosons in the Higgs sector yields the mass term for the Higgs, just as in the Wess-Zumino model. This produces

$$|\mu|^2 (H_u^i H_u^{i\dagger} + H_d^i H_d^{i\dagger}).$$

This implies firstly that both Higgs doublets are mass-degenerate. More seriously, it implies that this common mass is positive. In the conventional treatment of electroweak symmetry, however, a negative mass is mandatory to obtain a perturbatively valid description of electroweak symmetry. Since the appearance of the positive mass is a direct consequence of supersymmetry, as was seen in case of the Wess-Zumino model, there seems to be no possibility to have at the same time perturbative symmetry breaking and intact supersymmetry. One alternative would be non-perturbative effects, which has not been excluded so far. However, it seems somewhat more likely that it is not possible to have unbroken supersymmetry but so-called broken gauge symmetry simultaneously, and this has lead to a search for a common origin of both phenomena.

After analyzing the features which make the MSSM different from the standard model, it is instructive to see how the standard-model-type interactions are still present.

This is straightforward in the case of the gauge-boson self-interactions, as these are automatically included in the non-Abelian field-strength tensors appearing already in the supersymmetric version of Yang-Mills theory, (4.44). It is more interesting to investigate how the usual Dirac-type quarks and gluons and their coupling to the strong and weak interactions are recovered.

The simpler case is the parity-preserving strong interactions. Due to the gauge-covariant derivative, the coupling of a quark-type left-handed Weyl-fermion is

$$-\frac{1}{2} g \chi_q^\dagger \bar{\sigma}^\mu A_\mu \chi_q = -\frac{1}{2} g \chi_{qi}^\dagger \bar{\sigma}_\mu A_{ij}^\mu \chi_{qj}, \tag{4.62}$$

where in the second expression the gauge indices are made explicit, keeping in mind that $A_\mu = \tau^\alpha A_\mu^\alpha$ is matrix-valued. The right-handed contribution can be rewritten as a left-handed Weyl fermion by virtue of charge conjugation,

$$\begin{aligned}\psi_u &= \chi_{\bar{q}}^c = i\sigma_2\chi_{\bar{q}}^*.\\ \chi_{\bar{q}} &= -i\sigma_2\chi_{\bar{q}}^{c*}\end{aligned}$$

The original coupling of the right-handed fermion, which will be the anti-particle, is

$$\begin{aligned}\frac{1}{2}g\chi_{\bar{q}}^\dagger\bar{\sigma}^\mu A_\mu^*\chi_{\bar{q}} &= \frac{1}{2}g(-i\sigma_2\chi_{\bar{q}}^{c*})^\dagger\bar{\sigma}^\mu A_\mu^*(-i\sigma_2\chi_{\bar{q}}^{c*})\\ = -\frac{1}{2}g\chi_{\bar{q}}^{cT}\sigma_2\bar{\sigma}_\mu\sigma_2 A_\mu^*\chi_{\bar{q}}^{c*} &= -\frac{1}{2}g\chi_{\bar{q}}^{c\dagger}\sigma^\mu A_\mu\chi_{\bar{q}}^c. \qquad (4.63)\end{aligned}$$

In the last step, it has been used that $\sigma_2\bar{\sigma}^\mu\sigma_2 = -\sigma^{\mu T}$, and that $A_\mu^* = A_\mu^T$, since the A_μ are hermitian. The remaining step is just rewriting everything in indices and rearranging. Combining (4.62) and (4.63) yields

$$-\frac{1}{2}g(\chi_{\bar{q}}^{c\dagger}\sigma^\mu A_\mu\chi_{\bar{q}}^c + \chi_q^\dagger\bar{\sigma}^\mu A_\mu\chi_q) = -\frac{1}{2}g\bar{\Psi}\gamma^\mu A_\mu\Psi$$

with $\Psi^T = (\chi_{\bar{q}}^c\chi)$. This is precisely the way an ordinary Dirac quark Ψ would couple covariantly to gluons. Hence, by combining two chiral multiplets and one gauge multiplet, it is possible to recover the couplings of the standard model strong interactions.

The electromagnetic interaction, which is also parity-preserving, emerges in the same way, just that the photon field is not matrix-valued.

The weak interaction violates parity maximally by just coupling to the left-handed components. In the standard model, its coupling is given by

$$-\frac{1}{2}g'\bar{\Psi}_e\frac{1-\gamma_5}{2}\gamma^\mu W_\mu\frac{1-\gamma_5}{2}\Psi.$$

However, the action of $(1-\gamma_5)/2$ on any spinor is to yield

$$\frac{1-\gamma_5}{2}\begin{pmatrix}\chi_{\bar{e}}\\ \chi_e\end{pmatrix} = \begin{pmatrix}0\\ \chi_e\end{pmatrix},$$

and thus reduces the Dirac-spinor to its left-handed component. Thus, only the coupling

$$-\frac{1}{2}g\chi_e^\dagger\bar{\sigma}^\mu W_\mu\chi_e$$

remains. This is precisely the coupling of a left-handed chiral multiplet to a gauge multiplet. It is therefore sufficient just to gauge the left-handed chiral multiplets with the weak interactions to obtain the standard-model-type coupling. In addition also inter-generation mixing has to be included, but this proceeds in the same way as in the the standard-model, and is therefore not treated explicitly. This concludes the list of standard model couplings.

4.10.2 Breaking the supersymmetry in the MSSM

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

These parameters include masses for all the non-standard model superpartners, that is squarks and sleptons, as well as photinos, gluinos, winos, and binos. Only for the Higgsinos it is not possible to construct a gauge-invariant additional mass-term, due to the chirality of the weak interactions, in much the same way as for quarks and leptons in the standard model. One advantage is, however, offered by the introduction of these free mass parameters: It is possible to introduce a negative mass for the Higgs particles, reinstantiating the same way to describe the breaking of electroweak symmetry as in the standard model. That again highlights how electroweak symmetry breaking and supersymmetry breaking may be connected. These masses also permit to shift all superpartners to such scales as they are in accordance with the observed limits so far. However, if the masses would be much larger than the scale of electroweak symmetry breaking, i. e., 250 GeV, it would again be very hard to obtain results in agreement with the observations without fine-tuning. However, such mass-matrices should not have large off-diagonal elements, i. e., the corresponding CKM-matrices should be almost diagonal. Otherwise, mixing would produce flavor mixing also for the standard model particles exceeding significantly the observed amount.

In addition, it is possible to introduce triple-scalar couplings. These can couple at will the squarks, sleptons, and the Higgs bosons in any gauge-invariant way. Again, observational limits restrict the magnitude of these couplings. Furthermore, some couplings, in particular certain inter-family couplings, are very unlikely to emerge in any kind of proposed supersymmetry breaking mechanism, and therefore are usually omitted.

Finally, the mass-matrices in the Higgs sector may have off-diagonal elements, without introducing mixing in the quark sector. This are contributions of the type

$$b(H_u^1 H_d^2 - H_u^2 H_d^1) + h.c.,$$

where the mass parameter b may also be complex.

The arbitrariness of these enormous amount of coupling constants can be reduced, if some model of underlying supersymmetry breaking is assumed. One, rather often, invoked one is that the minimal supersymmetric standard model is the low-energy limit of a supergravity theory, the so-called mSUGRA scenario. Though by now experimentally ruled out, it is still interesting as the simplest representative of such types of models.

In the mSUGRA case, it is, e. g., predicted, that the masses of the superpartners of the gauge-boson superpartners should be degenerate at the breaking scale,

$$M_{\tilde{g}} = M_{\tilde{W}} = M_{\tilde{B}} = m_{1/2}, \tag{4.64}$$

and also the masses of the squarks and sleptons should be without mixing and degenerate

$$m^2_{\tilde{Q}} = m^2_{\tilde{q}} = m^2_{\tilde{L}} = m^2_{\tilde{l}} = m_0^2\mathbf{1}, \tag{4.65}$$

and these masses also with the ones of the Higgs particles,

$$m_H^2 = m_0^2.$$

Furthermore, all trilinear bosonic couplings would be degenerate, with the same value A_0. If the phase of all three parameters, A_0, $m_{1/2}$, and m_0 would be the same, also CP violation would not differ from the one of the standard model. The latter is an important constraint, as the experimental limits on such violations are very stringent, although not yet threatening to rule out the MSSM proper.

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

4.10.3 MSSM phenomenology

4.10.3.1 Coupling unification and running parameters

In the electroweak theory it is found that at some energy scale the electromagnetic and the weak coupling become both of the same value. This is known as electroweak unification. Of course, if also the strong coupling would become of the same value at the same energy, this would strongly indicate that all three couplings originate from one coupling at this unification scale, and become different at low energies because the gauge group to which the unified coupling corresponds becomes broken at this unification scale. This is also the idea behind so-called grand-unified theories (GUT) that at large energies there exists one gauge-group, say SU(5), which becomes broken by a Higgs effect, similar to that one in the standard model, to yield the product gauge group SU(3)×SU(2)×U(1) of the standard model at the unification scale. This unification is of course also interesting, as any evidence of it would support the idea of gauge-mediated supersymmetry-breaking. This idea will be taken up again in chapter 7. But supsymmetry enforces the idea, as will be discussed now.´

To check, whether such a unification actually occurs, it is necessary to determine the running of the effective coupling constants with the energy (renormalization) scale. There are three couplings[18], the strong one g, the weak one g', and the electromagnetic one. It is always possible to write the electromagnetic one as $e/\sin\theta_W$. The angle θ_W is known as the Weinberg angle. In the standard model its value originates from the electroweak symmetry breaking. Its measured value is 28.7°. The relevant couplings are then

$$\alpha_1(Q) = \tfrac{g(Q)^2}{4\pi} =_{Q=m_Z} 0.119 \tag{4.66}$$

$$\alpha_2(Q) = \tfrac{g'(Q)^2}{4\pi} =_{Q=m_Z} 0.0338 \tag{4.67}$$

$$\alpha_3(Q) = \tfrac{5}{3}\tfrac{e(Q)^2}{4\pi\cos^2\theta_W} =_{Q=m_Z} 0.0169 \tag{4.68}$$

In this case the appropriate mixed combination for the electromagnetic coupling has been used. The factor 5/3 appears as in an Abelian gauge theory there is a certain freedom in redefining the charge and the generator of gauge transformation not present in non-Abelian gauge theories. It has been set here to a conventional value, which would be expected if the standard model product gauge groups would indeed originate from a common one at high energies.

To one-loop order, all coupling constants evolve according to the renormalization group equation

$$\frac{d\alpha_i}{d\ln Q} = -\frac{\beta_i}{2\pi}\alpha_i^2, \tag{4.69}$$

where β_i is the first coefficient in a Taylor expansion of the so-called β-function for the coupling i. For a non-Abelian coupling its value is given by

$$\beta_i = \frac{11}{3}C_A - \frac{2}{3}C_A n_f^a - \frac{1}{3}n_f - \frac{1}{6}n_s,$$

where C_A is the adjoint or second Casimir of the gauge group with value N for a SU(N) group. The numbers n_f^a, n_f, and n_s are the number of adjoint fermionic chirality states, fundamental fermionic chirality states and the number of complex scalars charged under this coupling, respectively. In the Abelian case, the value of β_i is given by

$$\beta_i = -\frac{2}{3}\sum_f Y_f^2 - \frac{1}{3}\sum_s Y_s^2,$$

with Y_f and Y_s the charge of fermions and scalars with respect to the interaction. This stems from the fact that the charge for matter minimally coupled to an Abelian gauge-field

[18]Other couplings, like the Higgs-self-coupling or the Yukawa couplings do not unify at this scale.

can be chosen freely, while this is not the case for a non-Abelian case. The fact that the Y_f and Y_s are rational numbers in the standard model is another indication for the U(1) part of the standard model to be emerged from some other gauge group by symmetry breaking.

As the coefficients of the β-functions do not depend on the couplings themselves, the differential equation (4.69) is readily integrated to yield

$$\alpha_i(Q)^{-1} = \alpha_i(Q_0)^{-1} + \frac{\beta_i}{2\pi}\ln\frac{Q}{Q_0},$$

where $\alpha_i(Q_0)$ is the initial condition, the value of the coupling at some reference scale Q_0. As already indicated in (4.66-4.68), this reference scale will be the Z-boson mass, as all three couplings have been measured with rather good precision at this scale at LEP and LEP2.

The question to be posed is now whether the three couplings $\alpha_i(Q)$ have at some energy Q the same value. To obtain a condition, it is most simple to use the linear system of equations to eliminate Q and $\alpha_i(Q)$ in favor of the known measured values. This yields the conditional equation

$$B_x = \frac{\alpha_3(m_Z)^{-1} - \alpha_2(m_Z)^{-1}}{\alpha_2(m_Z)^{-1} - \alpha_1(m_Z)^{-1}} = \frac{\beta_2 - \beta_3}{\beta_1 - \beta_2} = B_t.$$

With the values given in (4.66-4.68) the left-hand side is readily evaluated to be $B_x = 0.72$. The right-hand side depends only on the β-functions, and therefore to this order only on the particle content of the theory.

In the standard model, there are no scalars charged under the $N = 3$ strong interactions, but 12 chirality states of fermions, yielding $\beta_1 = 7$. The positive value indicates that the strong sector is described by an asymptotically free theory. For the $N = 2$ weak interactions, there are 12 left-handed chiral states (6 quarks and 6 leptons), and one Higgs field, yielding $\beta_2 = 19/6$. Taking all electromagnetically charged particles, and the normalization factor 5/3, into account, $\beta_3 = -41/10$. Note that also some components of the Higgs field are electromagnetically charged. Evaluating B_t yields $115/218 \approx 0.528$. Thus, in the standard model, to this order the couplings will not match. This is also not changed in higher orders of perturbation theory.

The situation changes in the MSSM. For β_1, there are now in addition the gluinos, giving one species of adjoint fermionic chiralities, and 12 squarks, yielding $\beta_1 = 3$. The same applies to the weak case with one species of adjoint winos and zinos, 12 sleptons and squarks, two higgsinos and one additional Higgs doublet. Altogether this yields $\beta_2 = -1$. The change of sign is remarkably, indicating that the supersymmetric weak sector is no longer asymptotically free, as is the case in the standard model. The anti-screening contributed from the additional degrees of freedom is sufficient to change the behavior of the theory qualitatively. The case of β_3 finally changes to $-33/5$, after all bookkeeping is done. Together, this yields $B_t = 5/7 = 0.714$. This is much closer to the desired value of 0.72, yielding support for the fact that the MSSM emerges from a unified theory. Taking this result to obtain the unification scale, it turns out to be $Q_U \approx 2.2 \times 10^{16}$ GeV, which is an enormously large scale, though still significantly below the Planck scale of 10^{19} GeV. This result is not changing qualitatively, if the calculation is performed to higher order nor if the effects from supersymmetry breaking and breaking of the presumed unifying gauge group is taken into account. Thus the MSSM, without adding any constraint, seems a natural candidate for a theory emerging from a grand unified one, being one reason for its popularity and of supersymmetry in general. However, exact unification is never achieved in the MSSM, and therefore requires further contributions to occur.

These results suggest to also examine the running of other parameters as well. In particular, the parameters appearing due to the soft breaking of supersymmetry are particularly interesting, as their behavior will be the effect which obscures supersymmetry in nature, if it exists.

The first interesting quantity is the mass of the gauge-boson superpartners, the gauginos. The running is given by a very similar expression as for the coupling constants,

$$\frac{dM_i}{d\ln Q} = -\frac{\beta_i}{2\pi}\alpha_i M_i.$$

The β-function can be eliminated using the equation for the running of the coupling (4.69). This yields the equation

$$0 = \frac{1}{\alpha_i}\frac{dM_i}{d\ln Q} - \frac{M_i}{\alpha_i^2}\frac{d\alpha_i}{d\ln Q} = \frac{d}{d\ln Q}\frac{M_i}{\alpha_i}.$$

This implies immediately that the ratio of the gaugino mass for the gauge group i divided by the corresponding gauge coupling is not running, i. e., it is renormalization group-invariant.

If there exists a scale m_U at which the theory unifies, like suggested by the running of the couplings, then also the masses of the gauginos should be equal. As discussed previously, this would be the case in supergravity as the origin of the minimal supersymmetric standard model. This would yield

$$\frac{M_i(m_U)}{\alpha_i(m_U)} = \frac{m_{1/2}}{\alpha_U(m_U)},$$

and since the ratios are renormalization-group invariant it follows that

$$\frac{M_1(Q)}{\alpha_1(Q)} = \frac{M_2(Q)}{\alpha_2(Q)} = \frac{M_2(Q)}{\alpha_2(Q)}.$$

Since the α_i are known, e. g., at $Q = m_Z$, it is possible to deduce the ratio of the gaugino masses at the Z-mass, yielding

$$\begin{aligned} M_3(m_Z) &= \tfrac{\alpha_3(m_Z)}{\alpha_2(m_Z)} M_2(m_Z) \approx \frac{1}{2} M_2 \\ M_1(m_Z) &= \tfrac{\alpha_1(m_Z)}{\alpha_2(m_Z)} M_2(m_Z) \approx \frac{7}{2} M_2. \end{aligned} \tag{4.70}$$

This implies that the masses of the gluinos to the winos to the bino behave as $7:2:1$. This implies that the gluino is the heaviest of the gauginos. Note that the masses are not necessarily the masses of the original gauge bosons in the electroweak sector. The bino and the winos will together with the higgsinos form the final mass eigenstates. Therefore, the running of the parameters will provide already guiding predictions on how supersymmetry breaking is provided in nature, if the observed mass pattern of these particles matches this prediction, provided once more that higher-order and non-perturbative corrections are small.

An even more phenomenologically relevant result is obtained by the investigation of the scalar masses, i. e., the masses of the Higgs, the squarks, and the sleptons. Retaining only the top-quark Yukawa coupling y_t, which dominates all other Yukawa couplings at one loop, and a unified three-scalar coupling A_0, the corresponding third-generation evolution equations and Higgs evolution equation are given by

$$\frac{dm^2_{H_u}}{d\ln Q} = \frac{1}{4\pi}\left(\frac{3X_t}{4\pi} - 6\alpha_2 M_2^2 - \frac{6}{5}\alpha_3 M_3^2\right) \tag{4.71}$$

$$\frac{dm^2_{H_d}}{d\ln Q} = \frac{1}{4\pi}\left(-6\alpha_2 M_2^2 - \frac{6}{5}\alpha_3 M_3^2\right)$$

$$\frac{dm^2_{\tilde{t}_L}}{d\ln Q} = \frac{1}{4\pi}\left(\frac{X_t}{4\pi} - \frac{32}{3}\alpha_1 M_1^2 - 6\alpha_2 M_2^2 - \frac{2}{15}\alpha_3 M_3^2\right) \tag{4.72}$$

$$\frac{dm^2_{\tilde{t}_R}}{d\ln Q} = \frac{1}{4\pi}\left(\frac{2X_t}{4\pi} - \frac{32}{3}\alpha_1 M_1^2 - \frac{32}{15}\alpha_3 M_3^2\right) \tag{4.73}$$

$$X_t = 2|y_t|^2\left(m^2_{H_u} + m^2_{\tilde{t}_l} + m^2_{\tilde{t}_r} + A_0^2\right).$$

A number of remarks are in order. The quantity X_t emerges from the Yukawa couplings to the top quark. Therefore, the down-type Higgs doublet will, to leading order, not couple, and therefore its evolution equation will not depend on this strictly positive quantity. Furthermore, the Higgs fields couple to leading order not to the strong interaction, while the squarks do. Hence the former have no term depending on α_1, while the latter do. Similarly, only squarks from the left-handed chiral multiplet couple directly to the weak interactions, and therefore receive contributions from the weak interaction proportional to α_2. Still, all of these particles couple electromagnetically, and therefore receive contributions proportional to α_3.

Since all couplings α_i are strictly positive, the values of the masses can only decrease by the top-quark contribution X_t when lowering the scale Q. In particular, this implies that the mass of H_d will, at this order, only increase or at best stay constant. The strongest decrease will be observed for the H_u contribution, as the factor three in front of X_t magnifies the effect. Furthermore, the largest counteracting contribution ($7^2 = 49$ and $2^2 = 4$ and the largest α_1!) due to the gluino term is absent for the Higgs fields. Therefore, the mass of the H_u will decreases fastest from its unification value $\sqrt{\mu^2 + m_0^2}$ at the unification scale. In fact, this decrease may be sufficiently strong to drive the mass parameter negative at the electroweak scale. This would trigger electroweak symmetry breaking by the same mechanism as in the ordinary standard model. Using parameters which prevent the squark masses from becoming negative, and thus preserving vitally the color gauge symmetry, this indeed happens. Again, explicitly broken, but unified, supersymmetry provides the correct low-energy phenomenology of the standard model.

4.10.3.2 The electroweak sector

One of the most interesting questions in contemporary standard model physics is the origin of the value of the mass of the Higgs boson. In the standard model, this is an essentially independent parameter. However, in the minimal supersymmetric extension of the standard model, this parameter is less arbitrary, due to the absence of the hierarchy problem, and the fact that part of the Higgs couplings are determined by supersymmetry, like the four-Higgs coupling.

To determine the Higgs mass, it is first necessary to determine the electroweak symmetry breaking pattern. To do this, the first step is an investigation of the Higgs potential. This is more complicated than in the standard model case, due to the presence of a second Higgs doublet.

At tree-level, the quadratic term for the Higgs fields is determined by the contribution from the supersymmetric invariant term, and the two contributions from explicit breaking, yielding together

$$V_1 = (|\mu|^2 + m_{H_u}^2)(H_u^+ H_u^{+\dagger} + H_u^0 H_u^{0+}) + (|\mu|^2 + m_{H_d}^2)(H_d^0 H_d^{0\dagger} + H_d^- H_u^{-\dagger}) + b(H_u^+ H_d^- - H_u^0 H_d^0) + h.c.,$$

where the Higgs fields are labeled by their electric instead of the hypercharge. The parameters $m_{H_i}^2$ can have, despite their appearance, both signs.

The quadratic part of the potential originates from two contributions. Both are from the D couplings and F couplings for the groups under which the Higgs fields are charged, the weak isospin group SU(2) and the hypercharge group U(1). Since these contributions are four-point vertices, there are no contributions from explicit supersymmetry breaking.

Thus, the quartic part of the potential takes the form

$$V_2 = \frac{e^2+g'^2}{8}(H_u^+H_u^{+\dagger} + H_u^0H_u^{0+} - H_d^0H_d^{0\dagger} - H_d^-H_d^{-\dagger})^2 + \frac{g'^2}{8}|H_u^+H_d^{0\dagger} + H_u^0H_d^{-\dagger}|^2,$$

where g' is the gauge coupling of the weak isospin gauge group and e the one of the weak hypercharge group.

To obtain the experimentally measured electroweak phenomenology, this potential has to have a non-trivial minimum, especially if all other fields vanish. The latter also implies that the cubic coupling appearing in the Lagrangian due to the supersymmetry breaking are not relevant for this question, as they always involve at least one other field.

It is possible to simplify this question. The expressions are invariant under a local gauge transformation, as is the complete Lagrangian. If therefore any of the fields has a non-vanishing value at the minimum, it is always possible to perform a gauge transformation such that a specific component has this, and the other ones vanish. Choosing then H_u^+ to be zero, the potential must be extremal at this value of H_u^+. Finding a minimum is then a classical analysis. Requiring further that electromagnetism is unbroken yields $H_d^- = 0$. Therefore, only the electrically neutral components of both Higgs doublets matter. To avoid CP violations, H_u^0 and H_d^0 must also be real. If there should exist a non-trivial minimum, this implies that the product $H_u^0H_d^0$ must be positive, as otherwise all terms would be positive. By a global U(1) gauge transformation, both fields can then be chosen to be positive.

The direction $H_u^0 = H_d^0$ is pathologically, as the highest-order term vanishes, a so-called flat direction. For the potential to be still bounded from below, thus providing a perturbatively stable vacuum state, requires

$$2|\mu|^2 + m_{H_u}^2 + m_{H_d}^2 > 2b(>0).$$

Therefore, not the masses of both H_u^0 and H_d^0 can be negative simultaneously.

It is not possible to restrict the solution and the parameters further just from the Lagrangian at this point. However, it is experimentally possible to fix at least a particular combination of the associated vacuum expectation values $\langle H_u^0\rangle$ and $\langle H_d^0\rangle$. Studying the coupling to the electroweak gauge bosons masses yields a mass term in the Lagrangian of

$$\left(\frac{1}{2}(e^2+g'^2)Z_\mu Z^\mu + \frac{1}{2}g^2(W_\mu^+W^{\mu+} + W_\mu^-W^{\mu-})\right)(\langle H_u^0\rangle^2 + \langle H_d^0\rangle^2). \tag{4.74}$$

The value for the combined condensate is therefore the same[19] as in the standard model, $(174\text{ GeV})^2$.

[19] Note a factor of $\sqrt{2}$ due to the MSSM conventions.

It is furthermore possible to determine mass-bounds for the Higgs. Defining

$$\tan\beta = \frac{\langle H_u\rangle}{\langle H_d\rangle},$$

the resulting expressions become relatively simple. It is furthermore useful that the mass-matrix assumes a block-diagonal form, so that it is possible to reduce the complexity further by studying only doublets at each time.

The first doublet to be studied are the (non-condensing) imaginary parts of H_u^0 and H_d^0. This yields two linear combinations as mass eigenstates. One is a massless mode, and will become effectively the longitudinal component of the Z^0. The orthogonal combination is commonly referred to as the A^0, and is a pseudo-scalar. This is an uncharged second Higgs field (the first one will be one of the condensed ones). It is one of the extra Higgs particles not present in the standard model. In practical calculations, the parameter b is often traded for the mass of this particle, $m_- = m_{A^0}$.

The next pair is H_u^+ and H_d^+. Linear combinations will have masses 0 and $m_W^2 + m_{A^0}^2$. The charged massless combinations will become the longitudinal component of the $W^\pm$. The other states, usually just called $H^\pm$, correspond to an electrical charged Higgs particle, which is not appearing in the standard model.

Finally, masses for the condensing real parts of H_u^0 and H_d^0 are

$$\begin{aligned} m_{h^0}^2 &= \frac{1}{2}\left(m_{A^0}^2 + m_Z^2 - \sqrt{(m_{A^0}^2 + M_Z^2)^2 - 4m_{A^0}^2 m_Z^2\cos^2(2\beta)}\right) \\ m_{H^0}^2 &= \frac{1}{2}\left(m_{A^0}^2 + m_Z^2 + \sqrt{(m_{A^0}^2 + M_Z^2)^2 - 4m_{A^0}^2 m_Z^2\cos^2(2\beta)}\right) \end{aligned}$$

for the twolinear combinations h^0 and H^0, giving two more neutral, scalar Higgs particles. So instead of the one of the standard model, there are two in the minimal supersymmetric standard model.

The masses for the fields A^0, $H^\pm$ and H^0 are all containing a contribution $m_{A^0}\sqrt{2b/sin(2\beta)}$, which is unconstrained, and could, in principle, become arbitrarily large. This is not the case for the mass of h^0. If the A^0 mass would be small, it is possible to expand the root, yielding

$$m_{h^0}^2 \approx m_{A^0}^2\cos^2(2\beta) \leq m_{A^0}^2.$$

For large masses of the A^0 it becomes

$$m_{h^0}^2 \approx m_Z^2\cos^2(2\beta) \leq m_Z^2, \tag{4.75}$$

where only the experimental known mass of the Z-boson enters. Thus, though the bounds are not known, the mass is constrained. Since the experimental bounds for the A^0 indicate a rather large mass, the second expression (4.75) is more appropriate. Unfortunately, such a low mass for the lightest neutral Higgs boson is excluded experimentally. Even if the more precise formulas would be used, instead of the approximate ones, the situation is not improving qualitatively.

That could have been already a dismissal of the minimal version of a supersymmetric standard model. However, the leading quantum correction to this bound yields

$$m_{h^0}^2 \leq m_Z^2 + \frac{3m_t^4}{2\pi^2(\langle H_u\rangle^2 + \langle H_d\rangle^2)} \ln \frac{\sqrt{m_{\tilde{t}_1}^2 + m_{\tilde{t}_2}^2}}{\sqrt{2}m_t}, \tag{4.76}$$

where m_t is the known top quark mass and $m_{\tilde{t}_i}$ are the masses of the staus after mixing between the left and right multiplets occurred, which, in principle, can be different. Already for masses of the order of the experimentally excluded stau masses the bound is increased by these radiative corrections above the Higgs mass. On the one hand, this is good for the minimal supersymmetric standard model, but on the other hand this implies that leading order corrections are large, and subleading corrections may be relevant. This reduces the predictiveness of the bound, as then the other parameters enter in various ways.

There is a further problem with the corrections (4.76) to the Higgs mass. The condition for forming a Higgs condensate can be reformulated as

$$\frac{1}{2}m_Z^2 = -|\mu|^2 + \frac{m_{H_d}^2 - m_{H_u}^2 \tan^2\beta}{\tan^2\beta - 1}.$$

For example let $\tan\beta$ become large, i. e., the condensate $\langle H_u\rangle$ is much larger than the condensate $\langle H_d\rangle$. Then the condition becomes

$$\frac{1}{2}m_Z^2 = -|\mu|^2 - m_{H_u}^2.$$

Both parameters on the right-hand side are not constrained immediately by physics. However, if both parameters would be much larger than the mass of the Z boson, it would be necessary for them to cancel almost exactly. If this would be the case, it would immediately raise the same questions as in the original fine-tuning problem of the standard model if the Higgs mass would be much larger than the ones of the weak bosons. This is therefore also called the little fine-tuning problem or little hierarchy problem. The experimentally established mass of the Higgs is, in fact, borderline. Together with the large lower limits for the masses of the other sparticles, the MSSM, though still consistent with experiment, becomes in fact increasingly fine-tuned, therefore abolishing its original motivation. Still, it is not excluded, and may yet describe nature.

That this is indeed somewhat of a problem becomes apparent when considering the renormalization constant for the mass of the Higgs boson, which is given by the integral of (4.71), and is approximately

$$\delta m_{H_u}^2 \approx -\frac{3y_t^2}{8\pi^2}(m_{\tilde{t}_1}^2 + m_{\tilde{t}_2}^2)\ln\left(\frac{\sqrt{2}\Lambda_U}{\sqrt{m_{\tilde{t}_1}^2 + m_{\tilde{t}_2}^2}}\right), \tag{4.77}$$

where it was assumed that the stau masses give the dominant contributions, due to the condition (4.76). The value of the unification scale is still of the order of 10^{16} GeV. If therefore the mass shift due to the leading quantum corrections should be small, this yields approximately that the geometric average of the stau masses should not be larger than about 150 GeV. Otherwise the quantum corrections alone would produce a fine-tuning problem. This condition is, however, in violation of the bound of 500 GeV necessary to shift the h^0 Higgs boson out of the current experimental reach. Therefore, it cannot be permitted. Shifting then the stau masses to the required value yields a mass correction large compared to the Z-mass, actually, it becomes exponentially worse due to the logarithmic dependence. Thus a fine-tuning problem arises. Whether this constitutes a problem, or just an aesthetic displeasure, is not only a question of personal taste. It also is a challenge to understand whether nature prefers for what reasons theories with or without finetuning.

4.10.3.3 Mass spectrum

So far, the masses of the Higgs bosons and the electroweak bosons have been calculated. The gluons also remain massless, in accordance with observation. The remaining mass spectrum for the minimal supersymmetric standard model at tree-level will be discussed in the following.

The first particles are the remaining ones from the standard model. These have to acquire, of course, their observed masses. Due to the parity violating nature of the weak interactions, these masses can effectively arise only due to the Yukawa interaction with the Higgs particles. In unitary gauge, these couplings can be split into a contribution which contains only the Higgs vacuum expectation values

$$y\chi\chi H = y\chi\chi\langle H\rangle + y\chi\chi(H - \langle h\rangle)$$

and behave therefore as mass terms. Thus, the masses are given by

$$\begin{aligned} m_{uct} &= y_{uct}\langle H_u\rangle = y_{uct}\frac{\sqrt{2}m_W\sin\beta}{g'} \\ m_{dsb} &= y_{dsb}\langle H_d\rangle = y_{dsb}\frac{\sqrt{2}m_W\cos\beta}{g'} \\ m_{e\mu\tau} &= y_{e\mu\tau}\langle H_d\rangle = y_{e\mu\tau}\frac{\sqrt{2}m_W\cos\beta}{g'} \\ m_{\nu_e\nu_\mu\nu_\tau} &= y_{\nu_e\nu_\mu\nu_\tau}\langle H_u\rangle = y_{\nu_e\nu_\mu\nu_\tau}\frac{\sqrt{2}m_W\sin\beta}{g'}. \end{aligned}$$

The twelve Yukawa couplings are all free parameters of the minimal supersymmetric standard model. As is visible, it strongly depends on the value of β, whether these couplings are strong, preventing perturbative descriptions, or weak enough that at sufficiently high energies perturbation theory is adequate.

The situation for the gluinos is actually simpler. Since the color symmetry is unbroken, no other fermions exist with the quantum numbers of the gluinos, and they do not couple to the parity violating weak interactions. Thus, the only contribution comes from the explicit supersymmetry breaking term

$$-\frac{1}{2}M_1\tilde{g}\tilde{g} - \frac{1}{2}M_1^*\tilde{g}^\dagger\tilde{g}^\dagger.$$

The mass parameter can be complex in general, but the corresponding tree-level mass will just be its absolute value. Therefore, the value of the mass of the gluinos is unconstrained in the minimal supersymmetric standard model, but by virtue of relation (4.70) it is tied to the masses of the other gauginos in case of unification.

The situation becomes more complicated for the so-called neutralinos, the superpartners of W^0 and B, the wino $\tilde{W}^0$ and the bino $\tilde{B}$. These fermions are electrically uncharged, and can both mix, similar to their standard-model versions. Furthermore, the neutral superpartners of the two Higgs fields, the higgsinos $\tilde{H}_d^0$ and $\tilde{H}_u^0$ are both also fermionic and uncharged. Hence, these can mix with the binos and the winos as well. The gauge eigenstate is therefore

$$\tilde{G}^{0T} = (\tilde{B}, \tilde{W}^0, \tilde{H}_d^0, \tilde{H}_u^0).$$

The most direct mixing is due to the interaction mediated by the weak F-boson in the Wess-Zumino-like contribution to the superpotential, which leads to the contributions

$$\frac{1}{2}\mu(\tilde{H}_u^0\tilde{H}_d^0 + \tilde{H}_d^0\tilde{H}_u^0) + \frac{1}{2}\mu^*(\tilde{H}_u^{0\dagger}\tilde{H}_d^{0\dagger} + \tilde{H}_d^{0\dagger}\tilde{H}_u^{0\dagger}).$$

Furthermore, the weak gauge symmetry and supersymmetry demand the existence of couplings of the generic type

$$-\sqrt{2}g'H_u\tilde{H}_u\tilde{W} = -\sqrt{2}g'(H_u + \langle H_u\rangle)\tilde{H}_u\tilde{W} \tag{4.78}$$

to ensure supersymmetry in the super Yang-Mills part of the weak symmetry. If the Higgs fields condense, these yield a mixing term proportional to the condensates. The condensate is uncharged, and therefore this mixing can only combine two neutral fields or two of opposite charge. Hence, only the fields $\tilde{H}_d^0$ and $\tilde{H}_u^0$ will mix with the neutral wino and the bino by this mechanism. Finally, the wino and the bino can have masses due to the explicit supersymmetry breaking, similar to the gluinos,

$$-\frac{1}{2}M_2\tilde{W}^0\tilde{W}^0 - \frac{1}{2}M_3\tilde{B}\tilde{B} + h.c..$$

No such contribution exist for the higgsinos, as it is not possible to write down such a term while preserving explicitly the weak gauge symmetry. Thus, all four fields mix. The mass eigenstates are denoted by $\tilde{\chi}_i$ with $i = 1, ..., 4$, and are called neutralinos. These particles interact only weakly, like the neutrinos, and hence their name. Since the a-priori unknown parameters M_2, M_3 and μ, as well as β, enter their mass matrix, their masses cannot be predicted. However, if any of the neutralinos would be the lightest supersymmetric particles, then by virtue of R-parity conservation it would be stable. Since it interacts so weakly, it would be a perfect candidate for dark matter, which cannot be provided by neutrinos since their mass is too small. In fact, at least for some range of parameters the masses of the neutralinos would be such that they are perfectly compatible with the properties required for cold, non-baryonic, dark matter.

A similar situation arises for the charged counterpart of the neutralinos, the charginos. These stem from the mixing of the charged higgsinos and the charged winos. Since the positively and negatively charged particles cannot mix, two doublets appear instead of one quartet,

$$\begin{aligned}\tilde{G}^+ &= \begin{pmatrix}\tilde{W}^+\\ \tilde{H}_u^+\end{pmatrix}\\ \tilde{G}^- &= \begin{pmatrix}\tilde{W}^-\\ \tilde{H}_d^-\end{pmatrix}\end{aligned} \tag{4.79}$$

where the charged winos are linear combinations $\tilde{W}^1 \pm i\tilde{W}^2$ of the off-diagonal winos. Similar as in the case of the neutralinos, there is a contribution from the weak F-term and the condensation of the neutral Higgs fields from the super Yang-Mills action, which mixes both components of each doublet. Performing the algebra leads to masses of the charginos

$$|m_{\tilde{\chi}_i^\pm}|^2 = \frac{1}{2}\left(M_2^2 + |\mu|^2 + 2m_W^2) \mp \sqrt{(M_2^2 + |\mu|^2 + 2m_W^2)^2 - 4|\mu M_2 - m_W^2 \sin(2\beta)|^2}\right)$$

where the upper and lower sign refer to the two members of each charged doublet. Therefore there is always one pair of oppositely charged charginos having the same mass, as would be expected from CPT.

Thus remains the largest group of additional particles, the sfermion superpartners of the fermionic standard model particles. These contain the squarks and sleptons. Fortunately, due to the presence of the strong charge, squarks and sleptons will not mix. However, the three families within each sector could in principle mix. Since such mixings would have to be very small to be not in conflict with flavor changing currents observed in experiments, these will be neglected[20]. Furthermore, compared to the scale of the supersymmetric parameters, only the Yukawa couplings of the third family can give any significant contribution, and will only be considered in this case. Therefore, the first two and the third family will be treated in turns.

The masses are in the end driven by the common scale at unification, m_0, and then various contributions from running. Most remarkable, the squark masses will tend to be larger than the slepton masses. Also, the weak isospin splittings give a measure of β as

$$m_{\tilde{d}_L}^2 - m_{\tilde{u}_L}^2 = m_{\tilde{e}_L}^2 - m_{\tilde{\nu}_e}^2 = -m_W^2 \cos(2\beta),$$

and in the same manner for the second family. This constraint could be used to experimentally verify and/or predict parameters of the theory. The present experimental constraints favor a rather large β, such that the down-type sfermions would be heavier.

The situation for the third family is similar, though the additional Yukawa coupling tends to drive the masses to smaller values than for the first two families. In fact, in most scenarios the stop will be the lightest of the squarks, and in some regions of the parameter space the mass would be driven to negative values and thus initiate the breaking of color symmetry, excluding these values for the parameters. The situation for the sbottom, stau, and stau sneutrino is similar, just that all parameters are exchanged for their equivalent of that type. Because their Yuakaw couplings are smaller the masses of these super-partners will be larger than the ones of the stops.

[20]In fact, the absence of strong mixing sets strong limits on the properties of the mass matrices in the squark and slepton sector. Since these are only parameters in the MSSM, this raises the questions why the should be so specifically shaped. It is often assumed that whatever mechanism drives the flavor physics of the standard model will also be responsible for this feature of the MSSM.

Chapter 5

Technicolor

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

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

Such a construction is actually rather intuitive, and even realized in the standard model already. In QCD, bound states of quarks occur which actually have the same quantum numbers as the Higgs, e. g. the σ meson or the $\eta_{c/b}$ mesons. In fact, already within QCD condensates with the quantum numbers of the Higgs condensate can be constructed, which induce the breaking of the electroweak symmetry. Only because the size of such condensates is then also given by the hadronic scale, and thus of order tens to hundreds of MeV, this is not sufficient to provide quantitatively for the observed electroweak symmetry breaking. Qualitatively it is the case.

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

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

5.1 Simple technicolor

5.1.1 General setup

The simplest version of technicolor is indeed just an up-scaled version of QCD, though with a more general gauge group SU(N_T), with N_f additional fermions Q, the techniquarks. The techniquarks are massless at tree-level. They are placed in the fundamental representation of SU(N_T), and there are, in addition, the $N_T^2 - 1$ gauge bosons, called technigluons. Therefore, the total gauge group of the such extended standard model is SU(N_T)×SU(3)×SU(2)×U(1). The techniquarks harbor, similar to the ordinary quarks, a chiral symmetry. In such a theory the elementary particles include, besides the techniparticles, all the fermions and gauge bosons of the standard model, but no Higgs.

Such a theory then looks very much like QCD, though may have a different number of colors. Therefore, its dynamics are ought to be quite similar. In particular, technicolor confines, and techniquarks can only be observed bound in technihadrons. This dynamics will therefore be determined by a typical scale. In QCD, this scale is Λ_{QCD}, which is of order 1 GeV. This number is an independent parameter of the theory, and essentially replaces the coupling constant g_s of the elementary theory by dimensional transmutation. Now, in technicolor therefore there exists also such a scale Λ_T, the technicolor scale. To be of any practical use, it must be of the same size as the electroweak scale, otherwise the hierarchy problem will emerge again, though possibly less severe as a little hierarchy problem. Assume then that this scale is of the size 1 TeV instead of 1 GeV like in QCD. Then the dynamics of the technicolor theory would be the same as that of QCD, though at a much higher energy scale, and possibly with a different number of colors and flavors.

Besides the technimesons, which will play an important role in the electroweak sector as discussed below, there are also technibaryons. If they are fermionic, i. e. if N_T is odd, the lightest one can be stable, similar to the proton, and may thus exist as a remnant particle, and in particular is a dark matter candidate. However, in general these particles would be too strongly interacting, at least by quantum loop effects, than to be undetected by now. Hence, their decay into standard model particle is desirable, requiring a violation of the associated technibaryon number. If the number of technicolors N_T is even, the technibaryons are bosons, and could in that case oscillate by mixing into mesonic states of the standard model, and would therefore decay also´ by such channels.

Another similarity with QCD is even more important. The techniquarks are so far massless. As in QCD, the chiral symmetry of the techniquarks is assumed to be broken spontaneously by the dynamics of the technigluons. The associated condensate will have a size of about Λ_T, and will give the techniquarks an effective constituent mass of the order of Λ_T as well. Thus, technihadrons will have in general masses of multiple times the constituent techniquark mass. The only exception are the arising number of Goldstone bosons, similar to the pions and other pseudoscalar mesons of QCD. How many such Goldstone bosons appear depends on the number of techniflavors. In the present setup, their number will be N_f^2-1. As in QCD, these will be pseudoscalar bound states of a techniquark Q and an anti-techniquark $\bar{Q}$. If the techniquarks have the same weak charges and electromagnetic charges as the ordinary quarks, these technipions will just have the correct quantum numbers such that they can become the longitudinal components of the weak isospin W-bosons[1], instead of the would-be Goldstone bosons of the Higgs mechanism. Mixing with the hypercharge interaction will then lead as usual to the electroweak interactions. The Higgs is actually not one of the Goldstone bosons, but will be a scalar meson, the analogue of the σ meson of QCD. Thus, it is expected to be more massive, but also more unstable than the Goldstone mesons.

Note that for massless techniquarks the Goldstone bosons will be exactly massless. This can give rise to problems, as discussed below. However, it is not possible to give the techniquarks an explicit mass, because they have to be coupled chirally to the weak isospin. Thus, this remains a problem for $N_f > 2$ in such simple technicolor theories, and how to resolve it will be discussed after illustrating other problems of this simplest setup.

The actual quantitative values for the various scales introduced can be estimated if the numbers of QCD are just scaled up naively to Λ_T, and the scaling to the number of technicolors N_T is done using the large-N_T approximation. The basic relation relates the electroweak condensate $v \approx 246$ GeV with the decay constant of the technipion. The later can then be related to the relation of the technicolor scale and the QCD scale with the pion decay constant in QCD f_{QCD}, which is measured to be about 92 MeV,

$$v = \langle \bar{Q}_L Q_R \rangle^{\frac{1}{3}} = f_T = \sqrt{\frac{N_T}{3}} \frac{\Lambda_T}{\Lambda_{\text{QCD}}} f_{\text{QCD}},$$

with the technichiral condensate $\langle \bar{Q}_L Q_R \rangle$. Solving for the technicolor scale yields

$$\Lambda_T \approx \sqrt{\frac{3}{N_T}} \frac{f_T}{f_{\text{QCD}}} \Lambda_{\text{QCD}} \approx \sqrt{\frac{3}{N_T}} 0.7 \text{ TeV}.$$

in the $\overline{\text{MS}}$ scheme with a Λ_{QCD} of about 250 MeV. Due to the breaking of the chiral symmetry, the effective mass of the techniquarks at lower energies is approximately also

[1] Which already mix with the original pions, as pointed out before.

given by

$$m_Q(0) \approx v.$$

Though these are rather small masses, the techniquarks are not observable alone, similar to quarks at low energies. Thus, their direct detection is complicated by bound states, and their respective masses rather sets the scale for observation.

The mass of the Goldstone technipion is about

$$M_{\pi_T} \approx \sqrt{\frac{N_f}{2}}v,$$

and thus in the right region for them to be components of the W and Z bosons, if the number of flavors is not too large. Of course, QCD-like dynamics imply more bound states. Thus, masses of the low lying non-Goldstone bosons would start at about $2v \gtrsim 500$ GeV, plus binding effects. Assuming a QCD-like hierarchy, the next lightest state would be the techniρ, which would have a mass

$$M_{\rho_T} \approx \sqrt{\frac{3}{N_T}}\frac{vM_\rho}{f_{\text{QCD}}} \approx \frac{3.3 \text{ TeV}}{N_T^{\frac{1}{2}}},$$

and therefore would be sufficiently heavy to escape detection so far.

After outlining these general properties of simple technicolor, it is worthwhile to investigate possible realization, and using them to discuss shortcomings of this type of technicolor. This will force one to consider other realizations of the technicolor idea.

5.1.2 Susskind-Weinberg-Technicolor

The simplest (and ruled out[2]) realization of the general setup is given by the Susskind-Weinberg version of technicolor. These theories have as a gauge group SU(N_T)×SU(3) ×SU(2)×U(1). There are $2N_f$ flavors in the fundamental representation of SU(N_T), each flavor being either a member of a left-handed weak isospin doublet or a right-handed weak isospin singlet of techniquarks, in analogy to the fermions of the standard model. Despite their name, the techniquarks are chosen singlets under color. Their weak hypercharge is then determined by requiring to have an anomaly-free theory. This requires that the electric charges of the flavors are $1/2$ and $-1/2$, for the $+1/2$ and $-1/2$ weak isospin charges of the weak isospin doublet, respectively.

For $N_T = 4$ and $N_f = 2$ this gives with the above formulas a techniscale Λ_T of about 600 GeV. Alternatively, by embedding this theory in a minimal GUT, a value of the

[2]Note that in the context of extended technicolor such theories for $N_T = 3$ and $2N_f$ between 6 and 12 become interesting again, as will be discussed in section 5.3.2.1.

electroweak scale of 270 GeV can be obtained. Both numbers are in rather good agreement with the expectations.

The techniquarks will then acquire an effective mass of about 260 GeV, already in disagreement with current observational limits. Furthermore, the techniquarks can form technimesons with about twice this mass, and technibaryons, for $N_T = 4$ containing four techniquarks, and thus of a mass of about 1-2 TeV. However, these technibaryons would be (almost) stable, since in such a theory techniquark number is in the same way conserved as ordinary baryon (or quark) number in the standard model. At first, this may seem like a candidate for dark matter, but since it is potentially both weakly and electromagnetically charged, it cannot fulfill the role of dark matter, and is actually rather a problem for the consistency with cosmological observations. Extended technicolor introduced later will make it again unstable, and therefore remove this burden from technicolor. In fact, once unstable, it will have a spectacular decay pattern, generating heavy quarks abundantly.

Arranging the numbers differently for clarity, take only $N_f = 2$. The chiral symmetry of the techniquarks will then be the exact global chiral group $\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R \times \mathrm{U}(1)_V \times \mathrm{U}(1)_A$. Like in QCD, the techniquark condense and break chiral symmetry, providing the massless techniquarks with mass. This will break the chiral group down to $\mathrm{SU}(2) \times \mathrm{U}(1)_V \times \mathrm{U}(1)_A$. This will be accompanied by massless Goldstone bosons, the technipions. If the broken SU(2) subgroup and the $\mathrm{U}(1)_V$ is actually gauged to become the product group of the electroweak sector of the standard model, the technipions will have the correct charge structure to become the longitudinal components of the $W^\pm$ and the Z bosons. As a consequence, the scattering of $W^\pm$ and Z bosons will become dominated by the strong techniforce at energies above Λ_T, one of the most important signatures for technicolor.

The remaining symmetry part, $\mathrm{U}(1)_A$, is actually anomalous, as in QCD, and therefore is not present on the quantum level. As a consequence, the techniη' will also be anomalously heavy, as the η' of QCD, about 1-2 TeV by upscaling. It is therefore also safe from detection.

More phenomenological interesting are the analogues of the QCD ρ-mesons, the techniρs. Its mass in this case is expected to be about 1-2 TeV, and should be the first new composite particle, which is sufficiently stable and distinct to be detectable as an unambiguous signal for technicolor. The only potentially lighter particle is the technicolor version of the QCD σ-meson. However, its quantum numbers are that of the Higgs, and thus cannot easily be distinguished from a standard model Higgs. Furthermore, if faithfully upscaled from QCD, its widths will be so large that it will be essentially not visible. Since the width of the actually observed Higgs is small, this needs to be avoided.

The situation becomes even more awkward when including a larger number of techniflavors to obtain better agreement with the electroweak scale. As in QCD, the larger chiral symmetry group SU($2N_f$)×SU($2N_f$) will be broken down to SU($2N_f$), thus there are N_f^2-1 Goldstone bosons. Of these $N_f^2/4-1$ turn out to be uncharged under the standard model forces, and thus interact weakly. They therefore also do not acquire any mass, and are called therefore techniaxions. Unfortunately, though with them comes an additional source of CP violation, they are ultimately incompatible with cosmological observations. Even more problematic, the electromagnetically charged technipions not absorbed by the $W^\pm$ are massless up to standard model corrections, which amount to about 6 GeV. Such particles are experimentally ruled out.

5.1.3 Farhi-Susskind-Technicolor

As pointed out, this simplest versions of technicolor have a number of shortcomings. A bit more useful are the more general Farhi-Susskind versions of technicolor. In this a fourth generation is added to the standard model, though having possibly a different electric charge structure within the ranges permitted by anomaly freedom. Furthermore, this additional generation is gauged under the technicolor gauge group SU(N_T).

As a consequence, the associated chiral symmetry group is SU(8)×SU(8)×U(1)$_V$×U(1)$_A$. Since with respect to the techniforce all the technifermions are equal, all will condense, breaking the chiral group down to SU(8)×U(1), with the anomaly-mediated breaking of U(1)$_A$, and including the gauged subgroup SU(2)×U(1) of the electroweak sector. This gives for each of the four flavors, technitop, technibottom, technielectron and technineutrino a chiral condensate and in total 63 Goldstone bosons, and one massive one due to the axial anomaly. The four condensates can then act together to give the electroweak condensate, making a lower mass of about roughly 150 GeV for each possible.

Classifying the Goldstone bosons, there are again four excitations having the quantum numbers of the ordinary Higgs field, and thus three of them provide the longitudinal degrees of freedom of the $W^\pm$ and Z bosons, and one appears similar to the standard model Higgs, though with a potentially large mass due to the axial anomaly. For example, the Goldstone boson giving the Z boson's longitudinal component, the neutral technipion, is given by the combination

$$\bar{t}_T^i t_T^i - \bar{b}_T^i b_T^i + \bar{\nu}_T \nu_T - \bar{e}_T e_T,$$

where the index on the technitop and the technibottom correspond to their QCD color charge.

Besides the one appearing like the standard model Higgs, there are two more neutral electrically ones, and two electrically charged ones. The remaining ones have weak and/or color charges. Some of them are expected to be almost massless, making the model not viable in its current form. Depending on the assignment of the electric charge to the technifermions, these particles can be either stable or decay. However, stable colored technigoldstone bosons would be expected to bind to ordinary nuclear matter, thus setting strong limits on their existence. However, the quantum numbers of these objects are the same as the ones of leptoquarks in GUTs, making a distinction, if found, complicated. Similarly, technivectormesons will have the same charge structure as gluons, but have in general masses of the order of a few hundred GeV. They will therefore appear in radiative corrections of strong processes, and can thus be accessed at such energies in principle, though the QCD background may make this in practice complicated.

However, such simple setups run in general into problems with precision tests of electroweak observables, like the S, T, and U parameters. In particular, such theories permit that the techniquarks would appear in intermediate states. Since the flavors are mixed in the standard model, a further flavor would permit to enhance flavor-changing neutral currents, leading to a much too large splitting of the mass of the short-lived and long-lived kaons, if the technicolor scale would not be too high to provide electroweak symmetry breaking. Another problem is the top mass, which is almost of the same size as the technicolor condensate. Thus, top quarks should be sensitive to the composite structure of the Higgs to an extent which is incompatible with current experimental constraints. Also, if the techniquarks carry a conserved technibaryonic quantum number, this yields problems with cosmological observations.

All of these problems appear predominantly because of the assumption that technicolor is just QCD at a higher scale. Therefore, most attempts to remedy these problems aim at distorting these similarities.

5.2 Extending technicolor

There are several proposals how to deal with the problems introduced by adding a simple technicolor sector to the standard model, without affecting the virtues of such an extension. The most successful so far is extended technicolor, though here also some other proposals will be discussed.

5.2.1 Extended technicolor and standard model fermion masses

There is another reason that the simplest technicolor models are not sufficient. This is that not only electroweak symmetry is broken by the Higgs, but also the fermion masses are generated by Yukawa couplings to the Higgs. To obtain the different masses of the standard-model fermions requires the condensate of the techniquarks to be coupled differently to all the standard model fermions. This is usually done by adding further massive gauge bosons with mass of a second scale $\Lambda_{\mathrm{ETC}} > \Lambda_T$, the extended technicolor scale, and coupled also to the standard model fermions. By this also the flavor group of the standard model becomes gauged in the extended technicolor gauge group, say, $\mathrm{SU}(\geq 6)_F \times \mathrm{SU}(N_T) \times \mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$. In general, this is achieved, very similar to GUTs, by having a large extended technicolor (ETC) master gauge group, which contains all the other gauge groups, including the technicolor gauge group. All fermions, standard model ones and technifermions alike, are then embedded in representations of this master gauge group.

The breaking of the flavor group provides then a mechanism for the generation of the fermion mass, by their coupling to the now heavy flavor gauge bosons. However, despite giving a mechanism how the standard model fermion masses are generated, it is still a problem how to generate their relative sizes without the introduction of either new parameters or new fields. Note that this also explicitly breaks the chiral symmetry of QCD by fermion masses, as in the standard model.

Another problem is that the resulting effective couplings have to be very specific such that the hierarchy of fermion masses is obtained. E. g., quarks of mass a few GeV require a Λ_{ETC} of 2 TeV, while the top quark would rather require much less.

Nonetheless, such extended technicolor models are an important building block for promising technicolor theories, and therefore will be discussed here.

In general, the setting is to start with the master gauge group of extended technicolor G. It is broken by strong interactions at some scale Λ_{ETC} into the gauge group of one of the technicolor theories described previously. Then, at some lower scale Λ_{TC}, the technicolor interactions become strong, leading to electroweak symmetry breaking as before. To avoid the hierarchy problem, it is often convenient not to make a single step from extended technicolor to technicolor, but have one or more intermediate steps, which in a natural way generate a hierarchy of scales. This is also known as a tumbling gauge theory scenario. The initial driving mechanism of the first breaking is not necessarily specified. A possibility would be that the master gauge group is part of a supersymmetric gauge theory, which provides naturally a hierarchy-protected Higgs mechanism, as an initial starting point.

A variation on this theme are triggering models. In this case the fact that QCD breaks the electroweak symmetry is used to plant a seed of breaking also an extended technicolor gauge group. This seed is then amplified by a suitable arrangement of interactions such that the right hierarchy of scales emerge. This can also be done with other triggers than QCD.

The extended technicolor gauge bosons, which have become massive on the order of Λ_{ETC} still interact with all particles. In particular, they mediate four-fermion couplings purely between techniparticles, between ordinary standard model particles and techniparticles, and between standard model particles. Similarly to the electroweak interactions, they are perceived at the scale Λ_{TC} and below as effective four-fermion couplings. Since the technifermions condense, the mixed couplings have a contribution which couple the standard-model particles to the technicolor condensate, schematically

$$g^2\bar{T}\gamma_\mu T D^{\mu\nu}_{\mathrm{ETC}}(p^2)\bar{q}\gamma_\nu q \overset{p^2\ll\Lambda_{\mathrm{ETC}}}{\to} \frac{g^2}{M^2_{\mathrm{ETC}}}\bar{T}\gamma_\mu T\bar{q}\gamma^\mu q \overset{p^2\ll\Lambda_T}{\to} \frac{g^2\langle\bar{T}T\rangle}{M^2_{\mathrm{ETC}}}\bar{q}q,$$

with extended technicolor coupling constant g, the mass of the extended technigluon M_{ETC}, and the techni fermion condensate $\langle\bar{T}T\rangle$. Thus the techniquark condensate indeed also generates the masses of the standard model fermions on top of the W and Z boson masses. The size of the quark masses is given approximately by

$$m_q \approx \beta\frac{N_T\Lambda^3_{\mathrm{TC}}}{\Lambda^2_{\mathrm{ETC}}} \tag{5.1}$$

where the factor β depends on the structure of the theory. For $\beta \approx 1$, a 'natural' size, this yields an upper limit on Λ_{ETC} from the masses of the light quarks, to be about not much more than an order of magnitude larger than Λ_{TC}, and this only if Λ_{TC} is not too large itself. On the other hand, if Λ_{ETC} should not be too large, this is an upper bound for the quark masses, which can be produced. In fact, if Λ_{TC} is not much smaller than one to two orders of magnitude than Λ_{ETC}, and N_T is not too large, reproducing the bottom quark mass, and much less the top quark mass, is hardly possible.

An advantage of such an interaction is that, depending on the detailed structure of the interactions, this can provide some of the mixing of the standard model CKM matrices. However, the same holds for the effective four-techniquark coupling, coupling states of two techniquarks also to the techniquark condensate. This gives rise to larger masses for the technigoldstone bosons, and in particular of techniaxions. Unfortunately, the size of this effect is essentially given by the ratio of the extended technicolor scale and the technicolor scale. If both scales are not very far apart, the effect is unfortunately not large enough to give those particles too light in technicolor theories a sufficiently large mass to be compatible with experimental bounds. Increasing the extended technicolor scale is not a solution, since this spoils the masses of the light standard model fermions. A solution to this will be the walking technicolor theories discussed in section 5.3.

A further downside is that also a coupling between four standard model fermions is induced, which contributes to, e. g., flavor-changing neutral currents. As a result, e. g., the mass difference between the two neutral kaon states K_S^0 and K_L^0, δm_K^2, is modified by ETC contributions to

$$\frac{\delta m_K^2}{m_K^2} \to \frac{\delta m_K^2}{m_K^2} + \gamma \frac{f_K^2 m_K^2}{\Lambda_{\mathrm{ETC}}^2}.$$

Herein is γ the effective coupling between standard model quarks. If assumed to be of the same size as the Cabibbo angle, which mediates the mixing in the standard model and is of order 10^{-2}, Λ_{ETC} has to be of order 10^3 TeV for this to be compatible with experiment. This substantially exceeds the expected size. It is one of the persisting challenges of extended technicolor theories to provide at the same time the mass of the top quark without having such currents to be so large that they are in conflict with experiments. Actually, this problem also affects other beyond-the-standard-model theories, most notably supersymmetry. It thus makes evident that one of the greatest challenges is to understand the flavor structure of the standard model.

A further problem is that such an interaction yields corrections to quantities like the coupling of the Z boson: If the Z boson is first converted into a techniquark pair, and then these convert via extended technigluon exchange into ordinary standard model gauge bosons, this will yield vertex corrections. These are essentially given by ratios of the technicolor scale, giving the coupling to the Z boson of the techniquarks, and the extended technicolor scale, relevant for the conversion ratio of techniquarks to standard model fermions. Since this ratio is not too large, the corrections are significant, and indeed ruled out.

Thus, it is a challenge to construct extended-technicolor models which are consistent with observations.

5.2.2 Techni-GIM

Techni-GIM models try to solve the problem of flavor-changing neutral currents by imitating the Glashow-Illiopoulos-Maiani (GIM) mechanism of the standard model. This mechanism has been proposed to explain why no strangeness-changing neutral currents have been observed. Such currents would exist if there would be only three quarks, up, down, and strange, with the strange quark being a singlet under the weak interactions. The GIM mechanism shows that if there is a fourth quark, the charm quark, promoting the strange quark to being its weak isospin-doublet partner, interference effects will remove such currents. This also requires that the mixing of down and strange quarks is only due to the Cabibbo angle. Essentially that boils down to the fact that diagrams where initial state fermion lines and final state fermion lines are connected vanish (up to corrections proportional to the mass splittings) due to the mixing, and the only possibility is by an intermediate state with two weak gauge bosons. That is, e. g., the reason why the decay of K^+ to π^+ is suppressed compared to the decay into π^0.

Techni-GIM models capitalize on this idea by adjusting the particle content such that at tree-level no flavor-changing currents can occur. Radiative corrections can then be arranged such that they are not in conflict with experimental observations.

This is achieved by introducing instead of one common extended technicolor gauge group three, one for each weak multiplet. I. e., there is one extended technicolor gauge group coupled to the three generations of left-handed doublets, and one each for the three generations of two pairs of right-handed singlets. Thus, flavor-changing neutral currents are avoided, since they couple left-handed fermions and right-handed fermions differently. The price to be paid is a proliferation of gauge-groups. Furthermore, since the gauge groups are the same for both quarks and leptons, the gauge bosons act as leptoquarks. However, the effects can be adjusted such that, e. g., proton decay rates are not in violation of experimental bounds. Unfortunately, the relation (5.1) still holds, indicating that it is again a serious problem to obtain heavy quarks.

5.2.3 Non-commuting extended technicolor

Non-commuting extended technicolor is the first model to play with a recurring idea to solve the challenges imposed by the flavor structure of the standard model: To treat the third generation of standard model fermions differently. In this case the non-commuting implies that the third generation is actually charged under the extended technicolor gauge group but not under the ordinary weak isospin gauge group. By breaking the extended technicolor gauge group first down to a SU(2) group for the third generation, a sequence of breakings is generated which finally ends up with the appropriate structure for the standard model supplemented by some technicolor interaction to break the electroweak symmetry by the formation of a chiral condensate.

The sequence for the extended technicolor gauge group G is then

$$\begin{aligned}
& \mathrm{SU}(3)_c \times \mathrm{SU}(N_T) \times G \times \mathrm{SU}(2)_{1+2} \times \mathrm{U}(1) \\
\stackrel{f}{\to} \; & \mathrm{SU}(3)_c \times \mathrm{SU}(N_T) \times \mathrm{SU}(2)_3 \times \mathrm{SU}(2)_{1+2} \times \mathrm{U}(1)_Y \\
\stackrel{u}{\to} \; & \mathrm{SU}(3)_c \times \mathrm{SU}(N_T) \times \mathrm{SU}(2)_{1+2+3} \times \mathrm{U}(1)_Y \\
\stackrel{v}{\to} \; & \mathrm{SU}(3)_c \times \mathrm{SU}(N_T) \times \mathrm{U}(1)_{em}
\end{aligned}$$

where f, u, and v denote the condensates which hide the corresponding symmetries. The indices on the SU(2) groups denote which generations are charged under the corresponding gauge group. It then depends on the quantum numbers of these condensates how much of them enters in the hiding of the various groups, and therefore to which extent the mass generation of the individual standard model fermions is dominated by which interaction. It could either be that the third generation now indeed obtains the bulk of its mass from the electroweak symmetry breaking effect, but it is also possible to arrange it that this contribution is minor.

Irrespective of the details, in the end such a structure can be arranged such that the standard model fermion masses come out with roughly the right size. It is even possible to accommodate the two orders difference of magnitude of the τ and the top, despite that they have to be both charged under the ETC gauge group to provide an anomaly-free theory.

A distinct prediction of this theory is that by breaking $\mathrm{SU}(2)_3\times\mathrm{SU}(2)_{1+2}$ to $\mathrm{SU}(2)_{1+2+3}$ the three gauge bosons associated with the broken gauge group become massive with masses of the order of u, just above the electroweak scale v. These W' and Z' gauge bosons, since originally mediating a weak-like force between the third generation members, should have similar properties than the electroweak W and Z bosons. This gives quite unique signatures to be searched for, in particular in the form of effective four-point couplings of standard model fermions in weak channels. The lower mass limits for them are currently above 500 GeV, giving constraints on u. However, some related models like top-flavor models, and other theories having a further weakly interacting gauge group at the TeV scale, can also provide such heavy copies of the W and Z bosons. Indeed, nowadays generically new neutral vector bosons are denoted by Z', unless qualitatively very different for the Z in some particular model. In fact, even the techniρ would appear like such a Z'.

An extension of this idea (tumbling technicolor) plays with the possibility of a sequence of breaking theories, and each of the corresponding condensates is associated with one or more of the fermion masses, generating their hierarchy naturally.

5.3 Walking technicolor

5.3.1 Generic properties

The basic reason why a similarity of technicolor to QCD is problematic is that in QCD almost all non-trivial dynamics is concentrated in a narrow window around Λ_{QCD}. That is because the running coupling of QCD changes rather quickly from strong to weak over a very narrow range of energies. In the electroweak sector, however, the dynamics is spread out over a much larger range of energy scales in relation to its fundamental scale v, since both the masses of the fermions and electroweak symmetry breaking must occur. Thus, a viable realization of the technicolor idea of strong dynamics paired with electroweak phenomenology must reflect the slow evolution of the electroweak physics. This is the aim of walking technicolor by replacing the fast running QCD evolution with a much slower, walking, behavior.

As a consequence, such a theory has more intrinsic scales than QCD. QCD is essentially only characterized by the one scale when it becomes strong. A walking theory can have up to three scales. Assuming the walking theory to be also asymptotically free, there exists a scale where it changes from being a theory acting strongly enough to break electroweak symmetry to an almost free theory. A second scale must occur at low energy when it stops walking, and the third scale is the one where it becomes sufficiently strong to confine techniquarks. Of course, the latter two may coincide, but the first two may not, or the theory would no longer be walking anymore.

To implement such an idea, it is required that the running coupling becomes weak much slower. Since the coupling is given implicitly by the β function by

$$t = \int_g^{g(t)} \frac{dg'}{\beta(g')},$$

where $g = g(\Lambda_T)$ is some chosen initial conditions, and $t = \ln(\mu/\Lambda_T)$. The β function to three-loop order in a QCD-like setup is given by

$$\begin{aligned}
\beta(g) &= -\beta_0 \frac{g^3}{16\pi^2} - \beta_1 \frac{g^5}{(16\pi^2)^2} - \beta_2 \frac{g^7}{(16\pi^2)^3} + \mathcal{O}(g^9) \\
\beta_0 &= \frac{11}{3} C_A - \frac{4}{3} N_f T_R && (5.2) \\
\beta_1 &= \frac{34}{3} C_A^2 - \frac{20}{3} C_A T_R N_f - 4 C_R T_R N_f && (5.3) \\
\beta_2 &= \frac{2857}{54} C_A^3 - \frac{1415}{27} C_A^2 N_f T_R + \frac{158}{27} C_A (N_f T_R)^2 \\
&\quad - \frac{205}{9} C_A C_F N_f T_R + \frac{44}{9} C_F (N_f T_R)^2 + 2(C_F)^2 N_f T_R. && (5.4)
\end{aligned}$$

Here, N_f is the number of flavors in the fundamental representation, and T_R is the Dynkin index of the group. If the function β is very close to zero for some value of g, $g(t)$ becomes a very slowly varying function of t when it reaches this value.

E. g., to leading order of the β function of technicolor with techniquarks in the fundamental representation of an SU(N_T) techni gauge group with gauge coupling g_T this yields

$$\beta_T = -\frac{g_T^3}{16\pi^2} \left(\frac{11}{3} N_T - \frac{8}{3} N_f \right) + \mathcal{O}(g^5)$$

requiring about $N_f \approx 11 N_T / 8$. Note that the standard model charges of these techniquarks are not relevant at this order. Therefore, by judiciously choosing the gauge group and the number of flavors, it is possible to construct a β function giving a theory that has the desired walking behavior. More possibilities are offered by exchanging the representation of the techniquarks. Using instead of fundamental techniquarks adjoint techniquarks shows that for $N_T = 2$ the β function with $N_f = 2$ already vanishes to two-loop order. The existence of a zero of the β-function at two loop is known as the Banks-Zak fix-point.

It should be noted that this argumentation can only be superficial: The β function is dependent on the renormalization scheme, and the running coupling can be defined in many ways. It is therefore a much more subtle task to indeed show that a theory is walking than the outline discussed here. However, the general gist of finding a theory with, more or less, constant interaction strength over a large momentum range remains.

Staying for a moment with the assumption that the coupling and β-function give the correct picture, the big advantage is that the coupling evolves slowly with energies. Therefore, the theory stays strong over a wide range of energies. As a consequence, techni bound states can no longer spoil various electroweak precision measurements. Furthermore, when arranging this walking behavior for the range between Λ_{TC} and Λ_{ETC}, the interaction of the standard model fermions among each other and mediated by the extended technigluons will be essentially independent of energy, and thus remain small, while the electroweak dynamics only given by the technicolor dynamics is staying essentially unaltered. In fact, what happens is that (5.1) is modified to

$$m_q \approx \gamma \frac{N_T \Lambda_{\mathrm{TC}}^2}{\Lambda_{\mathrm{ETC}}} \tag{5.5}$$

and thus quark masses of order one to two GeV are possible for reasonably chosen values of Λ_{TC} and Λ_{ETC} between one and a few tens of TeV.

A similar replacement also takes place for the masses of the Goldstone boson masses which are not absorbed by the W and Z bosons. Their mass is now found to be of order $N_T \Lambda_{\mathrm{TC}}$, and thus sufficiently large to be not detectable yet.

A variation on the idea of walking technicolor is given by low-scale technicolor. In this case the techniquarks necessary to make the theory walk are in different representations of the gauge group. Since their respective energy scales are thus different, the corresponding condensates, which add up quadratically to form the electroweak condensate, form at different energies. These scales are widely separated due to the walking behavior. As a consequence, techni bound states could have masses of the same size as the top quark, though being sufficiently weakly coupled to escape detection so far. However, even with the relation (5.5) this is only marginally sufficient to obtain the bottom quark mass, and for the top quark an excessively fine-tuned value of γ would be required.

5.3.2 Realization of walking technicolor theories

Though there are thus still significant problems in realizing a phenomenologically fully consistent extended walking technicolor theory, it is quite likely that some walking behavior is an important part for many proposals of strong interactions beyond the standard model. Thus the classification of such theories, and the construction of viable models with them, has become an important goal in itself.

5.3.2.1 The conformal window

To identify viable technicolor sectors, it is important to understand the generic properties of gauge theories with a simple Lie algebra and a number of flavors in one or more representations. There are four different types of behaviors, which are expected to occur, and so far have been the only ones encountered.

If there are no fermions coupled to the theory, the resulting Yang-Mills theory shows for any Lie algebra the same qualitative behavior of a running theory, with a fast transition between weak interactions and strong interactions towards small energies.

When fermions are present, the following type of behaviors can emerge, depending on the number of massless flavors. However, for some Lie algebras some of the cases may merge, if the behavior evolves too quickly with the number of flavors. Still, if formally a fractional number of flavors is admitted, the following set of possibilities seems to be common to all gauge groups.

For a small number of flavors, all these theories remain running, and chiral symmetry breaks spontaneously. These theories behave essentially like QCD. When adding more flavors, the theories slow down, and become gradually more and more walking. At a critical number of flavors, even the walking stops altogether, and the theories become conformal, i. e., scaleless without chiral symmetry breaking and without any observable dynamics. This behavior persist for a range of flavors, and this range is also called the conformal window[3]. Finally, above a second critical number of flavors, the theories lose asymptotic freedom, and thus become more strongly coupled the larger the energies. For massive flavors, the theories follow a similar pattern, but such a theory can never be conformal, and walking will only be possible in a range where the energies are large compared to the fermion masses, as the walking behavior is similar to a conformal behavior.

What the precise number of flavors for a given gauge algebra and representation is, is a highly non-trivial question of current research. For some cases, rather good results

[3]It is not yet entirely clear, if this conformality is just a behavior persisting from the infrared up to a (large) scale, or for all energies. This would require an exact solution for the β function to test whether it is really or only almost constant.

have been obtained. E. g., for SU(3), the theory is QCD-like up to about 8-9 fundamental flavors, is walking up to 10-12 flavors, stays conformal up to 16 flavors, and looses asymptotic freedom for 17 or more flavors. For SU(2) with adjoint fermions, the theory is possible QCD-like or walking for one flavor, conformal for two flavors, and loses asymptotic freedom for three or more flavors.

For the purpose of extended technicolor theories, the technicolor sector can be chosen both as a walking theory or as a conformal theory. In case of a conformal theory, the coupling to the standard model with its intrinsic mass scale, like the one induced by QCD, will break the conformality, and make the theory walking.

However, having the right qualitative properties is not guaranteeing that the theory also exhibits the right quantitative properties. It is therefore, in principle, necessary to check for each theory whether its quantitative features are phenomenologically viable.

5.3.2.2 An example: Minimal walking technicolor

To give an example, one of the recently studied technicolor theories will be introduced here. This will be the so-called minimal walking technicolor. The name originates from the fact that the theory is tuned such as to yield minimal disagreement for the S, T, and U parameters.

The theory itself consists, besides the standard model, of an SU(2) technicolor sector with two flavors in the adjoint representation. Thus, the technicolor sector alone is a conformal theory, but this conformality is broken by the standard model. To avoid an anomaly, it is also necessary to couple to the theory a fourth generation of standard model leptons, but no fourth generation of quarks. The additional leptons and the techniquarks do not necessarily have, again for anomaly reasons, the expected charges for such particles with respect to the weak and the electromagnetic charges, and all are uncharged under color.

The detailed charges for the new particles are actually not uniquely fixed, but can be parametrized by a single parameter. E. g., a possible assignment for the hypercharge for the techniquarks is 1/2 for the left-handed techniquarks and 1 and 0 for the right-handed up-type and down-type techniquarks, yielding an electric charge of ± 1 for the techniquarks. The right-handed electron has the charge -2 and the right-handed neutrino the charge -1, while the left-handed ones have the charges of -2 and -1. Thus, these particles can have quite different signatures as the standard model particles. A more standard-model selection would be giving the neutrinos a charge of zero, yielding for the new leptons a charge of -1, as usual. The quarks would then have the conventional charges as well.

Such a theory has an interesting set of bound states. Combining a techniquark and a techniantiquark yields technimesons. There are three technipions, which will take the role of the longitudinal modes of the Ws and the Z. The technisigma will then act like a Higgs particle. Moreover, technibaryons in such a theory are also bosons. Because of a peculiarity of the group structure of SU(2), the fact that SU(2) is pseudo-real, technibaryons and technimesons can mix. This leads to the interesting possibility that a longitudinally polarized W or Z can oscillate into a technibaryon. In addition to these bosonic bound states, there are also fermionic ones, which consists out of a techniquark and a technigluon.

Thus, there will be a plethora of bound states at the TeV scale in such a theory. Nonetheless, at current energies there will be little observable of this theory by construction, at least to leading order in perturbation theory and in chiral effective models, which will not be discussed in detail here. Thus, such a theory is currently still in agreement with the standard model. However the additional neutrino must be very heavy, compared to the other neutrinos, at least above the Z mass. Also, for the additional lepton the lower mass bound is quite high, of the order of a few hundred GeV.

In the current setup of this theory, extended technicolor is not explicitly incorporated. Rather, a number of four-fermion terms appear with couplings adjusted to reproduce the standard model phenomenology. In this sense, minimal walking technicolor is currently an effective theory.

5.4 Topcolor-assisted technicolor

To also cope with the top quark, another proposal for a higgsless standard model, which alone fails, can be incorporated into the technicolor setup. This is the so-called topcolor approach.

Originally, to circumvent some of the problems appearing with the plethora of additional particles introduced by models, one approach, called topcolor, was to let instead a top quark condensate take the role of the Higgs. To provide such a mechanism, there is instead of $\mathrm{SU}(3)_c$ a double group $\mathrm{SU}(3)_{1+2}\times\mathrm{SU}(3)_3$. Only the top-quark is a triplet under the second gauge group, while all the other quarks are triplets under the first group. If this product group is broken at some scale to $\mathrm{SU}(3)_c$ there will be the ordinary gluons and in addition 8 massive topgluons. If the relative size of the couplings are chosen with hindsight, the massless $\mathrm{SU}(3)_c$ gluons will be predominantly from the group $\mathrm{SU}(3)_{1+2}$, thus not altering the strong interactions significantly for the light quarks, while the ones connected to the top quark are mostly the massive topgluons.

The interaction with the topgluons then induces only for the top-quark an effective four-top coupling involving the topgluon mass M_{topgluon}

$$-\frac{g_t^2}{M_{\text{topgluon}}^2}(\bar{t}_R t_L)(\bar{t}_L t_R).$$

Such couplings have been studied in various effective models, and it has been found that they induce rather generically bound states of t and $\bar{t}$, if the coupling g_t is sufficiently large at small energies. In particular, it also generically leads to a condensate of the same type, $\langle \bar{t}t \rangle = v_t$, and thus an effective Higgs as a bound state of tops with its condensate effectively obtained from a top condensate. The required size of the coupling is approximately

$$g_t^2 = \frac{g_3^2 \Lambda_t^2}{M_{\text{topgluon}}^2} > \frac{8\pi^2}{N_c},$$

where g_3 is the topcolor gauge coupling, and Λ_t is the scale associated with the breaking of SU(3)$_{1+2}\times$SU(3)$_2$. Since the bound-state has the quantum numbers of the Higgs, it can also be coupled by a Yukawa coupling to the top, therefore implying that it can also generate the mass of the top quark itself by the condensation requiring $m_t = g_t v_t$. However, in pure topcolor theories v_t has to be either too small to make up the entire electroweak condensate, or g_t is too small to induce symmetry breaking, or the top quark mass is too large. Therefore, top quark condensation can only be an additional mechanism.

An interesting opportunity appears when topcolor is used to supplement technicolor, leading to so-called top-color assisted technicolor theories. This has the advantage that the main technicolor sector is not needing so strong interactions that violations of experimental bounds become inevitable when attempting to cover also the top quark. At the same time, the combination of topcolor and technicolor condensates is sufficient to produce the large top quark mass. A drawback is that the bottom quark has to be also charged under topcolor, being the weak isospin partner of the top quark. Thus, its mass would also receive the same large contributions, in disagreement with experiments. A possibility to remove this is by also doubling the weak hypercharge group for the third and the other generations. Since top and bottom have different weak hypercharges, it is possible to rearrange the interactions such that the bottom mass is small compared to the top mass, if the additional weak hypercharge interaction is sufficiently strong. This is called tilting the vacuum, though it mainly distorts the condensate structure.

However, even models constructed in this way have the problem that the topgluons (or toppions) are usually too light to get everything else right, and therefore spoil consistency with the standard model results. To ameliorate this problem, but without introducing yet another gauge interaction, a possibility is to introduce another quark χ in the topcolor sector, which has left-handed components charged under SU(3)$_3$ and its right-handed contribution charged under $\mathrm{SU}(3)_{1+2}$. A gauge-invariant mass matrix for the top and χ quark can then be written as

$$(\bar{t}_L\bar{\chi}_L)\begin{pmatrix} 0 & g_t v \\ M_{t\chi} & M_{\chi\chi}\end{pmatrix}\begin{pmatrix} t_R \\ \chi_R\end{pmatrix} + \text{h.c.},$$

where $M_{t\chi}$ and $M_{\chi\chi}$ are free parameters. The obtained masses for the mass eigenstates are thus

$$m_i^2 = \frac{1}{2}\left(M_{t\chi}^2 + M_{\chi\chi}^2 + (g_t v)^2 \pm \sqrt{\left(M_{t\chi}^2 + M_{\chi\chi}^2 + (g_t v)^2\right)^2 - 4g_t^2 v^2 M_{t\chi}^2}\right) \tag{5.6}$$

Chosen appropriately, the lighter of the two eigenstates acquires the mass of the top quark, while the other is much more heavier, and can easily have a mass in the TeV range. Expanding the masses in this case gives

$$\begin{aligned} m_1^2 &= \frac{g_t^2 v^2 M_{t\chi}^2}{M_{t\chi}^2 + M_{\chi\chi}^2 + (g_t v)^2} + \mathcal{O}\left(\frac{g_t^4 v^4 M_{t\chi}^4}{M_{\chi\chi}^6}\right) \\ m_2^2 &= M_{t\chi}^2 + M_{\chi\chi}^2 + (g_t v)^2 + \mathcal{O}\left(\frac{g_t^2 v^2 M_{t\chi}^2}{M_{\chi\chi}^4}\right) \end{aligned}$$

For sufficiently large M_χ, the one state is much lighter, and a suitable top quark mass can be obtained. In fact, electroweak precision measurements favor a mass of about 4 TeV for M_χ. The corresponding Higgs particle contains now also contributions not only from technicolor and top quarks, but also from the χ quark, making it generically heavier, about 1 TeV. This requires tuning to make it again light enough to be compatible with experiments. Furthermore, this model can be be extended such that technicolor can be removed, and only a combination of a top and a χ-condensate account for electroweak symmetry breaking. In fact, the electroweak condensate is now proportional to the inverse of the sine of the mixing angle squared of the top and the χ quark, permitting an hierarchy of scales in agreement with experiment.

This mechanism of introducing a second partner state such that by mixing a heavy and a light particle emerge, is called see-saw. It is often used to provide a doublet with very different masses by appropriately mixing two similar particles.

The strong interactions among top quarks produces also further bound states, in particular relatively light toppions and topρs, which will couple strongly to the bound state which acts as the Higgs. Hence, these will show up strongly at the scale Λ_t in weak gauge boson processes.

Such models, of course, leave open the mechanism how to break $\mathrm{SU}(3)_{1+2}\times\mathrm{SU}(3)_3$ to $\mathrm{SU}(3)_c$ in the first place. Also, the scale Λ_t has to be of order a few TeV, and can therefore introduce a little hierarchy problem. However, this is usually attributed to other mechanisms at a higher scale. Another feature is that generically also a partner fermion to the bottom quark is necessary, also to cancel any anomalies. It then requires some further construction and adequate choice of parameters to prevent the bottom quark to acquire a mass comparable to the top quark, but it is possible to do so. As in case of the χ, also this partner will be a weak isospin singlet. The production of these particles would also be one of the prime signals for topcolor theories, in particular of the partner of the bottom quark which is generically lighter than the one of the top quark.

5.5 Partial compositness

Note that in general the introduction of additional elementary scalars, which can have varying technicolor and standard model charges, can remove many or even most of the problems technicolor theories have. The advantage compared to the Higgs boson of the standard model, for which to remove technicolor was invented in the first place, is that these Higgses can have a rather large mass, as they only contribute partly to the electroweak effects, the rest coming from technicolor. Therefore, they can be embedded in a higher-scale theory, like a supersymmetric one, where their masses become protected by additional symmetries from a hierarchy problem. This appears to be a valid alternative in case neither supersymmetric particles nor any other light new particles are found, but strong interactions in weak gauge boson scattering indicate a strongly interacting theory as the origin of electroweak symmetry breaking. In particular, even if these scalars do not condense, they can mediate additional interactions between the technicondensates, removing several of the large effects incompatible with the experimental observations in technicolor models, including the top mass. Such modification of technicolor theories usually go under the name of partial compositness. In case also the scalars interact strongly such theories are also known as Abbott-Farhi models.

5.6 Dualities

An important concept is dualities. This is the statement that two different theories are actually showing the same physics, if the involved quantities, fields, symmetries, and coupling constants, are reinterpreted. In particular, in the limit of a large gauge-group with at the same time limited matter content it is found that the perturbation series of different gauge theories coincide. This suggests that in this limit the theories could be identical. Since the proof is perturbative, this has the status of a conjecture.

This is particularly interesting as often corrections turn out to be small or negligible when going to smaller gauge groups. Since such theories are often also related by exchanging a weak coupling for a strong coupling, this implies that theories of different complexity can be used to describe each other.

This happens especially for strongly constrained theories, e. g. conformal theories. The best example is the AdS/CFT correspondence, which links a classical supergravity theory in a high-dimensional space to $\mathcal{N} = 4$ super Yang-Mills theory in the limit of an infinitely large gauge group at infinitely strong coupling. The later is a non-dynamical theory, as superconformality forbids non-trivial scattering. Thus, this is not yet useful. It is, however, conjectured that deforming the theory to become interesting could keep the duality still. As this links a quantum theory to a (comparatively) simpler classical theory, this is very useful. It is, however, not yet clear if this duality can be stretched to relevant theories.

In the same direction exist dualities between different theories in the conformal window of section 5.3.2.1. They are suspected to exist between theories at the upper edge and lower edge of the conformal window, which relates again strong-coupling theories and weak-coupling theories. This is not well established, but would be helpful, as the stronger interacting ones at the lower edge are more interesting for phenomenology.

Chapter 6

Other extensions of the standard model

In the following briefly some other possibilities to add particles to the standard model are presented, which still adhere to a four-dimensional space time and ordinary quantum field theories without gravity. In contrast to the theories discussed in the later chapters, these require less drastic changes at the electroweak or 1 TeV scale to our current picture of nature.

6.1 nHDM models

There is a generic trait for many BSM scenarios: The appearance of additional scalar particles, being them elementary or composite. All of these models have a very similar low-energy behavior, essentially the standard model with more Higgs-like particles. These can be either in the same representation as the standard-model Higgs, or also in a different one. This whole class of models is hence known as n-Higgs models. Particularly important are models which have copies of the standard model Higgs. These models are called n-Higgs doublet models (nHDM). Particularly important is the case of $n = 2$, so-called 2HDM. Of course, nH(D)M models can also be stand-alone models. The extended Higgs sector can have an enlarged custodial symmetry, which can be partly intact. This allows for further conserved quantum numbers.

A generic feature of 2HDM is that it is usually possible to have only one of the Higgs particles condense. This is achieved by a suitable choice of basis in the custodial space and gauge space, and this basis is called the Higgs basis. However, in this basis usually the Higgs particles are not eigenstates of the mass operator, and tree-level mixing is possible. To avoid this requires a different basis, the so-called mass basis, in which the vacuum expectation value will be distributed over multiple Higgs doublets/multiplets.

The other Higgs particles then form an additional quadruplet, of which one behaves like a heavier copy of the standard-model Higgs, two are electrically charged, and one is a pseudoscalar. The additional Higgs particles can partly even be lighter than the standard-model Higgs, and they can play the roles of axions in some cases. They can also be arranged to take part in the see-saw mechanism of section 5.4.

The situation quickly escalates if adding further Higgs doublets without having stringent symmetry conditions. Especially, already the 2HDM has five instead of two independent parameters in the Higgs sector, showing the strong growth of the parameter-space dimensionality with more Higgs particles. Special care has also to be taken that these do not accidentally break other symmetries, especially the electromagnetic gauge symmetry.

Depending on the details of the models, supersymmetric models, technicolor models, and the not yet discussed extra-dimensional models of chapter 8 can have nH(D)Ms as low-energy effective theory, as well as many others. In such cases the extended custodial symmetries and parameters are usually constrained compared to stand-alone nHDMs. Thus, nHDMs also play an important role in constructing non-minimal effective theories as a next step beyond the leading low-energy effective theories of section 3.6.6.

6.2 Little Higgs

The idea of the Higgs as an emergent state is also the primary guideline for the construction of little Higgs models. If the Higgs would be the Nambu-Goldstone boson of a broken global symmetry, it would naturally be light, in fact even massless if the symmetry-breaking would be only spontaneous, similar to the pion in QCD. The simplest case would be an additional global symmetry with some particles charged under it, which becomes broken at the TeV scale.

However, such a simple model is usually inappropriate, and more refined approaches are necessary. One of them is the idea of collective symmetry breaking. To become more formal, such physics is usually described using a non-linear σ-model

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi + \frac{1}{2}\frac{(\phi\partial_\mu\phi)(\phi\partial^\mu\phi)}{|f^2-\phi^2|}. \tag{6.1}$$

If f is zero, this reduces to a free scalar theory. This Lagrangian can be linearized to the linear sigma model by the introduction of another field σ to

$$\mathcal{L} = \frac{1}{2}\partial_\mu\Phi\partial^\mu\Phi - \frac{\lambda}{4}\left(\Phi\Phi - \frac{\mu^2}{\lambda}\right)^2. \tag{6.2}$$

where $\Phi = (\phi, \sigma)$. If the symmetry is broken, $f^2 = \phi^2 + \sigma^2$ and f is a function of the parameters λ and μ. The field ϕ is massless, and plays the role of the Goldstone boson, here the Higgs boson. The original Lagrangian (6.2) is invariant under a symmetry group G acting on Φ, while the effective Lagrangian is only so under a smaller symmetry group G/H acting on ϕ. Thus, to specify the non-linear sigma model, a strength f and a symmetry breaking pattern $G \to G/H$ is necessary. At an order $4\pi f \sim \Lambda$, the effective description in terms of (6.2) breaks down, as then the energy is sufficiently large to excite σs.

To achieve decent agreement with experiment is challenging with this concept. It is necessary to take a group G with a gauged subgroup SU(2)×U(1) to obtain a Higgs with correct properties. Also, there must be explicit breaking, which can be modeled by a mass-term $m^2\phi^2$ in (6.1). Though this approach gives a first possibility, it turns out that it endows a (little) hierarchy problem, since the emerging Higgs is again having a mass sensitive to corrections at the scale Λ.

As a remedy, the mentioned collective symmetry breaking was introduced. The basic idea is to use a product group $G_1 \times G_2$ which has a gauged electroweak group SU(2)×U(1) in each of the factor groups G_i. In such a setup, radiative mass corrections between both factor groups actually cancel, at least at (one-)loop level, such that the Higgs mass is protected. However, to obtain reasonable masses for the top quark and the Higgs simultaneously requires an additional vectorial partner of the top quark, usually denoted by T. Then the top quark can have a large mass, without its (large) Yukawa coupling to the Higgs leading to large radiative corrections of its mass, since the latter are canceled by contributions from its T partner.

Such subtle cancellations are a hallmark of the various little Higgs models. Popular examples for the choice of G are the minimal moose model $(\mathrm{SU}(3)_L\times\mathrm{SU}(3)_R/\mathrm{SU}(3)_V)^4$, where even a SU(3)×SU(2)×U(1) subgroup is gauged, leading to additional gauge bosons which become heavy by the symmetry breaking, the littlest Higgs model with SU(5)/SO(5) having a gauged $(\mathrm{SU}(2)\times\mathrm{U}(1))^2$ subgroup, and the simplest little Higgs model with group G being $(\mathrm{SU}(3)\times\mathrm{U}(1)/\mathrm{SU}(2))^2$ with gauged subgroup SU(3)×U(1).

However, in all cases some fine-tuning appears at some point to obtain results in agreement with experimental data. A radical approach to remedy the problem is by introducing an additional global Z_2 symmetry, under which standard model and additional particles are differently charged. The action of this symmetry is to exchange the subgroups G_i of G. This is called T parity. Provided T-parity is not developing an anomaly, what indeed happens for some models, the lightest additional particle is stable, and thus a dark matter candidate.

Note that the additional symmetry can also be part of a strongly-interacting theory, akin to the technicolor theories of chapter 5. In this context such theories are usually called compositness models, as the Higgs then becomes a composite Goldstone boson of the strongly interacting theory, rather than being an elementary scalar as in proper little Higgs theories.

6.3 Hidden sectors

The generic idea of hidden sectors is that in addition to the standard model there is a second set of particles which have very weak or no coupling to the standard model particles, and a set of very heavy messenger particles connecting this hidden sector to the standard model. Provided that these particles are not gravitationally bound in significant numbers to ordinary astrophysical objects, such a sector will not be detectable unless the energies reached become of order of the messenger masses.

A simple example for a hidden sector would be a hidden QCD with some gauge group[1] SU(N) and hidden quarks charged under this symmetry. The mediator is a U(1), i. e. a QED-like symmetry with a gauge boson Z'. This symmetry is broken at the TeV scale, making the Z' very heavy. If the hidden quarks also have mass of this size, but the hidden QCD is unbroken, a high-energetic Z' can be produced by standard model particles, and then decay into a hidden hadron, which decays to the lightest state, generically a hidden pion, which can then decay through a virtual Z' to standard model particles. Though this scenario is not solving any of the problems of the standard model, it is neither in contradiction to any observation, and has therefore to be taken into account when developing possible search strategies at experiments.

Another possibility is the quirk scenario. In this case the hidden quarks, called quirks, are in addition also charged like standard model quarks, but very heavy compared to the intrinsic scale Λ_{hidden} of the hidden gauge theory. Then the quirks themselves act as mediators. As a consequence, hidden glueballs would be quasi-stable on collider time-scales, giving unique missing energy signatures.

Another possibility is if the hidden theory is almost conformal, and only coupled weakly to the standard model. In this case the conformal behavior of the hidden particles will generate very distinctive signatures, as their kinematic behavior is quite distinct from anything the standard model offers on these energy scales. Such hidden (quasi-)conformal particles are also called unparticles.

Finally, mirror world scenarios have essentially a copy of the standard model as hidden

[1] N has to be larger than two for compatibility with experiment.

particles. Thus, this allows the existence of mirror worlds, which do, however, only interact weakly through the messenger and gravitationally with the usual standrad model. This allows to have whole mirror galaxies gravitationally bound to ours as candidates for dark matter. Variations on this idea keep the interaction strengths, but vary the masses of the particles.

Most of such hidden sectors, or sometimes also called hidden valley theories, have very specific signatures. These include long-lived particles, which decay to standard model particles on distances of meters to kilometers, and (partly) dark jets. The latter refer to the generation of a jet in stronhly interacting hidden sector theories, which are then transformed (partly) to ordinary strongly-interacing particles, and thus jets, appearing possibly substantially removed from an interaction site. Searching for such scenarios requires expeirment to be sensitive away from an interaction point at colliders.

6.4 Flavons

A serious obstacle in technicolor theories, as well as many other scenarios, had been the generation of the mass spectrum of the fermions. To remedy this problem, a variation of a hidden sector can be introduced.

In this case all Yukawa couplings of the standard model are dropped, i. e., all fermions are exactly massless. Then there exists an additional global symmetry, the flavor symmetry, which has symmetry group $U(3)^6$. Some part of it is broken by QCD due to chiral symmetry breaking, generating most of the mass for the up, down and strange quark, but (almost) nothing for leptons, and not enough for the heavier quarks. To provide it, quarks and leptons are coupled to a further field, called flavon, by a messenger particle. Neither are charged under the standard model gauge interactions. The flavon then condenses, and the messenger couples the condensate back to the standard model fermions, providing their masses. Though this is not explaining the mass hierarchy, this splits the dynamics of the fermion mass generation from electroweak symmetry breaking, which could then, e. g., be provided by a simpler technicolor theory than the top-color assisted extended walking technicolor.

Integrating out the messenger field will generate couplings between the Higgs (or whatever replaces it) and the standard model fermions, which will essentially look like the standard-model Yukawa couplings. The Yukawa couplings will then be proportional to the ratio of the electroweak condensate squared and the mass of the messenger particle. Assuming the lightest particle, neutrinos, to have a coupling to the messengers of order 1 then yields a mass for the messenger of order Λ_{GUT}, and therefore further consequences of these particles will not be harmful to present electroweak precision measurements. This scenario is also known as the Froggat-Nielsen mechanism.

6.5 Higgs portal

Dark matter is generically rather simply realized by a hidden sector of, more or less, arbitrary structure. In its simplest form this sector is only gravitationally coupled to the standard model, making dark matter only observable by its gravitational action. While possible, this is not very attractive.

On the other hand, direct detection places stringent limits on the interaction of dark matter with the standard model. Strong interactions are ruled out, and weak interactions only marginally allowed. Electromagnetic interactions are only possible if the electric charge is very small compared to the other standard-model particles, so-called milli-charged particles. If such a case is undesirable, e. g. because of it being hard to reconcile with a GUT structure, there is only one possibility left. This is the Higgs.

Because it is possible to construct a gauge-invariant and otherwise symmetry-compatible operator $\phi^\dagger\phi$ from the standard-model Higgs field, it is possible to construct, e. g. for a scalar dark matter particle d, a renormalizable coupling

$$\mathcal{L}_{\mathrm{HP}} = g d^\dagger d \phi^\dagger \phi$$

with an undetermined and free coupling constant g. Since now the Higgs interaction is the mediator to the dark matter sector, this is called a Higgs portal. If the dark matter particle carries a conserved symmetry, in the simplest case a parity for a real scalar particle, this also allows for a very massive dark matter particle, a so-called weakly-interacting massive particle (WIMP), without having all of the dark matter decaying during the evolution of the universe to standard-model particles.

Such scenarios are not easy to exclude, as g is in such simple models not constrained. It is also possible to have different Lorentz structures, internal symmetries, or even gauge symmetries for the dark matter sector. Also, multiple dark matter particles could all be coupled in this way. In experiment, this would show up as a too large invisible decay width of the Higgs in missing energy signatures. Given that the Higgs width cannot yet be measured directly, this gives relatively large freedom to create Higgs portals.

6.6 Left-right-symmetric models

A very constraining feature of the standard-model is the weak parity violation, as it forbids independent masses for fermions and imposes strong anomaly cancellation features. It is also quite cumbersome in supersymmetric extensions of the standard-model.

These problems are avoided in so-called left-right symmetric models. In such models the weak interaction is embedded in a larger gauge symmetry such that the uncharged right-handed standard-model fermions and the charged left-handed fermions are put into a common multiplet. This is arranged such that the right-handed particles correspond to charges then broken by a Brout-Englert-Higgs effect, and are thus no longer charged under the remaining gauge symmetry. The left-handed fermions are. Thereby, a variation on the GUT idea provides this effect. It is also possible to enlarge this scenario to have this effect in a full GUT[2].

Such scenarios therefore give rise to additional heavy gauge bosons and heavier Higgs siblings, at a, more or less, arbitrary scale at or above the TeV scale. This can therefore be tested by finding such particles.

6.7 Axions

A problem yet only briefly mentioned is the insufficient breaking of CP symmetry in the standard model. In fact, there exists another possible source of CP symmetry violation in the standard model. For both the weak interactions and the strong interactions it is possible to add a term

$$\mathcal{L} = \theta\epsilon_{\mu\nu\rho\sigma}F^{\mu\nu}F^{\rho\sigma} \tag{6.3}$$

to the standard model, a so-called topological term with the vacuum angle θ. The latter is bounded for topological reasons. Since effects of such a term are genuine non-perturbative, and suppressed like $\exp(1/g)$, they are irrelevantly small for the weak interaction. However, for the strong interaction an upper limit of the order of 10^{-10} for θ exists[3].

While θ is an independent parameter of the standard model, and its value therefore needs to be taken from experiment, its value is so close to zero that it is suspicious. It is not simple to find a structural embedding of the standard model such that θ becomes

[2]In fact, the SU(5) GUT of section 7.2 treats left-hand fermions and right-hand fermions both vectorial, but in different multiplets of the gauge symmetry.

[3]Formal reasons suggest that actually a term like (6.3) is generically not contributing. However, even in absence of this motivation for axions, the ensuing phenomenology for axions remains valid as an independent (less motivated) scenario.

zero, or at least very small. Thus, there is no easy solution to this so-called strong CP problem.

One possibility is that this is actually a dynamical effect. To achieve this, an additional scalar field, the so-called axion, is introduced, which couples to the topological term (6.3). Adding a suitable symmetry breaking to the axion sector, the term becomes dynamically suppressed, and therefore compensates strong CP violation. This symmetry, known as the Peccei-Quinn symmetry, is usually a global U(1).

In addition, such axions can be designed such that they can also act as dark matter candidates, therefore resolving two problems at once. Such axions would be produced in strong interactions, but usually strongly suppressed as they only couple through (6.3). Especially, this is (at least) a dimension five operator, and therefore suppressed by a scale related to the axion. Still, this implies that sources with a lot of strong interactions, e. g. the sun, will produce eventually axions, and these are therefore accessible in direct detection experiments.

A generalization of axions are axion-like particles (ALPs). They are usually not motivated by any issue in particular, but merely denote particles which weakly interact with the standard model. This can happen, as in the axion case, by a higher-dimensional operator or by utilizing the fact that hypercharge is an Abeian interaction. In the latter case, they receive a very small electric charge, of order 10^{-4} electron charges, or less. These are then also called milli-charged particles. Also known as feebly interacting particles (FIMPs), such particles can be both light or heavy. They can either be considered as one more possibility of new physics or as a dark matter candidate.

A unique signature of such particle is their ability to regenerate photons. By shining a laser on a wall, the laser is absorbed. However, milli-charged particle can be created from this beam, and then travel through the wall. Interacting with a strong electromagnetic field afterwards, they can be agained converted into photons, and thus a weak laser beam emerges after the wall. Such light-shining-through a wall experiments have been done, but so far without any hints of a signal.

6.8 Inflaton and quintessence

Another problem solvable by one or more additional scalar fields is the inflation problem of section 3.4.2. It can be shown that already the electroweak phase transition and the strong phase transition[4] leads to an inflationary period, but in both cases far too short and too ineffective as they are not first order. Thus, having a third phase transition in an

[4]Actually, to the best of our current knowledge both are only crossovers rather than phase transition.

additional sector could solve the problem. For this an additional scalar field with suitable potential and symmetry can be introduced, the inflaton[5].

However, it turns out that there is a reason for why both known cases are far too inefficient. In both cases the potential rises at large field values (classically) like a fourth power. Such a steep potential is accelerating phase transitions too much to yield long inflationary periods. To avoid this problem, slower rising potentials are necessary, e. g. of type $\phi^2 \ln \phi$, giving rise to the so-called slow-roll mechanism. While such potentials are in a quantum-field theoretical setting difficult to handle, they show in a quasi-classical treatment promising results.

Such a mechanism, depending on the energy scale it would act upon, could be discovered in the properties of the cosmological microwave background, especially its polarization (so-called r-mode). This is essentially an imprint of the gravitational waves created by the process. A direct observation of the gravitational wave background radition would also be able to provide information about such a mechanism.

[5]There are many similar scenarios, and the name quintessence is also attached to them.

Chapter 7

Grand unified theories

The first example of a high-scale BSM scenario will be the grand-unified theories (GUTs). While they become only relevant at rather high scales, compared e. g. to 1 TeV, they are very often needed to make extensions at a lower scale complete or more consistent. Thus, it is worthwhile to start with them, and have them available later. Note that supersymmetric GUTs form a field of its own, not touched upon in this lecture.

7.1 Setup

The most important motivation for GUTs is the following: As outlined before, the fact that the electromagnetic couplings have small ratios of integers for quarks and leptons cannot be explained within the standard model. However, this is necessary to exclude anomalies, as has been discussed beforehand. This odd but important coincidence suggests that possibly quarks and leptons are not that different as it is the case in the standard model. The basic idea of grand unified theories is that this is indeed so, and that at sufficiently high energies a underlying symmetry relates the gauge interactions of quarks and leptons, enforcing these ratios of electric charge. This is only possible, if the gauge interactions, and thus the gauge group $\mathrm{SU}(3)_{\mathrm{color}}\times\mathrm{SU}(2)_{\mathrm{weak}}\times\mathrm{U}(1)_{\mathrm{em}}$ is also embedded into a single group, since otherwise this would distinguish quarks from leptons due to their different non-electromagnetic charges. Another motivation besides the electromagnetic couplings for this to be the case is that the running couplings, the effective energy dependence of the effective gauge couplings, of all three interactions almost meet at a single energy scale, of about 10^{15} GeV, the GUT scale, as has been discussed in section 4.10.3.1. They do not quite, but if the symmetry between quarks and leptons is broken at this scale, it would look in such a way from the low-energy perspective. If all gauge interactions would become one, this would indeed require that all the couplings would exactly match at some energy scale.

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

The basic idea is that such a GUT is as simple as possible. The simplest version compatible with just the structure of the standard model requires to have a Yang-Mills theory with a single, simple gauge group, and the matter fields belong to given representations of it. As noted, the standard model gauge group is $\mathrm{SU}(3)_{\mathrm{color}}\times\mathrm{SU}(2)_{\mathrm{weak}}\times\mathrm{U}(1)_{\mathrm{em}}$. This is a rank 4 Lie group, 2 for SU(3) and 1 for SU(2) and U(1). Thus, at least a group of rank 4 is necessary, excluding, e. g., SU(4) with rank 3 or G_2 with rank 2. Furthermore, fermions are described by complex-valued spinors, and thus complex-valued representations must exist. This would be another reason against, e. g., G_2, which has only real representations. Another requirement is that no anomalies appear in its quantization.

Taking everything together, the simplest Lie groups admissible are SU(5) or SO(10), both having rank 4, as well as the rather popular cases of rank 6, 7, and 8, the groups E_6, E_7, and E_8, respectively.

Now, take SU(5) for example. It has 24 generators, and thus 24 gauge bosons are associated with it. Since the standard model only offers 12 gauge bosons, there are 12 too many. These can be removed when they gain mass from a Brout-Englert-Higgs effect, if the masses are sufficiently large, say of order of 10^{15} GeV as well. Thus, in addition to new heavy gauge bosons, a number of additional Higgs fields, or other mediators of symmetry breaking, are necessary in GUTs.

It should be noted that the idea of GUTs in this simple version, i. e., just be enlarging the gauge group, cannot include gravity. This is forbidden by the Coleman-Mandula theorm of section 4.1. To circumvent it requires again to move to supersymmetric GUTs.

7.2 A specific example

Lets take as a specific example SU(5) for the construction of a GUT. It has 24 generators, and therefore there are 24 gauge bosons. 8 will be the conventional gluons G_μ, 3 the W_μ of the weak isospin bosons, and 1 the hypercharge gauge boson B_μ, leaving 12 further gauge bosons. These 12 additional gauge bosons can be split in four groups of three X, Y, X^+, and Y^+, making them complex in contrast to the other gauge bosons for convenience. The general gauge field A_μ can then be split as

$$\begin{aligned} A_\mu = A_\mu^a \tau_a &= G_\mu^i \mathrm{diag}(\lambda^i, 0, 0) + W_\mu^i \mathrm{diag}(0,0,0,\sigma^i) \\ &\quad - \frac{1}{\sqrt{15}} B_\mu \mathrm{diag}(-2,-2,-2,3,3) + \sqrt{2}(X_\mu^c x^c + Y_\mu^c y^c + X_\mu^{c\dagger}\xi^c + Y_\mu^{c\dagger}\chi^c) \\ &= \sqrt{2}\begin{pmatrix} & & & X_\mu^{1\dagger} & Y_\mu^{1\dagger} \\ & \frac{1}{\sqrt{2}} G_\mu^i \lambda^i & & X_\mu^{2\dagger} & Y_\mu^{2\dagger} \\ & & & X_\mu^{3\dagger} & Y_\mu^{3\dagger} \\ X_\mu^1 & X_\mu^2 & X_\mu^3 & & \\ Y_\mu^1 & Y_\mu^2 & Y_\mu^3 & \multicolumn{2}{c}{\frac{1}{\sqrt{2}} W_\mu^i \sigma_i} \end{pmatrix} - \frac{1}{\sqrt{15}} \begin{pmatrix} -2 & & & & \\ & -2 & & & \\ & & -2 & & \\ & & & 3 & \\ & & & & 3 \end{pmatrix} B_\mu \end{aligned} \tag{7.1}$$

where λ are the Gell-Mann matrices, σ are the Pauli matrices, and the remaining generators of SU(5), the matrices x^μ, y^μ, ξ^μ, and χ^μ, have no entries on the diagonal. This assignment is necessary to obtain the correct charges of the known gauge bosons. This can be seen as follows. Gauge bosons transform under an algebra element as

$$[\tau^a, A_\mu] = [\tau^a, \tau^b] A_\mu^b = i f^{abc} \tau^c A_\mu^b.$$

Take for example τ^a to be the generator of B_μ. This matrix commutes with the ones of all the normal gauge field bosons, so their contributions are zero[1]. These particles are not charged. However, the matrices associated with the new gauge bosons do not. The appearing coefficients f^{abc} then show that the gauge bosons X_μ and Y_μ carry electric charge 5/3, while their complex conjugate partners have the corresponding anti-charge $-5/3$. In much the same way it can be shown that the three elements of each of the four fields can be arranged such that these gauge bosons carry the same color and weak isospin as the (left-handed) quarks and leptons. Since they can therefore couple leptons and quarks directly, they are referred to as leptoquarks, mediating e. g. proton decay as discussed below.

Arranging the fermions turns out to be a bit more complicated. Each family consists of 16 fermionic particles, 12 quarks and 4 leptons, counting two quark and lepton flavors

[1] The mixing of the weak fields and the photon has not been performed, therefore the weak isospin bosons are electrically neutral.

with three colors for the quarks and left-handed and right-handed chiralities separately. The fundamental representation of SU(5) is only of dimension 5, and can therefore not accommodate this number of particles. Also, the assignment in multiple copies of the fundamental representation cannot yield the correct quantum numbers. Therefore, the matter fields must be arranged in a non-trivial way.

It is an exercise in group theory, not to be repeated here in detail, that the simplest possibility is to assign the 16 particles to three different multiplets. In this construction the right-handed neutrino ν_R become a singlet under SU(5). Since already in the standard model it couples to the remaining physics only by the Yukawa coupling to the Higgs, and thus with a strength measured by its very small mass, this appears appropriate. The remaining particles of a family are put in two further multiplet structures. The right-handed down quarks and left-handed electron and electron neutrino can be put into an anti-5 (anti-fundamental) multiplet ψ

$$\psi = \begin{pmatrix} d_1^c \\ d_2^c \\ d_3^c \\ e_L \\ -\nu_L \end{pmatrix},$$

while the remaining particles can be arranged in a 10-multiplet χ

$$\chi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & u_3^c & -u_2^c & -u_L^1 & -d_L^1 \\ -u_3^c & 0 & u_1^c & -u_L^2 & -d_L^2 \\ u_2^c & -u_1^c & 0 & -u_L^3 & -d_L^3 \\ u_L^1 & u_L^2 & u_L^3 & 0 & -e^c \\ d_L^1 & d_L^2 & d_L^3 & e^c & 0 \end{pmatrix},$$

where c denotes the charge-conjugate of a right-handed particles ψ_R, i. e.,

$$\psi^c = i\gamma_2\gamma_0(\bar{\psi}_R)^T.$$

This multiplet structures is in fact necessary to provide an anomaly-free theory.

This appears to be a quite awkward way of distributing particles, and also not be symmetric at all. However, without proof, this distribution yields that all fermions have the correct quantum numbers. In particular, the correct electric charges are assigned - and exactly those. Hence, this embedding implies the mysterious relation of quark charges and lepton charges of the standard model in a natural way. Furthermore, it also implies that right-handed quarks are not interacting weakly, in this sense yielding parity violation of the weak interactions as well.

The remaining problem is now the presence of the X and Y gauge bosons. At the current level, these are massless. Even if they would be coupled to the Higgs field of the standard model, this would have to occur in the same way as with the W and Z bosons, thus yielding approximately the same masses. That is in contradiction to experiments, and therefore some way has to be found to provide them with a sufficiently heavy mass as to be compatible with experiments.

The simplest possibility is to have again a Brout-Englert-Higgs mechanism, like in the electroweak sector. Since the latter may not be affected, two sets of Higgs fields are necessary. The simplest possibility is to have one multiplet of Higgs fields $\Sigma = \Sigma^a \tau^a$ in the 24-dimensional adjoint representation of SU(5), and another one H in the fundamental five-dimensional one. The prior will be used to break the SU(5) to the unbroken standard model gauge group, and the second to further break it to the broken standard model.

To have the correct breaking of SU(5), the vacuum expectation value for Σ must take the form $\langle\Sigma\rangle = w\mathrm{diag}(1,1,1,-3/2,-3/2)$. The 3-2 structure is necessary to guarantee that the condensate is invariant under SU(3)-color and SU(2)-weak-isospin rotations. Such a condensate can be arranged for with an appropriate self-interaction of the Higgs fields.

That this condensation pattern removes only the X and Y gauge bosons can be directly seen from the interaction of Σ with the gauge bosons, which is given by

$$\mathcal{L}_\Sigma^{\text{kinetic}} = \frac{1}{2}\mathrm{tr}((\partial_\mu\Sigma - ig[A_\mu,\Sigma])^+(\partial^\mu\Sigma - ig[A^\mu,\Sigma])) \overset{\Sigma\to\langle\Sigma\rangle}{=} \frac{25}{8}g^2w^2(X_\mu^+X^\mu + Y_\mu^+Y^\mu),$$

where g is the SU(5) gauge coupling. The structure of the remaining term is then just that of a mass-term for the X and Y gauge bosons, and because of the particular structure chosen only for them. The corresponding masses can be read off directly and are

$$M_X = M_Y = \frac{5}{2\sqrt{2}}gw.$$

Choosing a potential such that w is sufficiently large thus makes the additional gauge bosons unobservable with current experiments. At the same time, any Higgs interactions having such a signature will give 12 of the 24 Higgs bosons of Σ a mass of order w as well. The other 12 are absorbed as longitudinal degrees of freedom of the X and Y gauge bosons. Thus no trace of them remains at accessible energies. Similarly choosing $\langle H\rangle = (0,0,0,0,v/\sqrt{2})$ yields

$$\begin{aligned}\mathcal{L}_H^{\text{kinetic}} &= (\partial_\mu H - igA_\mu H)^\dagger(\partial^\mu H - igA^\mu H)\\ &\overset{H\to\langle H\rangle}{=} \frac{g^2v^2}{4}\left(Y_\mu^+Y^\mu + W_\mu^+W^{\mu-} + \frac{1}{2\cos^2\theta_W}Z_\mu Z^\mu\right).\end{aligned}$$

This provides an additional mass shift for the Y bosons, and for the W and Z bosons their usual standard model masses. Out of the 10 independent degrees of freedom in the fundamental representation 3 are absorbed as longitudinal degrees of freedom of the W and Z bosons, leaving seven Higgs bosons. One of them has the quantum numbers of the standard model Higgs boson, while the other six decompose into two triplets (like quarks) under the strong interactions. By introducing appropriate couplings between Σ and H bosons, it is possible to provide these six with a mass of order w, thus also making them inaccessible at current energies.

It remains to show how fermion masses are protected from becoming also of order w. Actually, it is not possible to construct a coupling between Σ and the fermions which is renormalizable. However, it is possible to construct a Yukawa coupling to the H Higgs bosons. For example, for the first generation a mass term for the fermions is generated of type

$$\mathcal{L}_1 = -\frac{g_d v}{\sqrt{2}}(\bar{d}d + \bar{e}e) - 2\frac{g_u v}{\sqrt{2}}\bar{u}u,$$

where g_d and g_u are (arbitrary) Yukawa couplings. Since these are expressions at the GUT scale, this implies the same mass for down quarks and electrons at this scale. Transferring these results to the scale of the Z mass yields results which are in good agreement with experiment for some mass ratios, notably the bottom-to-τ ratio is about three, close to the experimental value of 2.4. However, in particular the light quark masses are not obtained reasonably well, showing that this most simple GUT is not sufficient to reproduce the standard model alone.

7.3 Running coupling

After this very specific example, it is also possible to make some more general statements, which will be done in this and the next section.

One of the motivation to introduce a grand-unified theory was the almost-meeting of the running couplings of the standard model when naively extrapolated to high energies. Because of the requirement that the structure of GUTs should be simple, all matter fields couple with the same covariant derivative to the GUT gauge bosons as

$$D_\mu = \partial_\mu - ig\tau_a A_\mu^a, \tag{7.2}$$

where g is the gauge coupling, A_μ^a the gauge boson field, and τ^a are the generators of the group, e. g. SU(5) yielding (7.1), in the representation of the matter fields. Here, it will be assumed that each generation of standard model matter fields fill exactly one (or more) multiplet(s) of the theory, but no further additional particles are needed to fill up the multiplets, and no multiplets contain particles from more than one generation.

The expression (7.2) has to be compared to the covariant derivatives of matter fields in the standard model, which is given by

$$D_\mu = \partial_\mu - ig_s\tau_a^s G_\mu^a - ig_i\tau_a^w W_\mu^a + ig_h\frac{y}{2}B_\mu.$$

The strong interaction is parametrized by g_s, the strong coupling constant, G_μ^a are the gluons, and the τ^s are either Gell-Mann matrices for quarks or zero for leptons. For the electroweak sector, only energies are considered at which the weak symmetry is essentially manifest. Therefore, it is useful to employ the corresponding notations. Then, the weak isospin bosons W_μ^a come with the weak isospin coupling g_i and the Pauli matrices τ^w. The influence of parity violation is neglected here for the sake of the argument[2]. Finally, there is the hypercharge gauge boson B_μ with the corresponding coupling g_h. Since the hypercharge group is the Abelian U(1), instead of representation matrices the hypercharge quantum numbers y appear, depending on the particle species in question, and have to be determined from experiment in the standard model.

Choosing a suitable basis with the same normalization of τ^s and τ^w unification implies that at the unification scale $g_s = g_i = g$. It is a bit more tricky for the hypercharges. One of the generators of the unified gauge group, say τ^h, must be proportional to the hypercharges y, $c\tau^h|_{\text{flavor}} = y$ for any given element of the matter multiplet. An example for SU(5) is given in (7.1). Staying with the assumption that each family belongs to one multiplet of the GUT implies that the corresponding hypercharges of the family members are essentially the eigenvalues of the generator τ^h. Taking the squared trace then yields

$$\frac{1}{4}\text{tr}(yy) = \frac{10}{3} \overset{!}{=} c^2\text{tr}(\tau^h\tau^h) = c^2 T_R$$

where no sum over h is implied but over the multiplet, and $T_R = 2$ is the Dynkin index[3]. Thus, $c = \sqrt{5/3}$, and hence $g = cg_h$. As emphasized earlier, the values y are not constrained by the standard model to have the prescribed values. Here, however, the values of y are fixed by the generator τ^h. This automatically requires the electric charges to have their values of the standard model. GUTs provide the quantization of electric charge observed in the standard model automatically, implying in particular that the electric charges of different particles have rational ratios.

The electric charge e is then given by

$$e = \frac{g_i g_h}{\sqrt{g_i^2 + g_h^2}} = \sqrt{\frac{5}{3}}g \tag{7.3}$$

[2]In case of the gauge group SU(5), as discussed in the previous section, the parity violation is actually manifest in the multiplet structure of the matter particles.

[3]Here a direct embedding of the SU(2) Pauli matrices for the gauge group of the GUT is used, requiring this normalization.

and the Weinberg angle θ_W by

$$\sin^2\theta_W = \frac{g_h}{\sqrt{g_h^2+g_i^2}} = \frac{3}{8}. \tag{7.4}$$

These relations only hold when the GUT's gauge symmetry is manifest, i. e. supposedly at the GUT scale. To check, whether this actually makes sense, it is necessary to let these values run down to the scale of the standard model and see whether the predictions agree with the observed values. The Weinberg angle (7.4) is useful here, as it known experimentally quite well.

Using (7.4), it follows that

$$\sin^2\theta_W = \frac{3}{8} - \frac{\alpha_i^{-1}+\alpha_h^{-1}}{2\pi}\frac{109}{24}\ln\frac{\Lambda_{\mathrm{GUT}}}{\mu}.$$

To eliminate the unknown scale Λ_{GUT} another of the evolution equations of section 4.10.3.1 can be used. Particularly convenient is the combination

$$\alpha_i^{-1}+\alpha_h^{-1}-\frac{8}{3}\alpha_s = \frac{67}{6\pi}\ln\frac{\Lambda_{\mathrm{GUT}}}{\mu},$$

yielding

$$\sin^2\theta_W = \frac{23}{134} + \frac{\alpha_i^{-1}+\alpha_h^{-1}}{\alpha_s}\frac{109}{201}.$$

Using the experimental values $\alpha_i^{-1}+\alpha_h^{-1} = 128$ and $\alpha_s = 0.12$ at the Z-boson mass, $\mu = M_Z$ yields $\Lambda_{\mathrm{GUT}} \approx 8\times 10^{14}$ GeV, $\alpha_s(\Lambda_{\mathrm{GUT}}) \approx 1/42$, and $\sin^2\theta_W(M_Z) = 0.207$. The latter number is uncomfortably different from the measured value of 0.2312(2), implying that at least at one-loop order this GUT proposal is not acceptable.

Unfortunately, this problem is not alleviated by higher-order corrections, and turns out to be quite independent of the particular unification group employed, and many other details of the GUT. This implies that unification cannot occur with the simple setup discussed here. Only when other particles, in addition to the minimum number needed to realize the GUT, are brought into play with masses between the electroweak and the GUT scale a perfect unification can be obtained. Supersymmetry, e. g., provides such a unification rather naturally, though also not in the simplest setup. Of course, the fact that Λ_{GUT} is much closer to M_P than to the electroweak scale could also be taken as a suggestion that quantum gravity effects may become relevant in the unification process. These are open questions.

7.4 Baryon number violation

A reason not to easily abandon GUT theories after the disappointment concerning the running couplings is that they naturally provide baryon number violation, which is so necessary to explain the matter-antimatter asymmetry of the universe. That such a process is present in GUTs follows immediately from the fact that quarks and leptons couple both to the same gauge group as one multiplet. Thus, gauge bosons can mediate transformations between them, just as the weak gauge bosons can change quark or lepton flavor individually. Whether some quantum numbers are still conserved depends on the details of the GUT. A GUT with gauge group SU(5), e. g., preserves still the difference of baryon number B and lepton number L, $B - L$. For the gauge group SO(10), not even this is conserved.

The profound consequence of baryon and lepton number violation is the decay of protons to leptons. It is, in principle, a very well defined experimental problem to measure this decay rate, though natural background radiation makes it extremely difficult in practice. To estimate its strength, assume for a moment that the masses of the gauge bosons mediating this decay are much heavier than the proton, which is in light of the experimental situation rather justified. Then the decay can be approximated by a four-fermion coupling, very much like a weak decay can be approximated by such a coupling at energies much smaller than the masses of the mediating W and Z bosons.

The corresponding interaction is then encoded in the Lagrangian

$$\mathcal{L} = \frac{4G_{\mathrm{GUT}}}{\sqrt{2}}(\bar{u}\gamma^{\mu}u\bar{e}\gamma_{\mu}d) \tag{7.5}$$

for quark fields u and d and the electron field e. This vertex permits that a d quark and a u quark scatter into a u quark and an electron. The corresponding decay channel of the proton would then be into a positron and a neutral pion, the lightest one permitted by electric charge and energy conservation involving charged leptons. The effective coupling G_{GUT} is then given by

$$\frac{G_{\mathrm{GUT}}}{\sqrt{2}} = \frac{g^2}{8m_{\mathrm{GUT}}^2} = \frac{\pi}{2m_{\mathrm{GUT}}^2(\alpha_i^{-1}(m_{\mathrm{GUT}}) + \alpha_h^{-1}(m_{\mathrm{GUT}}))},$$

in complete analogy to the weak case, and $m_{\mathrm{GUT}} \approx \Lambda_{\mathrm{GUT}}$ the mass-scale of the leptoquark gauge bosons. The life-time at tree-level can then be calculated in standard perturbation theory in leading order to be

$$\tau_{p\to e^+\pi^0} \sim \frac{192\pi^3}{G_{\mathrm{GUT}}^2 m_p^5} \sim \frac{m_{\mathrm{GUT}}^4(\alpha_i^{-1}(m_{\mathrm{GUT}}) + \alpha_h^{-1}(m_{\mathrm{GUT}}))}{m_p^5}.$$

Plugging in the previous numbers, the formidable result is about 10^{31} years. That appears quite large, but the current experimental limit in this channel is about 10^{34} years, clearly exceeding this value. Thus, at least this very simple approximation would yield that a GUT is in violation of the experimental observation by about three orders of magnitude.

However, pre-factors and higher order corrections depend very much on the GUT under study, and can raise the decay time again above the experimental limit. Finding proton decay, or increasing further the limit, would provide therefore information on the structure of permitted GUTs. However, the search becomes experimentally more and more challenging, so that pushing the boundaries further is an expensive and demanding challenge. Nonetheless, this is a worthwhile problem: If no proton decay should be observed ere reaching the standard model decay rate, this would imply that baryon number violation would proceed in a rather unexpected way. Or would need to be an initial condition of the big bang.

7.5 Flavor universality violations and leptoquark phenomenology

The explicit examples of GUTs so far used fermion multiplets which do not mix generation. However, larger representation or larger unification gauge groups would allow to embed multiple generations into a multiplet. This could also explain the existence of generations. Such a setup has one additional feature compared to the previous ones. Because the different generations reside in the same multiplet, leptoquarks can no mediate intergeneration decays. As a consequence, a GUT would enhanced such effects compared to those in the standard model, where they stem from the off-diagonal Higgs interaction, yielding the CKM/PMNS matrix[4].

In the standard model, the only difference between the generations stems from the masses. The intergeneration mixing in the quark sector due to the CKM matrix creates well-established effects. However, this also implies that there is quite some background for any searches. In the lepton sector, however, the smallness of the neutrino masses make intergeneration transitions negligbly small in the charged lepton sector[5]. Eliminating mass effects, leptons should be treated (almost) universally in the same way in the standard model. Lepton flavor universality violations (LFUV) are thus a very clean signal for some GUT scenarios. It would also provide a clean access to leptoquark physics, as it is parametrized by similar couplings as (7.5).

Despite some hints in (semi-)leptonic decays of heavy mesons, however, so far no unambigous signal of lepton-flavor universlity violations have been found. Still, it remains an important search channel, due to its small standard model background. This is also true for oscillations experiments, e. g. $e \to \mu\gamma$.

[4]Of course, rather than to enhance, they could replace them, making such a scenario harder to detect

[5]Neutrino oscialltions are actually the corresponding effect in the neutrino sector

7.6 Asymptotic safety

There is one interesting feature of the β-function (2.7). Formally, there can be β-functions such that the running coupling goes to a finite value if the energy scale is send to infinity. This is a third option compared to asymptotic freedom, in this context also called a Gaussian fixpoint, where the coupling vanishes, and an infinitely strongly coupled theory. This third scenario is called asymptotic safety.

Studying the perturbative expressions for the coefficients (5.2-5.4) already suggests that it should be possible to construct such a solution by a judicious choice of the theory content even at weak asymptotic coupling. This has indeed been done. It is yet not clear whether any such constructed theory could resolve any of the problems of the standard model in a convincing way, but it is certainly an interesting option.

Even more interesting is that theories which are perturbatively ill-defined by the appearance of a Landau pole could change to an asymptotically safe theory once the non-perturbative β-function is used. Of course, this requires to know the later to check. Fortunately, in recent times progress has been made in non-perturbative calculations. In some cases, this is also possible by adding suitable matter content even at weak coupling.

As a consequence, unified models with many fermions and scalars, and sufficiently large gauge groups, appear as a viable, ultraviolet stable scenario. It remains to be seen whether they are also phenomenologically viable. However, so far they do not seem to encounter more serious problems than other GUT candidates.

7.7 The physical spectrum of GUTs

The discussion of the BEH effect in the standard model in section 2.1, and those following on BSM physics, was performed in perturbation theory. This is actually not quite correct, as will be discussed now. While this has (likely) implications for the selection of which theories are exactly suitable extensions of the standard model, this does not touch upon the qualitative properties discussed, for which perturbation theory remains therefore a suitable guideline.

The problem arises as that in a non-Abelian gauge theory the asymptotic state space can, in principle, not contain any elementary particles. The reason is that the asymptotic fields cannot be free fields, since otherwise the state space has changed from a space of gauge-dependent objects to one of gauge-singlets, and thus a local symmetry would become a global symmetry. These two spaces are not unitarily equivalent, and therefore this is strictly speaking not possible beyond perturbation theory where all results are by construction smooth in the gauge coupling. A simple example is already QED: In perturbation theory only electrons, protons, and photons appear, but no hydrogen atom, despite being a stable state.

This point can be formalized in the context of axiomatic field theory, and is known as Haag's theorem: The state spaces of an interacting theory and a non-interacting theory are not unitarily equivalent, no matter how weak the coupling. Hence, strictly speaking perturbation theory expands around the wrong vector space. However, this theorem does not make any statements about the quantitative size of the non-analytic contributions. It is thus well possible that they are a negligible effect, and thus perturbation theory implicitly assumes that this is the case, and the dominant contribution comes actually from the analytic part. In the standard model, this seems to be true, vindicating the discussion of section 2.1, and indeed perturbation theory describes exceedingly well observations. But it will be seen that this does not need to be true beyond perturbation theory.

While in the standard model this is found to be a small effect, this is due to the special structure of it. This does not need to hold beyond, and turns out indeed to be a special problem for GUTs.

Hence, in the following a correct construction will be provided, and in the end shown why, and under which conditions, perturbation theory can still give the dominant part of the answer. To establish the answer, it is useful to start just with the standard model, and neglect for the moment all non-essential parts. This amounts to the weak gauge fields, now yielding degenerate masses for the $W^{\pm}$ and Z because of the absence of QED, or more precisely the hypercharge, and the Higgs.

7.7.1 The Osterwalder-Seiler-Fradkin-Shenker argument

If now the symmetry cannot be broken, the question is what is about the apparent symmetry breaking by the vacuum expectation value of the Higgs field. The answer is that it was actually a gauge condition which gave the Higgs a vacuum expectation value. E. g. the 't Hooft gauge condition of section 2.1.2 singles out a particular direction by explicitly introducing a choice of direction for the vacuum expectation value of the Higgs field. However, this choice is part of the gauge choice, and any choice of direction would yield an equally valid, though possibly more cumbersome, result.

Now, rather than fixing a direction once and for all, it is equally possible, just as in the construction of general linear covariant gauges, to average over all possible such choices. Then, the result would be that the vacuum expectation values would be the average over all possible direction, but this is zero, as all directions are equally preferred. Actually, without fixing this global degree of freedom the same result would be ensuing.

This seems to have drastic consequences, as without vacuum expectation value the whole construction breaks down, and especially there is no tree-level mass for the gauge bosons. This is in fact correct, and actually it can be shown that in such a gauge the masses of the gauge bosons remain massless to all orders in perturbation theory. But this is not a consequence of picking somehow a 'wrong' gauge: All gauge choices, which can be satisfied by all orbits[6] are equally acceptable. Thus, this cannot be a conceptual problem. In fact, in such gauges the fluctuations of the Higgs field are no longer small enough to justify perturbation theory, and hence the applicability of perturbation theory rests on the choice of a suitable gauge. In a more simple diction, this is just the statement that only in suitable coordinates perturbation theory makes sense.

In this section, the main question is different: Since the non-vanishing of the Higgs expectation value is apparently only due to the choice of a particular gauge, how it is still possible to identify the Brout-Englert-Higgs effect? This question has two layers.

The first is how to construct a quantity, which is still identifying the Higgs effect, even if the direction of the Higgs condensate is not fixed by the gauge choice. In the analogy of a magnet, on any single field configuration in the path integral, the Higgs field will still be aligned. Thus, the relative orientation of the Higgs field would not be influenced, especially as the different possibilities of direction in the 't Hooft gauge condition are connected by a global gauge transformation. Thus, an observable like

$$\langle v_2 \rangle = \left\langle \left| \int d^d x \phi(x) \right|^2 \right\rangle \tag{7.6}$$

would have the desired property. Note that a quantity like

$$\langle v^2 \rangle = \left\langle \int d^d x \left| \phi(x) \right|^2 \right\rangle,$$

would not work. Though it is non-zero for non-vanishing relative local alignment, it will actually never vanish, expect when the Higgs field is only in a measure-zero region of space non-zero, and vanishes otherwise. However, especially in a scalar-QCD-like phase, this can hardly be expected, and thus this observable cannot distinguish between a QCD-like behavior and a BEH-like behavior.

However, in a gauge theory this is not enough[7]. To show that this really distinguishes

[6] Or actually by all orbits up to a measure zero set.

[7] Note that for global symmetries similar considerations apply, and without explicit symmetry breaking a quantity like (7.6) would be more appropriate than the usual local order parameters, which do not involve an integration. Indeed, the ordinary local order parameters vanish without any external disturbance breaking explicitly the symmetry, and the symmetry remains unbroken. Thus, parameters like (7.6) should be rather seen as an indication for a metastability against external explicit symmetry breaking, rather than a real breaking of symmetries.

between the BEH case and any alternatives, the observable must also be gauge-invariant under local gauge transformations, and (7.6) is not.

Thus, the question is, whether there is any gauge-invariant possibility to detect the BEH effect. The answer to this appears to be that it is not the case. However, the reasoning, the so-called Osterwalder-Seiler-Fradkin-Shenker argument, is not entirely trivial, and there is at least one loophole.

The problem is that to answer this question it is necessary to go beyond perturbation theory, as it was already seen that perturbation theory provides not even for the restricted case of only differing global gauge choices the correct answer. But calculations beyond perturbation theory are always more involved, and often require assumptions and/or approximations.

The probably strongest statement about the situation in the present theory can be obtained using a so-called lattice discretization, i. e. an approximation where rather than to consider the ordinary space-time, the situation is considered on a discrete and finite lattice of space-time points. The original theory is then obtained in the limit of infinite volume and zero spacing between them. For asymptotically free theories, it can be shown that there is some neighborhood around infinite volume and zero discretization where the approach becomes smooth, and thus this is a valid approach to deal with them[8]. But for not asymptotically free theories, especially those suffering from the triviality problem of section 3.1, no such statement exists[9].

Thus, for the following it is necessary to make the assumption that either the limit exists and is smooth, or if not, this has no direct implication for the result. The latter is not a too high a hope: Since this only states that it should be valid up to at least some maximum discretization, which corresponds to some maximum energy, this is the statement that the results should be true in the sense of a low-energy effective theory.

The steps for the construction will only be outlined, as the technical details are too involved to present them here, and would require a thorough discussion of a discrete formulation of the theory. The first restriction is to work at fixed Higgs length $\phi^\dagger\phi = 1$. This is actually only a technical simplification, and can be dropped. This situation is obtained when sending the Higgs-self-coupling to infinity.

The next step is to switch to unitary gauge. This is always possible, since unitary gauge does not require the BEH effect to be active to be well-defined, in contrast to 't

[8] Though in practice it is usually impossible to make reliable statements on how large this neighborhood is.

[9] Actually, this can be an indication that the theory just does not exist without an explicit cutoff, and then the theory is ill-defined, no matter the method.

Hooft gauges[10]. Since the length of the Higgs field is fixed, there are no Higgs degrees of freedom left in the action, and the action is classically minimized by a vanishing gauge field. It is for this fact important that there is a Higgs field and that the Higgs field fully breaks the gauge symmetry. Otherwise other configurations could minimize the action.

Consider now any gauge-invariant operator[11]. Since the only gauge-invariant operators possible are compositions of the terms in the Lagrangian, any such operator can also be written as composition of such gauge-invariant operators. Thus, the full expectation value must be equivalent to a path integral over such gauge-invariant operators.

In the next step, expand the exponential in a series in these operators around vanishing fields, and thus vanishing field-strength tensors. On a finite, discrete lattice, this will always result in a convergent series.

The series can be merged with the expression for the gauge-invariant operator. Thus, the result is some series in gauge-invariant operators. Each term of the series is analytic. On a finite lattice, it can then be shown that this series is, for any gauge-invariant operator, bounded from above by a geometric series parametrized by the parameters. This is again only possible because of the additional potential term induced by the Higgs effect, and thus the presence of one additional parameter. The series is therefore uniformly bounded, and since every term is analytic, a general mathematical theorem guarantees then that the whole expression is an analytic function.

The whole argument fails only if any parameter of the theory either vanishes or diverges. Thus, on the boundaries of the phase diagram it is still possible to have a phase transition, but there can be no phase transition cutting the phase diagram in separate disconnected pieces. Thus, the phase diagram is connected, though may have phase transitions with end-points, and, of course, cross-overs.

It is visible that being on the lattice is important in the argument. It was also important that all Higgs degrees of freedom could be removed by either freezing or using the unitary gauge in an intermediate step. If the number of Higgs degrees of freedom is such that this is not possible, the argument does not hold. Thus, if the gauge symmetry is only partly broken by the Higgs field, a separation may still exist. Also, if there are surplus Higgs fields or other BSM structures, the minimum structure may be more complicated, and the argument may not apply. Finally, when adding the remainder of the standard model, the

[10]Fixing a gauge is permitted, as only gauge-invariant statements are made, and no approximations are performed which would break gauge invariance. Thus, the final result is gauge-invariant even though a gauge has been fixed in an intermediate step.

[11]The so-called Gribov-Singer ambiguity in gauge-fixing beyond perturbation theory is one of the reasons why this proof does not pertain to gauge-dependent quantities, and they may, and do, change non-analytically in the phase diagram, providing the perturbative picture of the BEH-QCD separation.

situation is more involved, especially due to the presence of the fermion fields, and there is no similar simple argument. Thus, the phase diagram of more complicated theories has not yet been classified with the same level of rigor.

7.7.2 The Fröhlich-Morchio-Strocchi mechanism

In the previous subsection the problem arose that the Higgs and W/Z fields are actually not really gauge-invariant, and in fact the whole Higgs mechanism is not. The question thus arose what is actually measured when seeing peaks associated with electroweak particles in cross sections. As before, it is simpler to first discuss only the case with the Higgs and the gauge bosons and afterwards continuing to include the remainder of the standard model, which in this case is actually possible. Finally, it will be discussed how this gives rise to conflicts in BSM theories.

The first realization necessary is that to describe physical objects requires operators which are manifestly gauge-invariant. For a non-Abelian gauge theory, like the one under discussion, this is only possible in case of composite operators, i. e. operators involving more than a single field, since any single-field operators are gauge-dependent.

Such gauge-invariant operators can then only be classified in terms of global quantum numbers, i. e. in the present case spin and parity as well as the custodial structure. Any open gauge index would yield that the quantity in question would change under a gauge transformation.

The simplest example of such an operator would be

$$\mathcal{O}_{0^+}(x) = \phi_i^\dagger(x)\phi_i(x),$$

created from the Higgs field ϕ and being a scalar and a singlet under the custodial symmetry, as well as a gauge-singlet. This operator creates a Higgs and an anti-Higgs at the same space-time point, and therefore corresponds to a bound state of two Higgs particles, just like a meson in QCD. It is a well-defined physical state, and therefore observable.

So far, this is formally all correct. However, the immediate question appearing is that the description of the observed Higgs agrees very well with the one obtained in perturbation theory, and thus the elementary Higgs. However, such a bound state, as is shown in QCD, can have widely different properties from its constituents. Thus, the two views seem to be at odds with each other.

However, there is a resolution for this apparent paradox, the so-called Fröhlich-Morchio-Strocchi (FMS) mechanism. The mechanism itself will actually not be the explanation, as it is actually only a description of how to determine perturbatively the mass of this state.

To do this, consider the propagator of the composite state,

$$\langle \mathcal{O}_{0^+}(y)^\dagger \mathcal{O}_{0^+}(x)\rangle = \langle \phi_j^\dagger(y)\phi_j(y)\phi_i^\dagger(x)\phi_i(x)\rangle.$$

As usual, the poles of this correlation function will give the mass of the particle. As the next step, select a gauge, like the 't Hooft gauge, in which the vacuum expectation value vn_i of the Higgs field does not vanish, and rewrite $\phi_i(x) = vn_i + \eta_i(x)$. Then perform a formal expansion in the quantum fluctuation field η, yielding to leading order

$$\langle \phi_j^\dagger(y)\phi_j(y)\phi_i^\dagger(x)\phi_i(x)\rangle = v^4 + v^2\langle \eta_i^\dagger(y)\eta_i(x)\rangle + \mathcal{O}\left(\left(\frac{\eta_i}{v}\right)^3\right).$$

Neglecting the higher order contributions, the only pole on the right-hand side is the one of the propagator of the fluctuation field. Thus, to this order, the masses coincide[12], and the bound state has the same mass as the elementary particle, showing why the perturbative result provides the correct mass for the observable state. Thus, this justifies why it is correct to use perturbation theory, and the perturbative spectrum, to obtain the mass of the Higgs[13].

In the same way, it is possible to construct a non-perturbative partner state for the gauge bosons,

$$\begin{aligned} \mathcal{O}^a_{1^-\mu}(x) &= \mathrm{tr}\tau^a X^\dagger D_\mu X \\ X &= \begin{pmatrix} \phi_1 & -\phi_2^* \\ \phi_2 & \phi_1^* \end{pmatrix}, \end{aligned} \tag{7.7}$$

which is a custodial triplet, and a gives the corresponding index. Using that the vacuum expectation value is constant, this yields

$$\langle \mathcal{O}^{a\dagger}_{1^-\mu}(y)\mathcal{O}^a_{1^-\mu}(x)\rangle \sim v^4 g^2\langle W_\mu^i(y) W_i^\mu(x)\rangle + \mathcal{O}\left(\frac{\eta}{v}\right)$$

and thus the mass of the W and Z are obtained, as well as the correct number of states, trading a custodial triplet for a gauge triplet. Note that because the masses of the gauge bosons are both scheme-invariant and gauge-parameter-invariant in perturbation theory in 't-Hooft-type gauges, this is actually an even stronger statement than for the Higgs itself.

[12]Beyond leading order in the weak coupling constant the mass of the Higgs becomes scheme-dependent. It is then necessary to do this comparison in the pole scheme.

[13]The validity of the expansion, and whether for a given set of parameters, the expansion is actually valid is a dynamical question, and requires to determine both sides non-perturbatively, or the left-hand-side by experiment. It works for the ones in the standard model, but by far not for all possible parameter sets of the theory.

It is possible to construct also operators for other quantum numbers, but only these two channels have a leading non-zero contribution given by one of the elementary fields. This also implies that in this expansion there are no other bound states than just these two[14].

This shows why the perturbative predictions provide the correct results. In fact, also scattering processes are dominated by the higher-order perturbative corrections, if the ratio η/v is sufficiently small. Hence, to a very good approximation a perturbative description of this theory can be sufficient. Given the good accuracy of the perturbative description of the most recent experimental results, the non-perturbative corrections for the investigated processes are at most at the percent level, at least at currently accessible energies.

7.7.3 Adding the rest of the standard model

Adding the remainder of the standard model is possible, but requires a careful distinction of the various cases. Right-handed neutrinos, if the neutrinos are also Dirac fermions, are anyhow gauge singlets, and therefore pose no problems.

For left-handed (or Majorana) neutrinos and leptons a problem arises. These particles are not confined, and carry a weak charge. However, a similar solution exists as for the Higgs and the weak gauge bosons. Form the composite operator

$$\mathcal{O}_{\frac{1}{2}}(x) = \phi_i(x)\psi_i(x),$$

where the field $\psi_i(x)$ is a (left-handed) fermion field of any of the above enumerate types. Because the Higgs is a scalar, this hybrid is still a spin-1/2 fermion. The correlation function expands then as

$$\langle(\phi_i(y)\psi_i(y))^\dagger\phi_j(x)\psi_j(x)\rangle \sim v^2\langle\psi_i(y)^\dagger\psi_i(x)\rangle + \mathcal{O}\left(\frac{\eta}{v}\right),$$

and therefore to the elementary fermion propagator, showing in the same way that the bound state has the same mass as the elementary fermion. Again, beyond leading order, the elementary mass has to be evaluated in the pole scheme.

Colored particles are forced asymptotically into hadrons due to confinement. Hadrons, like mesons, which are also with respect to the weak gauge symmetry singlets are therefore gauge-invariant. However, this is not the case for those states which are intrageneration-non-flavor-singlets, like nucleons. Since intrageneration flavor is actually the weak gauge charge - up and down are gauge indices - these are again exchanged for custodial indices working very much as for the vector bosons and leptons, but on the level of hadrons.

[14] Whether this is true beyond leading order is still an open question. Since no formal proof exists, this requires to perform actual non-perturbative calculations, which is quite non-trivial.

Somewhat trickier is the situation with the U(1) hypercharge, or the electric charge. Electric charge is an observable quantity, in contrast to the weak (and color) charge. The reason for this originates from the Abelian nature of this interaction. Given a field $\phi(x)$ with an Abelian charge, it is possible to construct an operator of type

$$\exp\left(i\int ds_\mu A_\mu\right)\phi(x)$$

where A_μ is an Abelian gauge field, and the path is a closed path[15] originating at infinity and ending at x. Such a phase factor is also called a Dirac phase. This object is actually gauge-invariant, but carries a conserved charge, the electromagnetic charge. This is possible for an Abelian gauge theory, because the gauge fields are not matrix-valued, and therefore commute, which is the key in making the phase factor cancel in any gauge transformation. In a non-Abelian gauge theory, it is no longer possible[16] to construct such a canceling phase factor, and hence there is no gauge-invariant charge. Physically, this corresponds to an infinite superposition of particles described by the field ϕ and arbitrary many photons, and thus it is a combination of the particle and a photon cloud, which creates a state which is both gauge-invariant and charged. But again, this is only possible for Abelian symmetries[17].

This completes the standard model.

7.7.4 Beyond the standard model

The same considerations apply beyond the standard model. However, the key in the standard model was the global custodial symmetry could become a proxy for the weak gauge interaction, because it is the same group. Thus, a problem with multiplicities may arise, not to mention dynamical effects, if this is no longer the case.

Indeed, in some toy theories, like toy-GUTs with SU($N>2$) with a single Higgs field in the fundamental representation, this leads to qualitative differences in the physical spectra and the spectrum of elementary particles, which becomes arbitrarily bad with increasing N. The reason is that in this case only a U(1) custodial symmetry exists, which creates no non-trivial degeneracies, and especially not the triplet structure needed for the weak gauge bosons. This can be seen by considering the generalization of (7.7). Because there

[15]This is somewhat symbolically, and requires a more precise formulation to avoid a path dependence.

[16]There is no full proof yet, but the evidence is overwhelmingly substantial.

[17]There are non-Abelian gauge theories for which a finite number of gauge bosons and matter fields create gauge-invariant states. These are, however, conventional bound states, and especially do not create a physical gauge charge.

is no non-trivial custodial symmetry, the corresponding operator is

$$O_\mu = \phi^\dagger D_\mu \phi.$$

The correlator then expands to leading order to

$$\langle O_\mu^\dagger O_\mu \rangle = n^a n^b \langle W_\mu^a W_\mu^b \rangle + \mathcal{O}\left(\frac{\eta}{v}\right),$$

where n is the direction of the vacuum expectation values. Herein n is the direction of the vacuum expectation value, and thus only the correlator $\langle W_\mu^N W_\mu^N \rangle$ in the direction of the Higgs vacuum-expectation value contributes. However, following the perturbative construction of section 7.2, it turns out that this is only the most massive gauge boson in the spectrum. Hence, especially no massless vector particles, which could play the role of photons or gluons, appear, and only a single state is present. This is not changed by higher orders. Thus, a different low-energy spectrum arises.

In other theories, different results arise, but generically such mismatches appear. But this does not need to happen. E. g. for the 2HDMs of section 6.1 no conflict, as in the standard model, arises. It is not yet generally clear, what is the decisive structural feature leading to agreement or disagreement between the perturbative and physical spectrum, but this is likely connected to the combination of gauge group, custodial group, and available representations.

The bottom line is that the possibility that a purely perturbative determination of the observable spectrum can fail. This implies that a careful (re)analysis of models are necessary to ensure that their observable spectrum can coincide with what is already known, the spectrum of the standard-model. This remains to be done for most of the theories discussed in this lecture.

Chapter 8

Large extra dimensions

As will be seen, the large difference in scale between gravity and the standard model can be explained by the presence of additional dimensions. Also, string theories, as discussed in chapter 9.1, typically require more than just four dimensions to be well-defined. Such extra dimensions are not (yet) seen, and therefore their effects must not (yet) be detectable. The simplest possibility to make them undetectable with current methods is by making them compact, i. e., of finite extent. Upper limits for the extensions of such extra dimensions depend strongly on the number of them, but for the simplest models with two extra dimensions sizes of the order of micrometer are still admissible. Such cases with extensions large compared to the Planck length are called large extra dimensions. They should be contrasted to the usually small extensions encountered in string theory, which could be of the order of the Planck length. Here, the observable consequences of such large extra dimensions will be discussed.

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

Here, a number of these models will be discussed briefly, and one particular simple example also in a certain depth to introduce central concepts like the Kaluza-Klein tower of particle states, in some detail.

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

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

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

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

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

8.1 Separable gravity-exclusive extra dimensions

8.1.1 General setup

The simplest example of large extra dimensions is given by theories which have n additional space-like dimensions, i. e., the metric signature is $\text{diag}(-1, 1, ..., 1)$. Furthermore, these additional dimensions are taken to be separable so that the metric separates into a product

$$g^{4+n} = g^4 \times g^n.$$

Furthermore, for the additional dimensions to be gravity exclusive the other fields have to be restricted to the 4-dimensional brane of uncompactified dimensions. In terms of the Einstein equation (2.18) this implies that the total energy momentum tensor T_{MN} takes the form

$$T_{MN} = \begin{pmatrix} T_{\mu\nu} & 0 \\ 0 & 0 \end{pmatrix}, \tag{8.1}$$

where the indices M and N count all dimensions and μ and ν only the conventional four. Furthermore, in such models the extra dimensions are compact, having some fixed boundary conditions.

The Einstein-Hilbert action is then

$$S_{EH} = -\frac{1}{2G_{N4+n}^{1+n/2}} \int d^{4+n}z\sqrt{|g^{4+n}|}R^{4+n} \tag{8.2}$$

with again the generalized Newton constant G_{N4+n}, the metric g and the Ricci scalar R. The action then factorizes as

$$S_{EH} = -\frac{M_s^{n+2}}{2}\int d^{4+n}z\sqrt{|g^{4+n}|}R^{4+n} = -\frac{1}{2}M_P^2\int d^4x\sqrt{-g^4}R^4.$$

The actual gravity mass-scale M_s is related to the perceived 4-dimensional Planck scale by

$$M_P = M_s(2\pi RM_s)^{\frac{n}{2}} = M_s\sqrt{V_nM_s^n},$$

with the volume of the additional (compact) dimensions V_n, which have all the same compactification radius R. For an M_s of order 1 TeV, the compactification radius for $n = 2$ to $n = 6$ ranges from 10^{-3} to 10^{-11} m, being at $n = 2$ just outside the experimentally permitted range.

Treating the theory perturbatively permits to expand the metric as

$$g_{MN} = \eta_{MN} + \frac{2}{M_s^{1+\frac{n}{2}}}H_{MN},$$

with the usual Minkowski metric $\eta_{AB} = \mathrm{diag}(-1, 1, ..., 1)$ and the metric fluctuation field H_{AB}. The Einstein-Hilbert action is then given by an integral over the Einstein-Hilbert Lagrangian

$$\mathcal{L}_{EH} = -\frac{1}{2}H_{MN}\partial^2H^{MN}+\frac{1}{2}H_N^N\partial^2H_M^M-H^{MN}\partial_M\partial_NH_L^L+H^{MN}\partial_M\partial_LH_N^L-\frac{1}{M_s^{1+\frac{n}{2}}}H^{MN}T_{MN}.$$

Since the additional dimensions are finite, it is possible to expand h_{MN} in the additional coordinates in a series of suitable functions f_n, embodying the structure of the extra dimensions

$$\begin{aligned}
H_{MN}(x_0, .., x_3, x_4, ..., x_{3+n}) &= \sum_{m_1,...,m_n} f_n(k^4_{m_1,...,m_n}x_4 + ...k^{3+n}_{m_1,...,m_n}x_{3+n})H_{MN}(x_0, ..., x_3) \\
k_{m_1,..,m_n} &= \left(K\left(\frac{\pi m_1}{R}\right), ..., K\left(\frac{\pi m_n}{R}\right)\right)^T
\end{aligned}$$

The energies k can be related in the usual way to masses, $k_m^2 = m_{KKm}^2$, as in quantum mechanics. These are the Kaluza-Klein masses. The field h_{MN} for a fixed mass can then be decomposed into four four-dimensional fields. These are a spin-2 graviton field $G_{\mu\nu}$, $i = 1, .., n-1$ spin-1 fields[1] A^i_μ, $i = 1, ..., (n^2 - n - 2)/2$ scalars S^i and a single further scalar h. These obey equations of motions

$$\begin{aligned}
(\partial^2 + m^2_{KKn})G^n_{\mu\nu} &= \frac{1}{M_P}\left(T_{\mu\nu} + \left(\frac{\partial_\mu\partial_\nu}{m^2_{KKn}} + \eta_{\mu\nu}\right)\frac{T^\lambda_\lambda}{3}\right) \\
(\partial^2 + m^2_{KKn})A^{ni}_\mu &= \\
(\partial^2 + m^2_{KKn})S^i_n &= 0 \\
(\partial^2 + m^2_{KKn})h_n &= \frac{\sqrt{\frac{3(n-1)}{n+2}}}{3M_P}T^\mu_\mu.
\end{aligned} \tag{8.3}$$

Soft modes are the zero-modes of the Fourier-transformed fields, i. e., those with $m^2_{KK} = 0$. The fields A and S do not couple to the standard model via the energy momentum tensor, and the graviton coupling is suppressed by the Planck mass, in agreement with the observation that gravity couples weakly. This also applies for the radion h, which couples to the trace of the energy-momentum tensor, corresponding to volume fluctuations. However, because of its quantum numbers, it will (weakly) mix with the Higgs.

Finally, since $m_{KKn} \sim k_n \sim n/R$ for a mode n, the level splitting of the Kaluza-Klein modes is associated to the size of the extra dimensions. The splitting is thus given by

$$\delta m_{KK} = m_{KKn} - m_{KKn-1} \sim \frac{1}{R} \approx 2\pi M_s \left(\frac{M_s}{M_P}\right)^{\frac{2}{n}}.$$

which is generically of order meV for $n = 2$ to MeV for $n = 6$. Thus, to contemporary experiments with their limited resolution of states the tower of Kaluza-Klein states will appear as a continuum of states.

8.1.2 An explicit example

The simplest example of the general discussion before has n additional dimensions with the same size $R/(2\pi)$ and periodic boundary conditions, i. e., they are torus-like, and the total space is $M^4 \times T^n$, with M denoting a Minkowski space endowed with a metric g.

First consider only the gravity sector. To exhibit the general properties it is useful to make a perturbative expansion. In this case, the metric is rewritten as

$$g_{MN} = \eta_{MN} + 16\pi G_N^{4+n} H_{MN},$$

[1]Originally, Kaluza and Klein in the 1930s aimed at associating this field with the electromagnetic one, which failed.

where g is the full $4+n$-dimensional metric, η is the $4+n$-dimensional flat Minkowski metric diag$(-1,+1,...,+1)$, h denotes the $4+n$-dimensional metric fluctuation field, G_N^{4+n} is the $4+n$-dimensional Newton constant, and α and β will run[2] in the following from 1 to $4+n$. Assuming $G_N^{4+n}h$ to be only a small correction to η permits to expand the higher-dimensional Lagrangian of general relativity

$$\mathcal{L} = \frac{\sqrt{|\det g|}R}{(16\pi G_N^{4+n})^2},$$

with the Ricci scalar

$$\begin{aligned} R &= R_M^M \\ R_{MN} &= \partial_K \Gamma^K_{MN} - \partial_M \Gamma^K_{NK} + \Gamma^K_{LK}\Gamma^L_{MN} - \Gamma^K_{LN}\Gamma^L_{MK} \\ \Gamma_{KMN} &= \frac{1}{2}\left(\partial_N g_{KM} + \partial_M g_{KN} - \partial_K g_{MN}\right). \end{aligned}$$

The linearized form is then

$$\mathcal{L} = \frac{1}{4}\left(\partial^K H^{MN}\partial_K H_{MN} - \partial^K H^M_M \partial_K H^N_N - 2\partial_K H^{KM}\partial^N H_{NM} + 2\partial_K H^{KM}\partial_M H^N_N\right)$$

and higher-order terms have been neglected.

This Lagrangian is invariant under the coordinate transformation

$$H_{MN} \to H_{MN} + \partial_M \zeta_N + \partial_N \zeta_M, \tag{8.4}$$

for some arbitrary functions ζ_M satisfying $\partial^2\zeta_M = 0$. For simplicity, this gauge freedom will be fixed to the de Donder gauge, imposing

$$\partial^M\left(H_{MN} - \frac{\eta_{MN}}{2}H^L_L\right) = 0, \tag{8.5}$$

and furthermore $H^M_M = 0$. With this, the equation of motion for H_{MN} becomes

$$\partial^2\left(H_{MN} - \frac{\eta_{MN}}{2}H\right) = 0. \tag{8.6}$$

Counting the number of constraint equations yields[3] that only $(3+n)(4+n)/2-1$ degrees of freedom are left unfixed. Simply speaking, there are $4+n$ constraints imposed by the de Donder condition, and further D conditions could be imposed on the ζ functions due to the arbitrariness still left. Thus the number of degrees of freedom for the graviton field

[2]The summation convention for these and other indices will always be over their respective subset only.

[3]Note that no space-time torsion appears in a perturbative treatment and that then $h_{\alpha\beta}$ is symmetric. It must also represent a (classically) massless field.

in $4+n$ dimensions. Hence, in four dimensions there are two, and in five dimensions five, and so on.

This field H_{MN} is then split as

$$H_{MN}=\frac{1}{\sqrt{V_n}}\begin{pmatrix} h_{\mu\nu}+\eta_{\mu\nu}\phi_{ii} & A_{\mu i} \\ A_{\nu j} & 2\phi_{ij}\end{pmatrix}_{MN}, \tag{8.7}$$

where $i,...$ denotes compactified dimensions, and $\mu,...$ the ordinary four space-time dimensions, and V_n is the volume $(2\pi R)^n$ of the compactified dimensions. This yields a redistribution of the degrees of freedom to one spin-2 field h, n four-dimensional (massive) spin-1 vector fields, and $n(n+1)/2$ scalars. Since the additional dimensions are just T^n, the expansion functions are the Fourier functions. This yields

$$\begin{aligned} h_{\mu\nu}(x_\mu,x_i) &= \sum_{\vec{n}} h^{\vec{n}}_{\mu\nu}(x_\mu)\exp\left(\frac{2\pi i n_i x^i}{R}\right) \\ A_{\mu i}(x_\mu,x_i) &= \sum_{\vec{n}} A^{\vec{n}}_{\mu i}(x_\mu)\exp\left(\frac{2\pi i n_i x^i}{R}\right) \\ \phi_{ij}(x_\mu,x_i) &= \sum_{\vec{n}} \phi^{\vec{n}}_{ij}(x_\mu)\exp\left(\frac{2\pi i n_i x^i}{R}\right), \end{aligned}$$

where the vector $\vec{n}$ contains the Fourier mode number in each extra dimension i. Defining the Kaluza-Klein mass of a state as

$$m^2_{\vec{n}}=\frac{4\pi\vec{n}^2}{R^2}$$

and inserting the mode-expanded field (8.7) in the equation of motion (8.6) yields

$$\begin{aligned} (\partial^2+m^2_{\vec{n}})\left(h^{\vec{n}}_{\mu\nu}-\frac{\eta_{\mu\nu}}{2}h^{\rho\vec{n}}_{\rho}\right) &= 0 \\ (\partial^2+m^2_{\vec{n}})A^{\vec{n}}_{\mu i} &= 0 \\ (\partial^2+m^2_{\vec{n}})\phi^{\vec{n}}_{ij} &= 0. \end{aligned}$$

The zero modes with $\vec{n}=\vec{0}$ are massless. They correspond therefore to the graviton, $n(n+1)/2$ massless scalars and n massless gauge bosons. In addition, there are an infinite tower, the so-called Kaluza-Klein tower, of massive spin 2, spin 1, and spin 0 states with masses $m_{\vec{n}}$. Of course, the number of degrees of freedom has not truly become infinite, but it merely appears that the fifth dimension has been traded for this tower.

The effective coupling of matter to this gravitational field can now be directly discussed with these four-dimensional fields. In the present case, the matter fields are only permitted to propagate in the four-dimensional space-time. Their coupling to gravity is therefore minimally given by

$$\int d^n x \int d^4 x \sqrt{\left|\det\left(\eta_{\mu\nu} + 16\pi G_N^{4+n}(h_{\mu\nu} + \eta_{\mu\nu}\phi)\right)\right|}\mathcal{L},$$

where $\phi = \phi_{ii}$ and the details of the standard model particles are encoded in $\mathcal{L}$, and none of them has a dependence on the x_i. It is useful to go to Fourier space. In this case the integral over the n extra dimensions becomes a sum over the Kaluza-Klein modes. In addition, a volume factor for the extra dimensions appears, V^n. This factor can be combined with the (small scale) G_N^{4+n} as $V_n^{-1/2} G_N^{4+n}$ to yield the large scale G_N, the ordinary Newton constant of four-dimensional physics.

Performing then an expansion in $16\pi G_N$ to leading order yields

$$\begin{aligned}
& \sum_{\vec{n}} \int d^4 x \left(1 + 8\pi G_N h + 32\pi G_N \phi\right) \\
& \times \left(\mathcal{L} + (\eta_{\mu\nu} - 16\pi G_N h_{\mu\nu} - 16\pi G_N \eta_{\mu\nu}\phi)\frac{\delta\mathcal{L}}{\delta g_{\mu\nu}}(g_{\mu\nu} = \eta_{\mu\nu})\right) \\
\approx & \sum_{\vec{n}} \int d^4 x \left(\mathcal{L} + \frac{\delta\mathcal{L}}{\delta g_{\mu\nu}}\left(2\eta_{\mu\nu} - 16\pi G_N (h_{\mu\nu} + \phi\eta_{\mu\nu})\right) + \mathcal{L}\left(8\pi G_N h_{\mu\nu} + 32\pi G_N \phi\eta_{\mu\nu}\right)\right) \\
= & \sum_{\vec{n}} \int d^4 x \left(\mathcal{L} + \eta_{\mu\nu}\frac{\delta\mathcal{L}}{\delta g_{\mu\nu}} + 16\pi G_N\left((h_{\mu\nu} + \phi\eta_{\mu\nu})\left(\eta_{\mu\nu}\mathcal{L} - 2\frac{\delta\mathcal{L}}{\delta g_{\mu\nu}}\right)\right)\right),
\end{aligned}$$

using

$$\begin{aligned}
1 &= \frac{1}{4}\eta^{\mu\nu}\eta_{\mu\nu} \\
h &= \eta_{\mu\nu}h^{\mu\nu}.
\end{aligned}$$

In this the energy-momentum tensor

$$T_{\mu\nu} = \left(-\eta_{\mu\nu}\mathcal{L} + 2\frac{\delta\mathcal{L}}{\delta g^{\mu\nu}}(g_{\mu\nu} = \eta_{\mu\nu})\right) \tag{8.8}$$

is recognized, yielding the final result

$$S - 8\pi G_N \sum_{\vec{n}} \int d^4 x \left(h^{\mu\nu}T_{\mu\nu} + \phi T_\mu^\mu\right).$$

Herein S denotes the action of the fields without gravitational interaction, which is obtained from the first two terms and resuming the expansion. The second term gives the coupling to the effective graviton, which is only mediated by the graviton and the trace of the ϕ_{ij}, which is called in this context the dilaton, and in general before radion. Thus, as discussed generally above, the gauge fields decouple in this approximation completely from the dynamics, and also all but effectively one of the scalar fields.

Before proceeding, it is worthwhile to take a look at the physical contents of the theory. The gravitative fields are still depending on the choice of coordinate system inherited from $H_{\alpha\beta}$ and given by the transformation (8.4). If aiming at a description in terms of effective particles, this is rather tedious. In particular, gauge-fixing and the introduction of gauge-fixing degrees of freedom would be necessary. It is therefore useful to define instead physical fields, which are invariant under the coordinate transformations.

Without going into details, the field redefinitions for the Fourier modes[4]

$$\begin{aligned}
\omega_{\mu\nu} &= h_{\mu\nu} - \frac{in_iR}{2\pi\vec{n}^2}\left(\partial_\mu A_{\nu i} + \partial_\nu A_{\mu i}\right) - \left(P_{ij}^T + 3P_{ij}^L\right)\left(\frac{2}{3}\frac{\partial_\mu\partial_\nu}{m_{\vec{n}}^2} - \frac{\eta_{\mu\nu}}{3}\right)\phi_{ij} \\
B_{\mu i} &= P_{ij}^T\left(A_{\mu j} - \frac{in_kR}{\pi\vec{n}^2}\partial_\mu\phi_{jk}\right) \\
\Phi_{ij} &= \sqrt{2}\left(P_{ik}^TP_{jl}^T + \frac{1}{1-n}\left(1-\sqrt{\frac{2+n}{3}}\right)P_{ij}^TP_{kl}^T\right)\phi_{kl} \qquad (8.9) \\
P_{ij}^T &= \delta_{ij} - \frac{n_in_j}{\vec{n}^2} \\
P_{ij}^L &= \frac{n_i}{n_j}\vec{n}^2
\end{aligned}$$

yields fields which are indeed invariant under coordinate transformations. As an example this will be checked for the scalar field. Since the extra dimensions are now compact, also the arbitrary functions ζ_i are expanded in Fourier modes, yielding

$$\zeta_\alpha(x_\mu, x_i) = \sum_{\vec{n}} \zeta_\alpha^{\vec{n}}(x_\mu)\exp\left(\frac{2\pi in_ix^i}{R}\right).$$

The scalar fields transform as

$$\phi_{ij} \to \phi_{ij} + \partial_i\zeta_j + \partial_j\zeta_i. \qquad (8.10)$$

Since the extra dimensions are discrete, the derivatives ∂_i with respect to extra-dimensional coordinates can be replaced after Fourier transformation by n_i. This yields

$$\phi_{ij}^{\vec{n}} \to \phi_{ij}^{\vec{n}} + n_i\zeta_j^{\vec{n}} + n_j\zeta_i^{\vec{n}}.$$

As a consequence, zero modes do not change since for these $n_i = 0$. For non-zero $\vec{n}$, the relation

$$n_iP_{ij}^T = n_i\left(\delta_{ij} - \frac{n_in_j}{\vec{n}^2}\right) = n_j - n_j\frac{\vec{n}^2}{\vec{n}^2} = 0$$

[4] Some care has to be taken for the zero modes.

holds. Thus, inserting the transformed field (8.10) into (8.9), the contribution from $n_i\zeta_j$ drop out, confirming that Φ_{ij} is invariant under coordinate transformations. The arguments for B_μ and $\omega_{\mu\nu}$ are similar, though more lengthy. In particular, also the zero modes of both fields are invariant without redefinition. Hence, a replacement is only necessary for non-zero modes.

This is simple for the ϕ field, since only its trace appears. Tracing the expression (8.9) yields

$$P^T_{ij}\phi^{\vec{n}}_{ji} = \frac{3}{2}\sqrt{\frac{2}{3(n+2)}}\Phi_{ii}.$$

The expression P^T contains a Kronecker-δ, yielding the trace of ϕ. The expressions n_i are just derivatives in the compact dimensions. By partial integration, and using that $T_{\mu\nu}$ is a conserved quantity, these do not contribute to the integral[5]. Thus, up to the pre-factor, ϕ can be replaced by Φ in the Lagrangian.

For the contribution from $h_{\mu\nu}$, it should first be noted that the contribution involving the A_μ are proportional to n_i, which is effectively a derivative once more and thus can be dropped. The term involving the ϕ is again either a derivative, which also vanishes, and terms containing either another δ_{ij} or n_i. Then, only the trace of δ_{ij} thus remains, multiplied with $\eta_{\mu\nu}$. But this just implies a further contribution to the $T^\mu_\mu\phi$ term. After once more replacing the ϕ with the Φ and sorting the pre-factors, the final Lagrangian in terms of the physical fields is

$$\sum_{\vec{n}}\left(\mathcal{L} + \eta_{\mu\nu}\frac{\delta\mathcal{L}}{\delta g_{\mu\nu}} + 8\pi G_N\left((\omega_{\mu\nu} + \xi^{\vec{n}}\Phi\eta_{\mu\nu})\left(\eta_{\mu\nu}\mathcal{L} - 2\frac{\delta\mathcal{L}}{\delta g_{\mu\nu}}\right)\right)\right). \tag{8.11}$$

where

$$\xi^{\vec{n}} = \begin{cases} 1 & \text{for} \quad \vec{n} = \vec{0} \\ \sqrt{\frac{2}{3(n+2)}} & \text{else} \end{cases}.$$

Thus, the only remaining ingredient is to specify the matter system to which the theory is coupled to and determine the energy-momentum tensor.

The (symmetrized) energy-momentum tensor for a theory of a gauge-field C_μ, a scalar

[5] Here it has been used that the compact dimension is just a torus. For more complicated spaces, possibly with non-trivial boundary conditions, the vanishing of boundary terms has to be checked explicitly.

Δ and a fermion ψ is given by[6]

$$\begin{aligned}
T_{\mu\nu} &= (-\eta_{\mu\nu}\eta^{\rho\sigma}+\eta^{\mu\rho}\eta_{\nu\sigma}+\eta^{\mu\sigma}\eta_{\nu\rho})(D_\rho\Delta)^\dagger D_\sigma\Delta+\eta_{\mu\nu}m_\Delta^2\Delta^\dagger\Delta\\
&+\frac{1}{4}\eta_{\mu\nu}F^a_{\rho\sigma}F^{a\rho\sigma}-F^{a\rho}_\mu F^a_{\nu\rho}\\
&-\eta_{\mu\nu}\left(\bar\psi i\gamma^\rho D_\rho\psi-m_\psi\bar\psi\psi+\frac{i}{2}\partial_\rho(\bar\psi\gamma^\rho\psi)\right)+\frac{i}{2}\bar\psi(\gamma_\mu D_\nu+\gamma_\nu D_\mu)\psi\\
&-\frac{i}{4}(\partial_\mu\bar\psi\gamma_\nu\psi+\partial_\nu\bar\psi\gamma_\mu\psi)\\
D_\mu &= \partial_\mu+igC^a_\mu\tau^a\\
F^a_{\mu\nu} &= \partial_\mu A^a_\nu-\partial_\nu A^a_\mu+gf^{abc}A^b_\mu A^c_\nu
\end{aligned}$$

where τ^a are the generators of the gauge group. Inserting, e. g., the scalar sector's energy-momentum tensor into (8.11) yields the description of the interaction of the scalar with the gravitational field for a mode $\vec{n}$ as

$$\begin{aligned}
\frac{1}{16\pi G_N}\mathcal{L}^{\vec n} &= -\left(\omega^{\vec n}_{\mu\nu}-\frac{1}{2}\eta_{\mu\nu}\omega^{\vec n\rho}_\rho\right)(D^\mu\Delta)^\dagger D^\nu\Delta-\frac{1}{2}h^{\vec n\mu}_\mu m_\Delta^2\Delta^\dagger\Delta\\
&+\xi^{\vec n}\Phi^{\vec n}\left((D_\mu\Delta)^\dagger D^\mu\Delta-2m_\Delta^2\Delta^\dagger\Delta\right),
\end{aligned}$$

and similarly for the gauge and fermion sector. With standard methods, it is possible to obtain Feynman rules and then calculate the influence of the additional particles to cross-sections. The generic features of such contributions will be discussed next.

For example, at tree-level the decay of a non-zero Kaluza-Klein mode of the graviton[7] to two massless gauge bosons is just obtained from the tree-level coupling. The calculations yield straightforwardly a decay width of

$$\Gamma_{\omega\to\gamma\gamma}=N_c\frac{(16\pi G_N m_\omega^{\vec n 2})}{160\pi}.$$

To obtain numerical values, it is necessary to specify G_N further. Generically, this $4+n$-dimensional Planck constant is given by the combination

$$\frac{1}{G_N}=M_s^{n+2}R^n,$$

and $M_s\sim 1/G_N^{4+n}$ is the intrinsic scale of the process causing the extra dimensions to be compactified, e. g., the scale of the string theory. This scale can then be rather low, if the compactification radius is sufficiently large. Multiplying with the n-dimensional volume

[6]The derivation, and why it has to be symmetrized, is a rather lengthy discussion, and can be found in most texts on relativistic field theory, and thus will be skipped here.

[7]Note that the couplings to the zero-modes is generically suppressed by the gravitational coupling.

factor is what makes from the small scale M_s a large scale G_N perceived in four dimensions. This also sets limits for the size of extra dimensions if M_s is fixed. Setting, e. g., M_s to about 1 TeV, the size of R varies between 10^{-4} eV (about a mm) for two additional extra dimensions[8] to a couple of hundred MeVs at $n = 6$ or 7, which is the number of additional dimensions suggested by string theories.

Entering its value of about 2.4×10^{18} GeV gives for the decay into two photons ($N_c = 1$) a life-time of

$$\tau_{\omega\to\gamma\gamma} \approx 3\times10^9 \left(\frac{100\text{ MeV}}{m_\omega^{\vec{n}}}\right)^3 \text{ years.}$$

Since $(m_\omega^{\vec{n}})^2 = 4\pi\vec{n}^2/R^2$ it now depends on the size of the additional dimensions for the final result. R compatible with precision measurements of small-distance gravity are of the size of eV to much larger scales, making a life-time of larger than the age of the universe easily possible, and thus the Kaluza-Klein state essentially stable. This makes it then also a viable dark-matter candidate.

A corresponding decay to gluons requires a follow-up hadronization, and therefore corresponds to at least a decay into two pions. Thus, this decay channel only opens up for masses starting at a few hundred of MeVs of ω. If the mass becomes even larger, there are also alternative couplings for real decays possible. First follow decays to light quarks and leptons, and then finally to heavy quarks and electroweak gauge bosons and finally to the Higgs. This permits a decrease of the life-time down to fractions of a year, but, very generically, the particle is still stable on collider time-scales, if not the compactification radius becomes very small.

There is an additional interesting possibility. The masses of the Kaluza-Klein tower of states is evenly spaced. Thus even, if the mass of the lowest state is small, say a couple of MeV, a highly excited Kaluza-Klein state could decay to it under the emission of a ladder of particles with energies of the order of the splitting. This could, under certain kinematic conditions, give a quite interesting signature of a shower of particles and a final missing energy at the endpoint of the shower in a collider.

The situation for the dilaton Φ is somewhat different. Since it couples to the trace of the energy-momentum tensor, it turns out not to have a tree-level coupling to gauge bosons. Thus it cannot, as the graviton, decay into two photons, and thus would be absolutely stable if light enough. If somewhat heavier, it could decay into two light leptons or quarks, but would have a very long life-time, as this becomes suppressed as $1/(m_f^2 m_\Phi)$ due to kinematics, with m_f the fermion mass. Thus, the decay to neutrinos is negligible, which would be the only real decay channel mandatory open by the maximum size of R

[8] One additional extra dimension gives a size significantly larger than a mm, which is excluded by experiments.

for any reasonable number of extra dimensions in accordance with experiments. If R is then sufficiently large, but not large enough to permit decay into two light quarks (and consecutively to two pions), the dilaton is essentially stable on the time-scale of the age of the universe, making it another dark matter candidate.

Another interesting effect of the presence of the Kaluza-Klein tower of states is the appearance of effective four-fermion couplings. For example, if four fermions couple by the exchange of a state $\vec{n}$ dilaton[9], the corresponding tree-level matrix element is given by

$$\mathcal{M} = -\frac{n-1}{n+2}\frac{4\pi C_4}{3} m_{f_1} m_{f_2} \bar{f}_2 f_2 \bar{f}_1 f_1. \tag{8.12}$$

The function C_4 encodes the details of the exchanged dilaton, and reads

$$C_4 = \frac{(16\pi G_N)^2}{8\pi} D(q^2, \vec{n}) = \frac{(16\pi G_N)^2}{8\pi} \frac{1}{q^2 - m_\Phi^{\vec{n}2} + i\epsilon},$$

where q^2 is the exchanged momentum. The problem is now that there is not only one possible exchange but instead an infinite tower of Kaluza-Klein states can be exchanged. Hence, the total amplitude is given by a sum over $\vec{n}$. This is particularly problematic, as in most cases the level spacing of Kaluza-Klein states are very narrow, and thus the corresponding masses are quite similar, given similar contributions to C_4, in particular if q^2 is much larger than $1/R^2$.

In fact, for the purpose of observing Kaluza-Klein states at a collider like the LHC the exchanged four-momentum q^2 can be safely taken to be much larger than $1/R^2$, if the string scale M_S should be at the TeV scale, and the number of extra dimensions be small, not more than ten. Then the level spacing is of order of (a couple of) MeV, while q^2 is deep in the GeV or higher range. It is then a rather good approximation to instead of performing a sum over all states to do an integration, i. e.,

$$D(q^2) = \sum_{\vec{n}} D(q^2, \vec{n}) \to \int dm_{\vec{n}}^2 \rho(m_{\vec{n}}^2) D(q^2, \vec{n})$$

with the density of Kaluza-Klein states

$$\rho(m_{\vec{n}}^2) = \frac{R^n m_{\vec{n}}^2}{(4\pi)^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)},$$

which is just the level-density for an n-dimensional sphere for states spaced as $\vec{n}^2$. The problem is now that the integral will diverge with this level density. Thus it requires regularization, and in principle renormalization. A common assumption is once more

[9] In principle, this interferes with the exchange of a graviton and standard model processes. These are neglected for the sake of simplicity, and do not change generically the result.

that the compactification is due to an underlying string theory. This can be most easily modeled by an explicit upper cutoff by the string scale M_s^2, and thus at the level of TeV.

The integral can then be performed, yielding

$$\begin{aligned} D(q^2) &= \frac{(q^2)^{\frac{n}{2}-1}R^n}{\Gamma\left(\frac{n}{2}\right)(4\pi)^{\frac{n}{2}}}\left(\pi - 2iI\left(\frac{M_s}{\sqrt{q^2}}\right)\right) \\ I(x) &= \begin{cases} \frac{1}{2}\ln(x-1) + \sum_{k=1}^{\frac{n}{2}-1}\frac{x^{2k}}{2k} & n \text{ even} \\ -\frac{1}{2}\ln\frac{\sqrt{x}+1}{\sqrt{x}-1} + \sum_{k=1}^{\frac{n-1}{2}}\frac{x^{2k-1}}{2k-1} & n \text{ odd} \end{cases}. \end{aligned}$$

The real part comes from resonant production of real Kaluza-Klein states, while the imaginary part stems from the continuum of other states. If M_S is large compared to q^2, which occurs at the LHC if the string scale is several TeV, the expression can be approximated by

$$D(q^2) \approx -iR^n \begin{cases} \frac{\ln\frac{M_s^2}{q^2}}{4\pi} & n = 2 \\ \frac{2}{(n-2)\Gamma\left(\frac{n}{2}\right)}\frac{M_s^{n-2}}{(4\pi)^{\frac{n}{2}}} & n > 2 \end{cases}.$$

Thus, the effectively induced four-fermion coupling is almost energy-independent, and looks like a contact interaction. This will give rise to corrections to the standard model processes. Combining the expression for $D(q^2)$ with the original matrix element (8.12) show that these additional interactions scale as $1/M_S^4$, and thus are strongly suppressed. However, if the large extra dimensions would not be present, the corresponding corrections due to gravity would be suppressed by the Planck mass instead, and therefore effectively irrelevant. The presence of the larger extra dimensions amplifies the effect of gravity in this case by sixty orders of magnitude. Thus, looking for signatures of this type has been done at experiments, in particular in two-to-two fermion scattering processes, providing further constraints on the presence of extra dimensions[10]. Similar calculations can be done for other processes, like the scattering of fermions to weak gauge bosons with their subsequent decays, and similar corrections arise.

A serious problem arises when the universally coupling Kaluza-Klein modes show up in processes forbidden, or strongly suppressed, in the standard model, like proton decay. The standard model limit for proton decay by an effective four fermion vertex is about 10^{15} GeV, thus much larger than the comparable effect from the larger extra dimensions if M_s should be of order TeV. Thus this leads to a contradiction if not either M_S is again set very large (or the number of dimensions n), and thus large extra dimensions become once more undetectable, or additional custodial physics is added to this simple setup. This usually leads, like in the case of technicolor, to rather complex setups.

[10] Of course, such corrections appear generically in almost all theories, see e. g. technicolor, and thus measuring them provides immediately constraints on many theories simultaneously.

8.1.3 Black holes

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

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

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

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

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

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

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

8.2 Universal extra dimensions

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

When compactifying the additional dimensions in such theories care has to be taken when imposing the boundary conditions. The reason is that fermions in a box with anti-periodic boundary conditions will develop an effective mass of order $1/L$, where L is the compactification scale. This is the same process as occurs at finite temperature, and is due to the fact that only odd frequencies in a Fourier-expansion of a fermion field have the right periodicity, $(2n+1)L$, instead of $2nL$ as for bosons, as required by the spin 1/2 nature of fermions. Therefore, chiral boundary conditions are required. The mass-spectrum of all standard model particles for a compactification along a single extra dimension with open (chiral) boundary conditions, a so-called orbifold, is then given by

$$M_j^2 = \frac{\pi^2 j^2}{L^2} + m_0^2,$$

where m_0 is the mass of the standard model particle, and its Kaluza-Klein excitations have mass M_j, and j counts the excitation.

The advantage of such universal extra dimensions is that they can provide a natural way of explaining the (flavor) hierarchies of the standard model by localizing the fermions on branes inside the bulk instead of the standard model brane. This idea will be repeated similarly in section 8.3 for warped extra dimensions. Here, it is sufficient to have a look at the action of a fermion propagating in the bulk described by the action

$$S = \int d^4xdy\bar{\psi}(i\Gamma^\mu\partial_\mu + i\Gamma^5\partial_y + \phi(y))\psi$$

where the Γ_μ denote the 4×4 five-dimensional version of the Dirac γ matrices[11],

$$\begin{aligned}
\Gamma_{0\ldots3} &= \gamma_{0\ldots3} = \begin{pmatrix} 0 & \sigma_{0\ldots3} \\ \bar{\sigma}_{0\ldots3} & 0 \end{pmatrix} \\
\Gamma_5 &= \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} \\
\sigma &= (1,\vec{\sigma}) \\
\bar{\sigma} &= (1,-\vec{\sigma}).
\end{aligned}$$

The field $\phi(y)$ denotes the brane, and its interaction with the fermions will localize them on the brane.

The idea is now to separate the fermion field as

$$\begin{aligned}
\psi(x,y) &= \sum_n (f_L(n,y)\psi_L(n,x) + f_R(n,y)\psi_R(n,x)) \\
i\Gamma_5\psi_L &= \psi_L \\
i\Gamma_5\psi_R &= -\psi_R
\end{aligned}$$

[11]In an odd number d of dimensions there exist two possible inequivalent representations of the Dirac algebra, one is $d-1$-dimensional, as chosen here, and one is $d+1$-dimensional. The latter is not lending itself easily for the purpose of obtaining the standard model on a brane.

which implies that ψ_L and ψ_R are left-handed and right-handed fermions with respect to four dimensions. A simple version of a brane is given by $\phi(y) = 2\mu^2 y$. It is then possible to get such a decomposition if $f_{L,R}$ are chosen as

$$\begin{aligned} f_L(n,y) &= (-\partial_y + \phi(y))^n f_L(0,y) \\ f_R(n,y) &= \frac{1}{\mu}(\partial_y + \phi(y)) f_L(n,y) \\ f_{L,R}(0,y) &= \sqrt{\frac{\mu}{\sqrt{\frac{\pi}{2}}}} \exp\left(\mp \int_0^y \phi(y')dy'\right). \end{aligned}$$

To have a normalizable mode, $f_{L,R}(0,\infty)$ must be finite, leaving only the left-handed solutions for now. Right-handed fermions thus require localization on a different brane, e. g. with $\phi_R = -\phi$. The remaining left-handed zero-mode of the fermion is thus exponentially localized at $y = 0$ due to this pre-factor function. Entering this expression into the Dirac equation shows that the zero-mode is furthermore massless. Choosing other functions permits to have fermions localized at different values of y. The non-zero modes have, as usual, a large mass of order the inverse size of the dimension, and are thus not (yet) observable.

The advantage is then the following. Assume that (only) the Higgs field $h(x)$ is not propagating into the extra dimension. Furthermore, take the right-handed fermions to be localized at a different position $y = r$. This setting is called split fermions. The standard-model Yukawa coupling then reads with a coupling matrix C_5

$$\begin{aligned} &\int d^4x dy (h(x)\psi_L^T(x,y) C_5 \psi_R(x,y) + \text{h.c.}) \\ \overset{j=0}{=} &\int d^4x h(x)\psi_L(0,x)\psi_R(0,x) \int dy f_L(0,y) f_R(0,y) = e^{-\frac{\mu^2 r^2}{2}} \int d^4x h(x)\psi_L(0,x)\psi_R(0,x). \end{aligned}$$

Thus the Yukawa coupling is exponentially suppressed if the fields are sufficiently far (but not exponentially so) separated, and thus give a natural explanation for the large mass hierarchies observed in the standard model, if the different flavors are located on different branes inside the bulk. Also, e. g., graviton-mediated proton decay, which has been a challenge for non-universal extra dimensions, is reduced exponentially by the reduced overlap with the standard-model brane. To prevent that the other standard-model interactions suffer a similar fate requires them to propagate also in the bulk, or requires other amendments.

As has already been encountered when discussing the sum-of-states for the explicit example of gravity-exclusive large extra dimensions, the higher-dimensional theories are usually not renormalizable prior to compactification. Furthermore, because compactification explicitly breaks the Lorentz invariance of the $4+n$-dimensional theory, boundary-terms appear which are usually also divergent. Both facts are usually taken to be an indication for these theories to be also only low-energy effective theories of, e. g., a string theory.

The problem of divergent boundary terms can be reduced by imposing boundary conditions such that this effect is minimized. As a consequence of these terms and their compensation usually states of different mass can mix. However, in general arbitrary mixing is not possible. In five-dimensional theories of this type the Kaluza-Klein states j acquire a conserved quantum number $(-1)^j$. Thus, a state with the lowest Kaluza-Klein mass with $j = 1$ cannot decay in a state with $j = 0$, and thus standard-model particles. As a consequence, such states provide dark matter candidates. This is especially attractive, as a compactification radius in such models of about $(1 \text{ TeV})^{-1}$ is well possible, giving such particles a mass of roughly the same size and making them therefore accessible at accelerator-based experiments.

8.3 Warped extra dimensions

In models with warped extra dimensions, also known as Randall-Sundrum models, the additional dimensions have an intrinsic curvature k in such a way that the energy scales depend exponentially on the separation Δy of two points in the additional dimensions, $\exp(-2\Delta yk)$. By positioning different effects at different relative positions, large scale differences can appear naturally, e. g., $M_H \sim \exp(-\Delta yk)M_P$. In particular, the different Yukawa couplings for the standard model fermions can be explained by having different wave functions for different fermion species in the additional dimension, which then have different overlap with the Higgs wave function, therefore permitting very different couplings to the Higgs, even if the difference is of order unity in a flat space. This is very similar to the concept of split fermions in the case of universal extra dimensions in section 8.2.

8.3.1 Minimal model

In the minimal version of warped extra dimensions there is only one additional dimension. This one is orbifolded, i. e., it is compactified on a radius πR with opposite points identified[12], giving the additional coordinate y the range from 0 to πR. The invariant length element is then

$$ds^2 = g_{MN}dX^M dX^N = e^{-2k|y|}\eta_{\mu\nu}dx^\mu dx^\nu - dy^2.$$

Taking the absolute value of y is necessary, because y can take also negative values to $-\pi R$, which are then identified with the original ones by the absolute value. This con-

[12]Topologically, this is S^1/Z_2.

struction is necessary to permit this metric to be a solution of the Einstein equations of the five-dimensional space. Indeed, such a space is obtained from an anti-de Sitter space with a cosmological constant. The details of the construction are not entirely trivial. In particular, the cosmological constant necessary to warp the extra dimension sufficiently strongly such that their size is compatible with measurements has to be almost canceled in the four-dimensional space to obtain one in agreement with experiment.

The fifth dimension is bounded by two end-points, which are four-dimensional. These two end-points are called branes. Due to the explicit exponential, both are not the same, but differ by a metric factor of $\exp(2k\pi R)$. One of the branes is identified with the present four-dimensional world. Its Planck mass is then related to the Planck mass of the bulk M_g, i. e., inside the total volume, by

$$M_P^2 = \frac{M_g^3}{k}\left(1 - e^{-2\pi kR}\right).$$

A natural size for k is about M_g itself, and it is also natural to have $kR \gtrsim 1$. Then the Planck mass and the bulk Planck mass are again of the same magnitude, despite that otherwise only natural scales appear.

Why there is nonetheless no discrepancy between the electroweak and the Planck scale is explained thus differently in such models than in the large extra dimensional models beforehand. Take the brane at $y = \pi R$ to be our world. Assume that the Higgs H is confined to this brane. It is then described by the action

$$\begin{aligned} S &= \int d^4x\sqrt{|g(y=\pi R)|}(g_{\mu\nu}(y=\pi R)(D^\mu H)^+(D^\nu H) - \lambda(HH^+ - V^2)^2) \\ &= \int d^4x(\eta_{\mu\nu}(D^\mu H)^+(D^\nu H) - \lambda(HH^+ - e^{-2\pi kR}V^2)^2), \end{aligned}$$

where in the second step the Higgs field has been rescaled by $H \to \exp(\pi kR)H$, to remove the exponential from the four-dimensional induced metric. As a consequence, the expectation value of the Higgs is $\langle H\rangle = \exp(-\pi kR)V = v$, and by this a quantity naturally of the same scale as the Planck mass is scaled down to the much smaller electroweak scale by the exponential pre-factor. To have the correct numbers for a V of the size of the Planck scale $kR \approx 11$ is needed. This solves the hierarchy problem, or actually makes it nonexistent. Such a value of kR can be obtained if a radion field, as part of the graviton field, acts in the bulk. For the reason that V is scaled down to v, our brane at $y = \pi R$ is usually called the infrared brane, in distinction to the ultraviolet brane at $y = 0$.

A further interesting distinction to the case of large extra dimensions is in the induced additional particle content. For large extra dimensions, the Kaluza-Klein states form almost a continuum. Here, this is not the case. After separating the graviton field, as in the case of large extra dimensions, in a four-dimensional graviton, the radion, and further states which do not couple to the standard model field, these can be Fourier expanded in the extra dimension as

$$h_{\mu\nu}(x,y)=\sum_{n=0}^{\infty}h^n_{\mu\nu}(x)g_n(y).$$

However, the base functions this time are not the ordinary Fourier functions $\exp(ik_nr)$, but more complex functions due to the warped geometry. The associated masses of the Kaluza-Klein gravitons are then

$$m^n_{KK}=x^G_nke^{-\pi kR},$$

where x^G_n are rather well approximated by the zeros of the Bessel function for $n>0$, thus 3.8, 7.0, 10,... going to $n\pi$ for large n. Depending on the precise size of k and kR, the lightest excitation has mass of size a few TeV, and thus the level spacing is of similar order.

Due to the warping, also the coupling is modified compared to the large-extra dimension case (8.3), with an effective Lagrangian

$$\mathcal{L}=-\frac{1}{M_P}T^{\mu\nu}h^{(0)}_{\mu\nu}-\frac{e^{\pi kR}}{M_P}T^{\mu\nu}\sum_{n>0}h^{(n)}_{\mu\nu}.$$

Hence, only the ordinary graviton couples weakly to matter, while the Kaluza-Klein gravitons couple at the TeV scale, and could therefore be much more easily observed.

8.3.2 Extra-dimensional propagation of standard model particles

So far, all the standard model fields have been restricted to the infrared brane. Permitting further particles to also propagate in the fifth dimensions requires some subtle changes.

Gauge fields will then have five components, instead of four. Furthermore, it is necessary to specify boundary conditions. Usually on either brane Dirichlet boundary conditions $A_5=0$ or von Neumann conditions $\partial_5A_\mu=0$ on the branes are imposed, even in mixed form. That these two boundary conditions are the most important ones can be seen by the example of a scalar field Φ. The associated current along the extra dimension J_5 is given as usual by

$$J_5=i\Phi^\dagger\partial_5\Phi$$

However, particles should not vanish or be created at the boundaries of the extra dimension. This is prevented by imposing either of the two boundary conditions, since then the current automatically vanishes. Note that imposing the boundary conditions corresponds to require the fields to be either odd (Dirichlet) or even (Neumann) under the transformation $y \to -y$.

Returning to the gauge field, choosing an appropriate gauge and Dirichlet boundary conditions make the fifth component vanish altogether. The remaining gauge field can then be decomposed into Kaluza-Klein modes. As they are spin one instead of spin two particles, the mass spectrum is slightly different then for the gravitons and given by

$$m_{KK}^n = x_n^A k e^{-\pi k R},$$

with x_n^A for $n > 0$ being 2.5, 5.6, 8.0, and moving also towards $n\pi$ for large values of n. Physically, the absence of the fifth component of the field can then be interpreted as that this component provides the necessary longitudinal degree of freedom for the massive Kaluza-Klein gauge bosons. Unfortunately, as in the graviton case, the Kaluza-Klein modes couple enhanced by a factor $k \exp(\pi k R)$ to the standard model particles as in case of the graviton. This can only be avoided at the cost of having the geometry such that all new physics is moved to rather large energies, reintroducing the hierarchy problem, or by rather subtle manipulations on the kinetic terms of the gauge bosons on the ultraviolet brane.

This changes, if also the fermions can propagate into the bulk. However, this is again complicated by the chiral nature of fermions. As noted, chirality of five-dimensional fermions is fundamentally different from the one of four-dimensional ones. This can be remedied by introducing a second set of fermions with opposite chirality of the standard model ones. To avoid that all of them are visible, it is necessary to give them different boundary conditions. Only fields with von Neumann boundary conditions on both branes are found to have (up to the Higgs effect) massless modes, and can therefore represent the standard model fermions. Fermions with Dirichlet boundary conditions on at least one boundary immediately acquire a Kaluza-Klein mass. Therefore, they will not be visible below the TeV scale.

There is another twist to this. On top of the Kaluza-Klein and Higgs mass, there is usually also a bulk mass of order ck for some constant c. This can be counteracted by choosing the mixed boundary conditions

$$(\partial_y + c_L k)\psi_L = 0 \tag{8.13}$$

on the branes at $y = 0$ and $y = kR$, for the desired left-handed fermion for the standard model. As a consequence, the five-dimensional Fourier mode of the solution of the Dirac equation still has a zero-energy/zero-mass mode. Furthermore, in the fifth dimension this implies a behavior of the fermion field as

$$\psi \sim e^{-\left(c_L-\frac{1}{2}\right)k|y|}.$$

Hence, the field is exponentially localized towards either of the branes, depending on the precise value of c_L. A similar calculation for the right-handed fermions in the standard model yields that $c_L - 1/2$ is replaced by $c_R + 1/2$. The masses of the Fourier-expansion, and thus the Kaluza-Klein modes, is then given by

$$m_{KK}^n \approx \pi \left(n + \frac{1}{2}\left(\left|c_{L,R} \mp \frac{1}{2}\right| - 1\right) - \frac{(-1)^n}{4}\right) ke^{-\pi kR} \tag{8.14}$$

for $n > 0$ and zero for $n = 0$. Since the gauge bosons have no such localization due to their boundary conditions, they will couple to all these fields equally. The exponential localization outside the standard-model brane then provides that from a four-dimensional perspective the interaction of fermions and gauge bosons is not appearing enhanced.

Since the Higgs boson is (yet) localized to a brane, the effective overlap of a fermion field, and thus its interaction strength, is strongly determined by how much it is localized on the brane. This is exponentially controlled by the parameters c_i. Thus, even very small differences in the c_i can yield huge effects, and thus naturally explain why the different masses of the fermions generated by the Higgs-Yukawa couplings are so very large without requiring the couplings to be actually very different.

Still, such scenarios require quite a number of amendments, like extra symmetries or particles, to make them compatible not only qualitatively but also quantitatively with experimental precision measurements.

8.3.3 Symmetry breaking by orbifolds

With the orbifolds it is also possible to provide symmetry breaking. Take for example the SU(5) GUT of section 7.2. In this case it was necessary to introduce numerous additional Higgs fields to remove the additional gauge bosons, acting as leptoquarks, from the spectrum to have a decent proton life time. This can also be achieved by orbifolded extra dimensions. Take for example a single extra dimensions with boundary conditions.

To see this note first that Dirichlet boundary conditions generate an expansion for a field of type

$$\phi(x,y) = \sqrt{\frac{2}{\pi R}} \sum_{n=1}^{\infty} \phi(n,x) \sin\frac{ny}{R},$$

while Neumann boundary conditions lead to

$$\phi(x,y) = \frac{1}{\sqrt{\pi R}}\phi(0,x) + \sqrt{\frac{2}{\pi R}} \sum_{n=1}^{\infty} \phi(n,x) \cos\frac{ny}{R}.$$

Hence, only for Neumann boundary conditions a zero-mode with zero Kaluza-Klein mass exist, while the lightest excitation in the Dirichlet case has a mass $1/R$. Finally in the case of a mixed boundary condition, i. e., Dirichlet at one end and Neumann at the other, again a zero-mode is forbidden, and the period is halved. E. g., Dirichlet conditions at $y = 0$ and Neumann conditions at $y = \pi R$ yields

$$\phi(x, y) = \sqrt{\frac{2}{\pi R}} \sum_{n=1}^{\infty} \phi(n, x) \sin \frac{ny}{2R},$$

and vice verse.

Now take the SU(5) gauge field. It is splitted into the standard model fields of gluons, W, Z, and the photon. Furthermore the X and Y gauge fields appear. In addition, universal extra dimensions dictate to have a further fifth component for all gauge fields. The fifth dimension is different then the ordinary four dimensions by its different structure. In addition to ordinary parity transformations in four dimensions, there is then also an exclusive five-dimensional parity transformation $y \to -y$. The fifth component of the gauge field must then have opposite parity under $y \to -y$. This can be seen from the fact that ∂_y is necessarily odd under $y \to -y$. The component $F_{\mu 5} = \partial_\mu A_5 - \partial_5 A_\mu$ of the field-strength tensor must have a definite parity under this transformation, or otherwise the theory would not respect the orbifold structure of the theory. Thus, A_μ and A_5 have opposite boundary conditions.

Choosing then Neumann boundary conditions for the standard model fields automatically makes their fifth component having Dirichlet boundary conditions, making them heavy. To also remove the X and Y gauge bosons together with their fifth component from the low-energy realm requires them to have mixed boundary conditions. By this, the GUT symmetry appears to be explicitly broken since all additional fields have become massive. This is the concept of symmetry breaking by orbifolding.

8.4 Deconstructed extra dimensions

An alternative flavor of (large) extra dimensions are obtained from so-called deconstructed extra dimensions. In this case the extra dimensions are not continuous, but are discrete, i. e., contain only a finite number of points, like a lattice. This removes the ultraviolet divergences encountered by having an infinite number of Kaluza-Klein states, making the theory renormalizable. This can also be viewed by a finite, in case of the extra dimension being compactified, or infinite set of four-dimensional space-times, which are distinguished by a discrete quantum number.

As an example, take only one additional dimension, with N points and radius R. Then each of the N points is a complete four-dimensional space-time, and is also called a brane. Take now a gauge-field $A_\mu(x, y)$ with y the (discrete) fifth coordinate. On a given brane, the fifth coordinate is fixed and denotes the brane. There are then four gauge-field components depending on the remaining four coordinates x, just as a normal gauge field would. This can, e. g., give the gluons of QCD. There is another field, the fifth component of the gauge field, depending on a fixed brane again on the four coordinates. It can be shown to behave like an adjoint Higgs field.

Expanding the gauge field in a discrete Fourier series shows the presence of further, heavier Kaluza-Klein modes as copies of these fields. From a low-energy perspective, like an experiment, these appear in addition to the gauge theory described by the zero modes as $N - 1$ copies of these gauge theory, which are broken by the additional $N - 1$ adjoint Higgs fields, giving the Kaluza-Klein modes of the gauge fields their mass. The remaining zero-mode of the A_5 component can be rearranged such that it can take the role of the standard model Higgs, breaking the electroweak symmetry[13].

Similarly, it is possible to introduce fermions having the correct chiral properties by choosing appropriate boundary conditions, as before[14]. As a bonus, tuning the parameters appropriately, it is possible to make Kaluza-Klein fermions condense, essentially realizing a topcolor mechanism, and thus providing the mechanism unspecified in topcolor theories.

[13] Theories exploiting the same mechanism to obtain the standard model Higgs for a continuous extra dimensions are sometimes called gauge-Higgs unifying theories or also holographic Higgs theories.

[14] In fact, the domain-wall fermions of lattice gauge theory are a very similar concept to a deconstructed theory. However, in this case the limit $R \to 0$ is taken, making all Kaluza-Klein modes infinitely heavy in the end.

Chapter 9

Quantum gravity

9.1 String theory

9.2 Non-commutative geometry

One further possibility to quantize gravity is to postulate the existence of a minimal length, similar to the postulate of a minimal phase space volume $\Delta x \Delta p \sim \hbar$ in ordinary quantum mechanics. This is also similar to the idea of a maximum speed in general relativity. As there, the existence of such a minimal length, which is typically of the order of the Planck length 10^{-20} fm, has profound consequences for the structure of space-time. Especially, coordinate operators do no longer commute, just like coordinate and momenta do not commute in quantum mechanics, i. e. $[X_i, X_j] \neq 0$.

The same effect can be reached by postulating canonical commutation relations for coordinates, in addition to the ones between coordinates and momenta. Thus, this ansatz is called non-commutative geometry. Since there is a minimal length, there is also a maximal energy, and hence all quantities become inherently finite, and renormalization is no longer necessary. On the downside of this approach, besides an enormous increase in technical complexity, is that in general relativity neither coordinates nor energies themselves are any longer physical entities, like in special relativity or in quantum (field) theories. Thus, the precise physical interpretation of a non-commutative geometry is not entirely clear. Furthermore, so far it was not possible to establish a non-commutative theory which, in a satisfactory manner, provides a low-energy limit sufficiently similar to the standard model. Particularly cumbersome is that it is very hard to separate the ultraviolet regime where the non-commutativity becomes manifest and the infrared, where the coordinates should again effectively commute. This problem is known as IR-UV mixing.

9.3 Loop quantum gravity

In contrast to asymptotic safety in section 7.6, loop quantum gravity goes a step further, and postulates that quantum gravity cannot be canonically quantized. Rather, different variables need to be used for quantization. Especially, the basic requirement is that the degrees of freedom in the path integral to be integrated over are diffeomorphism, i. e. coordinate-transformation, invariant.

This avoids many conceptually tricky problems, which are similar to those arising in (non-)Abelian gauge theories. In fact, a similar reformulation exists also for ordinary non-Abelian gauge theories, and thus it appears in principle possible. In the latter case, the gauge-invariant degrees of freedom are so-called Wilson loops, exponentiated line-integrals over gauge-fields. In the same way the new variables are loop integrals over the metric, and thus the name. However, the downside is that the ensuing theory is much more involved, and contains a substantial, probably infinite, number of degrees of freedom and potential non-localities. This makes work with this theory, even at the perturbative level, very much more involved. In particular, it may even be only possible in a genuine non-perturbative way.

9.4 Supergravity

The second important gauge theory, besides Yang-Mills theory with or without matter, is gravity. Gravity can be considered as a gauge theory for translations. Therefore, local supersymmetry will therefore create gravity. Without going into too much details, especially as many questions on off-shell supergravity have not been solved, here only a short introduction is made.

9.4.1 General relativity

Before talking about gravity in a particle physics setup, it seems appropriate to quickly repeat the basics of classical general relativity.

The basic dynamical variable is the metric, which describes the the invariant length element ds by

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu.$$

The inverse of the metric is given by the contravariant tensor

$$g^{\mu\nu} g_{\nu\lambda} = \delta^\mu_\lambda.$$

As a consequence, for any derivative δ

$$\delta g^{\mu\nu} = -g^{\mu\lambda} g^{\nu\rho} \delta g_{\lambda\rho} \tag{9.1}$$

holds. The metric is assumed to be non-vanishing and has a signature such that its determinant is negative,

$$g = \det g_{\mu\nu} < 0.$$

The covariant volume element dV is therefore given by

$$\begin{aligned} dV &= hd^4x \\ h &= \sqrt{-g} = \sqrt{-\det g_{\mu\nu}} > 0, \end{aligned}$$

implying that h is real (hermitian), and has derivative

$$\delta h = \frac{1}{2} h g^{\mu\nu} \delta g_{\mu\nu} = -\frac{1}{2} h g_{\mu\nu} \delta g^{\mu\nu} \tag{9.2}$$

as a consequence of (2.14).

The most important concept of general relativity is the covariance (or invariance) under a general coordinate transformation $x_\mu \to x'_\mu$ (diffeomorphism) having

$$\begin{aligned} dx'^{\mu} &= \frac{\partial x'^{\mu}}{\partial x^\nu} dx^\nu = J^\mu_\nu dx^\nu \\ \det(J) &\neq 0, \end{aligned}$$

where the condition on the Jacobian J follows directly from the requirement to have an invertible coordinate transformation everywhere. Scalars $\phi(x)$ are invariant under such coordinate transformations, i. e., $\phi(x) \to \phi(x')$. Covariant and contravariant tensors of n-th order transform as

$$\begin{aligned} T'_{\mu...\nu}(x') &= \frac{\partial x_\mu}{\partial x'_\alpha} ... \frac{\partial x_\nu}{\partial x'_\beta} T_{\alpha...\beta}(x) \\ T'^{\mu...\nu}(x') &= \frac{\partial x'^{\mu}}{\partial x^\alpha} ... \frac{\partial x'^{\nu}}{\partial x^\beta} T^{\alpha...\beta}(x) \end{aligned}$$

respectively, and contravariant and covariant indices can be exchanged with a metric factor, as in special relativity. As a consequence, the ordinary derivative ∂_μ of a tensor A_ν of rank one or higher is not a tensor. To obtain a tensor from a differentiation the covariant derivative must be used

$$\begin{aligned} D_\mu A_\nu &= \partial_\mu A_\nu - \Gamma^\lambda_{\mu\nu} A_\lambda \\ \Gamma^\lambda_{\mu\nu} &= \frac{1}{2} g^{\lambda\sigma} (\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}), \end{aligned} \tag{9.3}$$

where Γ are the Christoffel symbols. Only the combination hA_ν, yielding a tensor density, obeys

$$D_\mu(hA_\nu) = \partial_\mu(hA_\nu).$$

As a consequence, covariant derivatives no longer commute, and their commutator is given by the Riemann tensor $R_{\lambda\rho\mu\nu}$ as

$$\begin{aligned} [D_\mu, D_\nu]\, A^\lambda &= R^\lambda_{\rho\mu\nu} A^\rho \\ R^\lambda_{\rho\mu\nu} &= \partial_\mu \Gamma^\lambda_{\nu\rho} - \partial_\nu \Gamma^\lambda_{\nu\rho} + \Gamma^\lambda_{\mu\sigma}\Gamma^\sigma_{\nu\rho} - \Gamma^\lambda_{\nu\rho}\Gamma^\sigma_{\mu\rho}, \end{aligned}$$

which also determines the Ricci tensor and the curvature scalar

$$\begin{aligned} R_{\mu\nu} &= R^\lambda_{\nu\mu\lambda} \\ R &= R^\mu_\mu, \end{aligned}$$

respectively.

These definitions are sufficient to write down the basic dynamical equation of general relativity, the Einstein equation

$$R_{\mu\nu} = \frac{1}{2} g_{\mu\nu} R + g_{\mu\nu}\Lambda = -\kappa T_{\mu\nu}$$

which can be derived as the Euler-Lagrange equation from the Einstein-Hilbert Lagrangian[1]

$$\mathcal{L} = \frac{1}{2\kappa} hR - \frac{1}{\kappa} h\Lambda + h\mathcal{L}_M,$$

where $\mathcal{L}_M$ is the matter Lagrangian yielding the covariantly conserved energy momentum tensor $T_{\mu\nu}$

$$T_{\mu\nu} = \left(-\eta_{\mu\nu}\mathcal{L} + 2\frac{\delta \mathcal{L}_M}{\delta g^{\mu\nu}} (g_{\mu\nu} = \eta_{\mu\nu}) \right), \tag{9.4}$$

and again $\kappa = 16\pi G_N$ is Newton's constant, and Λ gives the cosmological constant (with arbitrary sign).

For the purpose of quantization it is useful to rewrite the first term of the Lagrangian, the Einstein-Hilbert contribution $\mathcal{L}_E$, as

$$\begin{aligned} \mathcal{L}_E &= \frac{1}{2\kappa} h g^{\mu\nu} (\Gamma^\lambda_{\sigma\lambda}\Gamma^\sigma_{\mu\nu} - \Gamma^\lambda_{\mu\sigma}\Gamma^\sigma_{\nu\lambda}) + \partial_\mu V^\mu \\ V^\lambda &= \frac{1}{2\kappa} h (g^{\mu\lambda} g^{\sigma\tau} - g^{\mu\sigma} g^{\lambda\tau}) \partial_\mu g_{\sigma\tau}. \end{aligned}$$

The second term is a total derivative, and therefore quite often can be dropped.

[1] In the following usually the cosmological constant term will be absorbed in the matter part.

There is an important remark to be made about classical general relativity. The possibility of making a general coordinate transformation leaving physics invariant has the consequence that both energy and three momentum loose their meaning as physically meaningful concepts, just like charge in a non-Abelian gauge theory. Indeed it is possible to alter the energy of a system by performing a space-time coordinate transformation. Only the concept of total energy (or momentum) of a localized distribution of particles when regarded from far away in an otherwise flat space-time can be given an (approximate) physically meaning, similarly to electric charge. Therefore, many concepts which are usually taken to be physical lose this meaning when general relativity is involved. This carries over to any quantum version.

9.4.2 The Rarita-Schwinger field

The metric, as a symmetric tensor, describes a spin 2 object. Supersymmetrizing gravity therefore requires a spin 3/2 field, which is the so-called Rarita-Schwinger field. This field will be the associated gauge field for the local supersymmetry, just as the metric field is the gauge field for the local translation symmetry.

In analogy to conventional gauge theories, the Rarita-Schwinger field is required to transform under a local transformation as

$$\Psi_\mu \to \Psi_\mu + \partial_\mu \epsilon$$

where ϵ is a spinor-valued function. This follows in essentially the same as for vector fields from teh representation theory of the Lorentz group. Thus, a Rarita-Schwinger field Ψ carries both, a vector index and a spinor index. This was to be expected, as this couples effectively a spin 1 and a spin 1/2 object to create spin 3/2, just as for the metric two spin 1 indices are coupled to spin 2. As the supercharges are Weyl/Majorana spinors, so are the Rarita-Schwinger components.

Since the transformation is linear, it is an Abelian gauge theory, and the corresponding field strength tensor

$$\Omega_{\mu\nu} = \partial_\mu \Psi_\nu - \partial_\nu \Psi_\mu$$

is therefore gauge invariant, but carries also a spinor index.

It is still necessary to postulate a Lagrangian for the theory, which is gauge-invariant. Introducing $\bar{\Psi} = \Psi^\dagger \gamma_0$, a possibility is

$$\begin{aligned}
\mathcal{L} &= -\bar{\Psi}_\mu \gamma^{\mu\nu\rho} \partial_\nu \Psi_\rho. \\
\gamma^{\mu\nu\rho} &= \frac{1}{2}\{\gamma^\mu, \gamma^{\rho\sigma}\} \\
\gamma^{\mu\nu} &= \frac{1}{2}[\gamma^\mu, \gamma^\nu]
\end{aligned}$$

As for the Maxwell case, there are no gauge-invariant, perturbatively renormalizable further interaction terms possible. Without interactions, only non-interacting Rarita-Schwinger fields are possible. The equation of motion is, similar to the Dirac equation,

$$\gamma^{\mu\nu\rho}\partial_\nu\Psi_\rho = 0.$$

It follows that the Rarita-Schwinger field can have (classically) physical modes only for $d > 3$, similar like the vector potential only for $d > 2$. This equation of motion also implies

$$\gamma^\mu\Psi_{\mu\nu} = 0,$$

which is Rarita-Schwinger form of the Maxwell equations. The equations of motions can be solved in a similar way as the free Dirac equation, and creates the free-field solutions.

It is possible to add a mass term, yielding

$$\mathcal{L} = -\bar{\Psi}_\mu(\gamma^{\mu\nu\rho}\partial_\nu - m\gamma^{\mu\rho})\Psi_\rho,$$

in contrast to the vector gauge fields.

9.4.3 Supergravity

The actual supergravity action is somewhat involved. Here, only the situation will be considered without additional matter fields, as they would have to be supersymmetrized as well. As shown above, the coupling of different matter multiplets leads to intertwining of those, which leads to a rather involved result. Also, the cosmological constant will be set to zero in the following.

The coupling between fermions and gravity actually is not a straightforward exercise in itself[2]. The approach taken here is based on exchanging the metric in favor of a different type of dynamical variables, the so-called vierbeins, defined as

$$g_{\mu\nu} = e_\mu^a\eta_{ab}e_\nu^b$$

where η is the space-time-constant Minkowski metric, and also the indices a and b run therefore from 0 to 3. This relation implies

$$e_a^\mu g_{\mu\nu}e_b^\nu = \eta_{ab},$$

i. e. the vierbein is the matrix field which yields a transformation of some given metric to the Minkowski metric. This field is therefore sometimes also called a frame field, as it

[2]In fact, there is more than one possibility, and they differ at the quantum level. Without experimental input, it is at the current time not possible to decide which one is correct. Here, the most prevalent construction is chosen.

locally transforms the metric to a Minkowski frame. Both indices of the vierbein can be raised and lowered using the Minkowski metric.

The Lagrangian of the simplest $\mathcal{N}=1$ supergravity is then

$$\begin{aligned}
\mathcal{L} &= \frac{\det e}{2\kappa}\left(e^{a\mu}e^{b\nu}R_{\mu\nu ab}-\bar{\Psi}_\mu\gamma^{\mu\nu\rho}D_\nu\Psi_\rho\right)\\
R_{\mu\nu ab} &= \partial_\mu\omega_{\nu ab}-\partial_\nu\omega_{\mu ab}+\omega_{\mu ac}\omega_\nu{}^c{}_b-\omega_{\nu ac}\omega_\mu{}^c{}_b\\
D_\nu &= \partial_\nu+\frac{1}{4}\omega_{\nu ab}\gamma^{ab}\\
\omega_{\nu ab} &= 2e_{\mu[a}\partial_{[\nu}e_{b]}^{\mu]}-e_{\mu[a}e_{b]}^{\sigma}e_{\nu c}\partial^\mu e_\sigma^c
\end{aligned}$$

where the brackets around the indices indicates that the expression has to be antisymmetrized with respect to the same-type indices. It is seen that the covariant derivative couples gravity and the Rarita-Schwinger field. This theory is therefore coupled. The involvement of the vierbein also makes the Einstein-Hilbert part more involved, and modifies the Riemann tensor, which now involves different types of indices.

The, now local, supersymmetry transformations of the fields are, without proof,

$$\begin{aligned}
\delta e_\mu^a &= \frac{1}{2}\bar{\epsilon}\gamma^a\Psi_\mu\\
\delta\Psi_\mu &= D_\mu\epsilon,
\end{aligned}$$

with the spinor 1/2 field $\epsilon(x)$.

9.5 Introduction

The following will discuss the quantization of the simplest possible string system, the simple, non-interacting, bosonic string. This will still be a formidable task, and will yield a number of rather generic properties of string theories, like the natural appearance of gravitons, the need for additional dimensions, and the problems encountered with, e. g., tachyons. In particular the natural appearance of the graviton makes string theories rather interesting, given the intrinsic problems of quantum gravity. Further advantages of more sophisticated string theories are that they have generically few parameters, are not featuring space-time singularities such as black holes on a quantum level, and often have no need for renormalization, thus being consistent ultraviolet completions. The price to be paid is that only rather complicated string theories even have a chance to resemble the standard model, their quantization beyond perturbation theory is not yet fully solved, and it is unclear how to identify a string theory which has a vacuum state which is compatible with standard model physics. Furthermore, in general genuine string signatures usually only appear at energy levels comparable to the Planck scale, making an experimental investigation, or even verification of stringy properties of physics, almost impossible with the current technology.

How comes the idea of string theory about? Generically, as motivates all the searches beyond the standard model, the understanding has been increased by looking at ever shorter distances and at ever high energies. The current confirmed state of affairs is then the standard model. Going back to quantum gravity, a similar insight can be gained. In the perturbative approach, the ratio of free propagation to a tree-level exchange of a graviton is essentially given by the interaction strength of gravity times a free graviton propagator, which is essentially given by the inverse of $G_N E^2$, with E the energy of the graviton. Thus the corresponding ratio is

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

where $M_P^2 = \hbar c^5/G_N$ is again the Planck mass, this time in standard units. Since M_P is once more of the order of 10^{19} GeV, this effect is negligible for typical collider energies of TeV. However, if the energy becomes much larger than the scale, the ratio of free propagation to exchange of a graviton becomes much smaller than one, indicating the breakdown of perturbation theory.

This is not cured by higher order effects. E. g., in case of the two-graviton exchange, the corresponding amplitude ratio becomes

$$\frac{A_{2g}}{A_{\text{free}}} \sim (\hbar G_N)^2 \sum_{\text{Intermediate states}} \int_0^E dE' E'^3 \sim \frac{1}{M_P^4} \int dE' E'^3 \to \infty \text{ for } E \to \infty \tag{9.5}$$

This gets even worse with each higher order of perturbation theory. Thus, perturbation theory completely fails for quantum gravity. Either non-perturbative effects kick in, or something entirely different. That might be string theory.

The basic idea behind string theory is to try something new. The problem leading to the divergence of (9.5) is that with ever increasing energy ever shorter distances are probed, and by this ever more gravitons are found. This occupation with gravitons is then what ultimately leads to the problem. The ansatz of string theory is then to prevent such an effect. This is achieved by smearing out the interaction over a space-time volume. For a conventional quantum field theory such an inherent non-locality usually comes with the loss of causality. String theories, however, are a possibility to preserve causality and smear out the interaction in such a way that the problem is not occurring.

However, the approach of string theory actually goes (and, as a matter of fact, has to go) a step further. Instead of smearing only the interaction, it smears out the particles themselves. Of course, this occurs already anyway in quantum physics by the uncertainty principle. But in quantum field theory it is still possible to speak in the classical limit of a world-line of a particle. In string theory, this world line becomes a world sheet. In fact, string theories can also harbor world volumes in the form of branes. However, a dynamical theory of such branes, called M(atrix)-theory, is still not known, despite many efforts. One of the problems in formulating such a theory is that internal degrees of freedom of a world volume are also troublesome, and can once more give rise to consistency problems. String theory seems to be singled out to be theory with just enough smearing to avoid the problems of quantum field theory and at the same time having enough internal rigidity as to avoid new problems. The details of this are beyond the scope of this lecture, which thus only introduces string theory.

One feature of string theory is that there is usually no consistent solution in four space-time dimensions, but typically more are required. How many more is actually a dynamical property of the theory: It is necessary to solve it to give an answer. In perturbation theory, it appears that ten dimensions are required, but beyond perturbation theory indications have been found that rather elven dimensions are necessary. Anyway, the number is usually too large. Thus, some of the dimensions have to be hidden, which can be performed by compactification, as with the setup for large extra dimensions. Indeed, as has been emphasized, large extra dimensions are rather often interpreted as a low-energy effective theory of string theory.

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

Here, these problems will be left aside in favor for a very simple string theory. This theory will exhibit many generic features of string theory, despite requiring 26 (large) dimensions and, at least perturbatively, will not have a stable vacuum state. The latter will be signaled by the existence of a tachyon, a particle traveling faster than the speed of light, which is another generic, though beatable, problem of string theories.

To give a more intuitive picture for the peculiarities and properties of string theory in the following a point particle and its quantization will be compared step-by-step to the quantization of the string theory.

9.6 Classic string theories

In the following the number of dimensions will be D, and the Minkowski metric will take the form

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & & 0 & \\ & 1 & & 0 \\ 0 & & \ddots & \\ & 0 & & 1 \end{pmatrix}.$$

It is actually a good question why the signature of the Minkowski metric should be like this, and also string theory so far failed to provide a convincing answer. But before turning to string theory, it makes sense to set the stage with a relativistic point particle.

9.6.1 Point particle

To become confident with the concepts take a classical particle moving along a world line in D dimensions. Classically, a trajectory is described by the $D-1$ spatial coordinates $x_i(t)$ as a function of time $t = x_0$. More useful in the context of string theory is a redundant description in terms of D functions $X_\mu(\tau)$ of a variable τ, which strictly monotonously increases along the world line. A natural candidate for this variable is the eigentime, which thus parametrizes the world line of the particle.

The simplest Poincare-invariant action describing a free particle of mass m in terms of the eigentime is then given by

$$S_{pp} = -m \int d\tau \sqrt{-\partial_\tau X^\mu \partial^\tau X_\mu}. \tag{9.6}$$

This thus tells that the minimum action is obtained for the minimum (geodetic) length of the world line. Variation along the world line

$$\delta \dot{X}_\mu \equiv \delta \partial_\tau X^\mu = \partial_\tau \delta X^\mu$$

yields the equation of motion as

$$
\begin{aligned}
\delta S_{pp} &= -m\int d\tau\left(\sqrt{-\dot{X}_\mu\dot{X}^\mu}-\sqrt{-\left(\dot{X}^\mu+\delta\dot{X}^\mu\right)\left(\dot{X}_\mu+\delta\dot{X}_\mu\right)}\right)\\
&= -m\int d\tau\left(\sqrt{-\dot{X}_\mu\dot{X}^\mu}-\sqrt{-\left(\dot{X}_\mu\dot{X}^\mu+2\dot{X}^\mu\delta\dot{X}_\mu\right)}\right)\\
&= -m\int d\tau\left(\sqrt{-\dot{X}_\mu\dot{X}^\mu}-\sqrt{-\dot{X}_\mu\dot{X}^\mu\left(1+2\frac{\dot{X}^\mu\delta\dot{X}_\mu}{\dot{X}_\mu\dot{X}^\mu}\right)}\right)\\
&\overset{\text{Taylor}}{=} m\int d\tau\frac{\dot{X}^\mu\delta\dot{X}_\mu}{\sqrt{-\dot{X}^\mu\dot{X}_\mu}}
\end{aligned}
$$

where in the last line use has been made of the infinitesimality of $\delta\dot{X}_\mu$ and the square root has been Taylor-expanded.

Defining now the D-dimensional normalized speed as

$$u^\mu=\frac{\dot{X}^\mu}{\sqrt{-\dot{X}_\mu\dot{X}^\mu}} \tag{9.7}$$

yields the equation of motion after imposing the vanishing of the action under the variation and a partial integration as

$$m\dot{u}^\mu=0 \tag{9.8}$$

This is nothing else then the equation of motion for a free relativistic particle, which of course reduces to the one of Newton in the limit of small speeds. This also justifies the interpretation of m as the rest mass of the particle.

With τ the eigentime the action is indeed Poincare-invariant. This can be seen as follows. A Poincare transformation is given by

$$X'^\mu=\Lambda^\mu_\nu X^\nu+a^\mu.$$

Inserting this expression for the argument of the square root yields

$$
\begin{aligned}
&\partial_\tau\left(\Lambda^\mu_\nu X^\nu+a^\mu\right)\partial_\tau\left(\Lambda_\mu^{\ \omega}X_\omega+a^\mu\right)\\
=&\left(\Lambda^\mu_\nu\Lambda_\mu^{\ \omega}\right)\partial_\tau X^\nu\partial_\tau X_\omega.
\end{aligned}
$$

Since the expression in parenthesis is just δ_ν^ω because of the (pseudo-)orthogonality of Lorentz transformations, this makes the expression invariant. Since the eigentime is invariant by definition, this shows the invariance of the total action.

Additionally, it is also reparametrization invariant, i. e., it is possible to transform the eigentime to a different variable without changing the contents of the theory, as it ought to be: Physics should be independent of the coordinate systems imposed by the observer. This is what ultimately leads to the diffeomorphism (diff) invariance of general relativity.

To show this invariance also for the action (9.6) take an arbitrary (but invertible) reparametrization $\tau' = f(\tau)$. This implies

$$\begin{aligned} \dot{\tau}' &= \frac{d\tau'}{d\tau} \\ d\tau &= \frac{d\tau'}{\dot{\tau}'}, \end{aligned}$$

yielding the transformation property of the integral measure. For the functions follows then

$$\dot{X}^{\mu'}(\tau') = \dot{X}^{\mu}(\tau)\frac{d\tau}{d\tau'} = \dot{X}^{\mu}\frac{1}{\dot{\tau}'}$$

Hence the scalar product changes as

$$\dot{X}^{\mu'}\dot{X}_{\mu'} = \frac{1}{\dot{\tau}'^2}\dot{X}^{\mu}\dot{X}_{\mu}.$$

One power of $\dot{\tau}'$ is removed by the square root, and the remaining one is then compensated by the integral measure.

Showing this explicitly for the action (9.6) was rather tedious, and it is useful to rewrite the action. For this purpose it is useful to introduce a metric along the world line. Since the world line is one-dimensional, this metric is only a single function $\gamma_{\tau\tau}(\tau)$ of the eigentime. This yields a trivial example of a tetrad η

$$\eta(\tau) := \left(-\gamma_{\tau\tau}(\tau)\right)^{\frac{1}{2}},$$

which in general is a set of N (by definition positive) orthogonal vectors on a manifold. However, the manifold is just one-dimensional for a world line, and thus the tetrad is again only a scalar. In analogy to the string case, $\gamma_{\tau\tau}$ can also be denoted as the world-line metric.

Taking the tetrad as an independent function a new action is defined as

$$S'_{pp} = \frac{1}{2}\int d\tau\left(\frac{\dot{X}^{\mu}\dot{X}_{\mu}}{\eta} - \eta m^2\right).$$

Under a reparametrization $\tau \to \tau'(\tau)$ it is defined that the functions X and η transform as

$$\begin{aligned} X(\tau) &= X(\tau'(\tau)) \\ \eta'(\tau') &= \eta(\tau)\frac{d\tau}{d\tau'} = \frac{1}{\dot{\tau}'}\eta(\tau) \end{aligned} \qquad (9.9)$$

This makes the expression invariant under diffeomorphisms: The transformation of η (9.9) takes care of the extra factor of $\dot{\tau}'$, and also makes the second expression invariant.

To show that the new action is indeed equivalent to the old, and that η is thus just an auxiliary function, can be shown by using the equation of motion for η. Using the Euler-Lagrange equation this time yields

$$\begin{aligned} 0 = \frac{d}{d\tau}\frac{\partial L}{\partial \dot{\eta}} - \frac{\partial L}{\partial \eta} &= \frac{\dot{X}^\mu \dot{X}_\mu}{\eta^2} + m^2 \\ &\implies \eta^2 = -\frac{\dot{X}^\mu \dot{X}_\mu}{m^2}. \end{aligned}$$

Thus knowledge of X determines η completely, since no derivatives of η appear. Inserting this expression into 9.9 leads to

$$\begin{aligned} S'_{pp} &= \frac{1}{2}\int d\tau \left(\frac{\dot{X}^\mu \dot{X}_\mu}{\sqrt{-\frac{\dot{X}^\mu \dot{X}_\mu}{m^2}}} - \sqrt{-\frac{\dot{X}^\mu \dot{X}_\mu}{m^2} m^2} \right) \\ &= \frac{m}{2}\int d\tau \left(-\frac{-\dot{X}^\mu \dot{X}_\mu}{\sqrt{-\dot{X}^\mu \dot{X}_\mu}} - \sqrt{-\dot{X}^\mu \dot{X}_\mu} \right) \\ &= -m\int d\tau \sqrt{-\dot{X}^\mu \dot{X}_\mu} = S_{pp}. \end{aligned}$$

Thus S'_{pp} is indeed equivalent to S_{pp}. However, one advantage remains to be exploited. By separation of the mass S'_{pp} can also be applied to the case of $m = 0$ directly, which is only possible in a limiting process for the original action S_{pp}.

9.6.2 Strings

For strings, the world line becomes a world sheet. As a consequence, at any fixed eigentime τ the string has an extension. This extension can be infinite or finite. In the latter case, the string can be closed, i. e., its ends are connected, or open. In string theories usually only finite strings appear, with lengths L of size the Planck length. Furthermore, open strings have usually to have their ends located on branes. This is not necessary for the simple case here, which will be investigated both for open and closed strings.

Analogous to the eigentime then an eigenlength σ can be introduced. Both parameters together describe any point on the world-sheet. The functions X_μ describing the position of the points of the world-sheet are therefore functions of both parameters, $X_\mu = X_\mu(\sigma, \tau)$. Furthermore, as for the point particle, these functions should be reparametrization invariant

$$X^\mu(\sigma, \tau) = X^\mu(\sigma'(\sigma, \tau), \tau'(\sigma, \tau)) \tag{9.10}$$

such that the position of the world sheet is not depending on the parametrization.

Derivatives with respect to the two parameters will be counted by Latin indices $a, ...,$

$$\begin{aligned}\partial_{a,b,...} &= \partial_\tau, \partial_\sigma \\ \partial_0 &= \partial_\tau \\ \partial_1 &= \partial_\sigma.\end{aligned}$$

It is then possible to define the induced metric on the world sheet as

$$h_{ab} = \partial_a X^\mu \partial_b X_\mu,$$

as a generalization of $h_{\tau\tau} = \dot{X}^\mu \dot{X}_\mu$, which as a metric has already been used to define the action (9.6) in analogy to the Einstein-Hilbert action. $\sqrt{-\det h_{ab}} d\tau d\sigma$ is then an infinitesimally element of the world sheet area.

The simplest possible Poincare-invariant action which can be written down for this system is the Nambu-Goto action

$$S_{NG} = \int_M d\tau d\sigma \mathcal{L}_{\text{NG}}$$

in which M is the world-sheet of the string and $\mathcal{L}_{\text{NG}}$ is the Nambu-Goto Lagrangian

$$\mathcal{L}_{\text{NG}} = -\frac{1}{2\pi\alpha'}\sqrt{-\det h_{ab}} = -\frac{1}{2\pi\alpha'}\sqrt{\partial_\tau X_\mu \partial_\sigma X^\mu \partial_\sigma X_\rho \partial_\tau X^\rho - \partial_\tau X_\mu \partial_\tau X^\mu \partial_\sigma X^\rho \partial_\sigma X_\rho},$$

again the direct generalization of the point-particle action. In particular, the minimum area of the world sheet minimizes the action.

The constant α' is the so-called Regge slope, having dimension Mass squared. In principle, it could be set to one in the following for the non-interacting string, but due to its importance in the general case, it will be left explicit. The Regge slope can be associated with the string tension T as $T = 1/(2\pi\alpha')$.

The Nambu-Goto action has two symmetries. One is diffeomorphism invariance. This can be seen directly, as in the case of the point particle, except that now the Jacobian appears. The second invariance is Poincare invariance, which leaves the world-sheet parameters τ and σ invariant. However, the functions X_μ transform as

$$\begin{aligned}X'^\mu &= \Lambda^\mu{}_\nu X^\nu + a^\mu \\ \partial_a \Lambda^\mu{}_\nu X^\nu \partial_b \Lambda_\mu^{\ \gamma} X_\gamma &= \overbrace{\Lambda^\mu{}_\nu \Lambda_\mu^{\ \gamma}}^{\delta_\nu^\gamma} \partial_a X^\nu \partial_b X_\gamma = \partial_a X^\mu \partial_b X_\mu.\end{aligned}$$

Thus, the induced metric is Poincare invariant, and hence also the action as well as the Lagrangian and any other quantity constructed from it is.

It is once more rather cumbersome to use an action involving a square root. To construct a simpler action, it is useful to introduce a world-sheet metric $\gamma_{ab}(\tau,\sigma)$. This metric is taken to have a Lorentz signature for some chosen coordinate system

$$\gamma_{ab} = \begin{pmatrix} + & 0 \\ 0 & - \end{pmatrix}.$$

Thus, this metric is traceless, and has a determinant smaller zero. With it the new action, the Brink-Di Vecchia-Howe-Deser-Zumino or Polyakov action,

$$S_P = -\frac{1}{4\pi\alpha'}\int_M d\tau d\sigma\, (-\gamma)^{\frac{1}{2}}\, \gamma^{ab} h_{ab} \tag{9.11}$$

is constructed, where γ denotes $\det \gamma_{ab}$.

As in case of the point particle, the world-sheet metric γ_{ab} has to have a non-trivial transformation property under diffeomorphisms,

$$\frac{\partial\omega'^c}{\partial\omega^a}\frac{\partial\omega'^d}{\partial\omega^b}\gamma'_{cd}(\tau',\sigma') = \gamma_{ab}(\tau,\sigma),$$

where the variables ω denote either σ and τ, depending on the index. This guarantees that for all invertible reparametrizations, which are continuous deformations of the identity transformation, the metric is still traceless and has negative determinant.

To obtain the relation of the Polyakov action to the Nambu-Goto action it is again necessary to obtain its equation of motion. This is most conveniently obtained using the variational principle. For this, the general relation for determinants of metrics

$$\delta\gamma = \gamma\gamma^{ab}\delta\gamma_{ab} = -\gamma\gamma_{ab}\delta\gamma^{ab}$$

is quite useful.

Abbreviating the Polyakov Lagrangian by L_P and performing a variation with respect to γ yields

$$\begin{aligned}
\delta S_P &= -\frac{1}{4\pi\alpha'}\int d\tau d\sigma \left(L_P - (-\gamma-\delta\gamma)^{\frac{1}{2}}\left(\gamma^{ab}+\delta\gamma^{ab}\right)h_{ab}\right)\\
&= -\frac{1}{4\pi\alpha'}\int d\tau d\sigma \left(L_P - \left(-\gamma+\gamma\gamma^{cd}\delta\gamma_{cd}\right)^{\frac{1}{2}}\left(\gamma^{ab}+\delta\gamma^{ab}\right)h_{ab}\right)\\
&= -\frac{1}{4\pi\alpha'}\int d\tau d\sigma \left(L_P - (-\gamma)^{\frac{1}{2}}\left(1-\gamma^{cd}\delta\gamma_{cd}\right)^{\frac{1}{2}}\left(\gamma^{ab}+\delta\gamma^{ab}\right)h_{ab}\right).
\end{aligned}$$

Expanding the term with indices cd up to first order in the variation leads to

$$\begin{aligned}
\delta S_P &= -\frac{1}{4\pi\alpha'}\int d\tau d\sigma \left(L_P - (-\gamma)^{\frac{1}{2}}\left(1-\frac{1}{2}\gamma^{cd}\delta\gamma_{cd}\right)\left(\gamma^{ab}+\delta\gamma^{ab}\right)h_{ab}\right)\\
&= -\frac{1}{4\pi\alpha'}\int d\tau d\sigma \left(L_P - (-\gamma)^{\frac{1}{2}}\left(\gamma^{ab}+\delta\gamma^{ab}-\frac{1}{2}\gamma^{cd}\gamma^{ab}\delta\gamma_{cd}\right)h_{ab}\right).
\end{aligned}$$

The second term is again the Polyakov Lagrangian, canceling the zero-order term. Then only

$$\delta S_P = -\frac{1}{4\pi\alpha'}\int d\tau d\sigma\,(-\gamma)^{\frac{1}{2}}\left(h_{ab}-\frac{1}{2}\gamma_{ab}\gamma_{cd}h^{cd}\right)\delta\gamma^{ab}$$

is left.

The condition that this expression should vanish yields the equation of motion for the world-sheet metric as

$$h_{ab} = \frac{1}{2}\gamma_{ab}\gamma_{cd}h^{cd} \tag{9.12}$$

Division of each side by its determinant finally yields

$$\begin{aligned}\frac{h_{ab}}{(-h)^{\frac{1}{2}}} &= \frac{1}{2}\frac{\gamma_{ab}\left(\gamma_{cd}h^{cd}\right)}{\left(\det -\frac{1}{2}\gamma_{ab}\gamma_{cd}h^{cd}\right)^{\frac{1}{2}}}\\ &= \frac{1}{2}\frac{\gamma_{ab}\left(\gamma_{cd}h^{cd}\right)}{\left(\left(\frac{1}{2}\gamma_{cd}h^{cd}\right)^2\det -\gamma_{ab}\right)^{\frac{1}{2}}}\\ &= \frac{\gamma_{ab}}{(-\gamma)^{\frac{1}{2}}}\end{aligned}$$

In the second line it has been used that $\gamma_{cd}h^{cd}$ is a scalar, permitting it to pull it out of the determinant. The result implies that h and γ are essentially proportional.

Inserting this result in the Polyakov action yields

$$S_P = -\frac{1}{4\pi\alpha'}\int d\tau d\sigma\gamma^{ab}\gamma_{ab}\,(-h)^{\frac{1}{2}} = -\frac{1}{2\pi\alpha'}\int d\tau d\sigma\,(-h)^{\frac{1}{2}} = S_{NG}$$

showing that it is indeed equivalent to the Nambu-Goto action, where the fact that the diffeomorphism invariant quantity $\gamma^{ab}\gamma_{ab}$ is two, due to the Lorentz signature of γ, has been used.

The Polyakov action thus retains the Poincare and diffeomorphism invariance of the Nambu-Goto action. The Poincare invariance follows since γ is Poincare invariant, since it is proportional to the Poincare-invariant induced metric, thus

$$\gamma^{ab\prime} = \Lambda\gamma^{ab} = \gamma^{ab}.$$

The diffeomorphism invariance follows directly from the transformation properties of the world-sheet metric, in total analogy with the point-particle case, but considerably more lengthy since track of both variables has to be kept.

The redundancy introduced with the additional degree of freedom γ grants a further symmetry. This is the so-called Weyl symmetry, given by

$$\begin{aligned} X'^{\mu}(\tau,\sigma) &= X^{\mu}(\tau,\sigma) \\ h'_{ab} &= h_{ab} \\ \gamma'_{ab} &= e^{2\omega(\tau,\sigma)}\gamma_{ab}. \end{aligned}$$

for arbitrary functions $\omega(\tau,\sigma)$. The origin of this symmetry comes from the unfixed proportionality of induced metric and the world-sheet metric. The expression of γ in terms of the induced metric h is invariant under this transformation,

$$\frac{\gamma'_{ab}}{(-\gamma')^{\frac{1}{2}}} = \frac{\gamma_{ab}e^{2\omega}}{(-\gamma')^{\frac{1}{2}}} = \frac{\gamma_{ab}e^{2\omega}}{(-\gamma e^{4\omega})^{\frac{1}{2}}} = \frac{\gamma_{ab}}{(-\gamma)^{\frac{1}{2}}}.$$

Also the action is invariant. To see this note that γ_{ab} is indeed a metric. Since $\gamma_{ab}\gamma^{ab}$ has to be a constant, as noted before, this implies that

$$\gamma'^{ab} = e^{-2\omega}\gamma^{ab}.$$

As a consequence, the expression appearing in the action transforms as

$$(-\gamma')^{\frac{1}{2}}\gamma'^{ab} = \left(-\gamma e^{4\omega}\right)^{\frac{1}{2}}\gamma^{ab}e^{-2\omega} = (-\gamma)^{\frac{1}{2}}\gamma^{ab}.$$

Thus, the Weyl invariance is indeed a symmetry.

The Polyakov action can also be viewed with a different interpretation. Promoting the world-sheet indices to space-time indices and taking the indices μ to label internal degrees of freedom, then the Polyakov action just describes D massless Klein-Gordon fields X_μ (with internal symmetry group SO($D-1$,1)) in two space-time dimensions with a non-trivial metric γ, which is dynamically coupled to the fields. This is an example of a duality of two theories, which plays an important role for more complicated theories. E. g., dualities between certain string theories on certain background metrics with so-called supergravity theories, the AdS/CFT correspondence, had an enormous impact recently on both string theory and quantum field theory.

9.7 Quantized theory

9.7.1 Light cone gauge

As the Poincare and Weyl symmetry introduce a gauge symmetry, it is easier to perform the quantization in a fixed gauge. Particularly useful for this purpose in the present context is the light-cone gauge. Though this gauge is not keeping manifest Poincare covariance, it is very useful (similar to the case of quantizing electrodynamics in Coulomb rather than linear covariant gauges). Proving that the theory is still covariant after quantization is non-trivial, but possible. Hence, this will not be shown here.

To formulate light-cone gauge light-cone coordinates are useful. They are introduced by the definitions

$$\begin{aligned} x^{\pm} &= \frac{1}{\sqrt{2}}\left(x^0 \pm x^1\right) \\ x^i &= x^i,\ i = 2, ..., D-1, \end{aligned}$$

and thus mix the time-coordinate and one, now distinguished, spatial coordinate. Since the zero-component is the only one involving a non-positive sign in the metric this yields the following relation between covariant and contravariant light-cone coordinates

$$\begin{aligned} x_{\pm} &= \frac{1}{\sqrt{2}}\left(x_0 \mp x_1\right) \\ x_- &= -x^+ \\ x_+ &= -x^- \\ x_i &= x^i. \end{aligned}$$

This implies the metric

$$a^\mu b_\mu = a^+ b_+ + a^- b_- + a^i b_i = -a^+ b^- - a^- b^+ + a^i b^i,$$

which is equivalent to the conventional one

$$\begin{aligned} -a^+ b^- - a^- b^+ + a^i b^i &= -\frac{1}{2}\left(a^0 + a^1\right)\left(b^0 - b^1\right) - \frac{1}{2}\left(a^0 - a^1\right)\left(b^0 + b^1\right) + a^i b^i \\ &= -a^0 b^0 + a^1 b^1 + a^i b^i = a^\mu b_\mu. \end{aligned}$$

Aim of the gauge fixing is to restore the original number of independent degrees of freedom. In case of the point particle this amounts to remove the eigentime τ. This is most conveniently done by the condition

$$\tau \equiv x^+,$$

thus being the light-cone gauge condition for the point particle. This is more convenient than the more conventional choice $\tau = x^0$. With this x^+ corresponds to the time and p^- to the energy . Correspondingly, x^- and p^+ are now longitudinal degrees of freedom while x^i and p^i are transverse ones. This immediately follows from the scalar product

$$\frac{\partial}{\partial a^+}\left(-a^+ b^- + ...\right) = -b^-,$$

and correspondingly for the derivative with respect to x^+ which produces p^-.

9.7.2 Point particle

To demonstrate the principles, it is once more convenient to first investigate the point particle. However, one should be warned that the resulting theory is actually flawed due to the appearance of unphysical (non-normalizable) states. It should therefore be taken rather as a mathematical than a physical discussion.

Returning to the parametrization of the point particle of section 9.6.1, the gauge condition to fix the diffeomorphism invariance becomes

$$X^+(\tau) = \tau.$$

The action is given by equation (9.9), thus

$$\begin{aligned}
S'_{pp} &= \frac{1}{2}\int d\tau \left(\frac{\dot{X}^\mu \dot{X}_\mu}{\eta} - \eta m^2\right) \\
&= \frac{1}{2}\int d\tau \left(\frac{1}{\eta}\left(-\dot{X}^+\dot{X}^- - \dot{X}^-\dot{X}^+ + \dot{X}^i\dot{X}^i\right) - \eta^2 m\right) \\
&= \frac{1}{2}\int d\tau \left(\frac{1}{\eta}\left(-2\dot{X}^-\dot{\tau} + \dot{X}^i\dot{X}^i\right) - \eta^2 m\right) \\
&= \frac{1}{2}\int d\tau \left(\frac{1}{\eta}\left(-2\dot{X}^- + \dot{X}^i\dot{X}^i\right) - \eta m^2\right).
\end{aligned}$$

As usual, the Lagrangian yields the canonical conjugated momenta by the expression

$$P_\mu = \frac{\partial L}{\partial \dot{X}^\mu}$$

yielding

$$\begin{aligned}
P_- &= -\frac{1}{\eta} \\
P_i &= \frac{\dot{X}^i}{\eta}
\end{aligned}$$

With this the Hamiltonian can be readily constructed as

$$\begin{aligned}
H &= \sum P\dot{Q} - L \\
&= P_-\dot{X}^- + P_i\dot{X}^i - L \\
&= -\frac{\dot{X}^-}{\eta} + \eta P_i P_i + \frac{\dot{X}^-}{\eta} - \frac{1}{2}\dot{X}^i\dot{X}^i + \frac{1}{2}\eta m^2 \\
&= \eta P_i P_i - \frac{1}{2}\eta P_i P_i + \frac{1}{2}\eta m^2 = \frac{P^i P^i + m^2}{2P^+}.
\end{aligned}$$

Where it has been used that

$$P^+ = -P_- = \frac{1}{\eta},$$

and it is thus possible to remove η and P_- from the expression.

In this result the variable X^+ is no longer a dynamical variable, and thus the gauge is fixed. Furthermore it follows that $P_\eta = 0$, since the Lagrangian does not depend on $\dot{\eta}$. Hence η is not a dynamical variable. This was expected, since it was already in the classical case only used to make the Lagrangian more easily tractable

For the quantization then the usual canonical commutation relations are imposed as

$$\begin{aligned} \left[P_i, X^j\right] &= -i\delta_i^j \\ \left[P_-, X^-\right] &= -i \end{aligned}$$

The relations for P_+ is provided by the other relations, since P^- is the energy and thus

$$H = P^- = -P_+. \tag{9.13}$$

That is essentially the relativistic mass-shell equation, implying once more that P^+ is not an independent degree of freedom. The resulting Hamilton operator is the one of a $D-2$-dimensional harmonic oscillator, but supplemented with the additional unconstrained degree of freedom P_-. The spectrum of this is known, being a relativistic scalar (with all its sicknesses) and states $|k_-, k_i\rangle$.

9.7.3 Open string

Again, the first step is to fix the gauge. For that purpose first the permitted range for the world-sheet parameters have to be chosen, which will be

$$\begin{aligned} -\infty &\leq \tau \leq +\infty \\ 0 &\leq \sigma \leq L. \end{aligned} \tag{9.14}$$

Thus, L is the length of the string. Again, it is chosen that

$$\tau = X^+. \tag{9.15}$$

This deals again with the diffeomorphism degree of freedom. To also take care of the Weyl freedom a second condition is necessary, which will be chosen to be

$$\partial_\sigma \gamma_{\sigma\sigma} = 0 \tag{9.16}$$

$$\det \gamma_{ab} = -1 \tag{9.17}$$

The conditions (9.15-9.17) fixes these degrees of freedom completely, provided that the world-sheet is parametrized by the eigenvariables in such a way that one and only one set of eigentime and eigenlength correspond to a given point on the world sheet. In the case of the point particle, it can be shown that this condition is actually superfluous, since even in case of a doublebacking world line this would not contribute to a path integral. For string theory, this is something not yet really simply understood.

A way to get an intuition for the significance of these gauge condition is by the use of the invariant length. The choice of $\tau = X^+$ is of course always possible. Then start by the definition

$$f = \gamma_{\sigma\sigma}\left(\frac{1}{-\det\gamma_{ab}}\right)^{\frac{1}{2}}.$$

Now perform a reparametrization which leaves τ invariant. This implies

$$f' = f\frac{d\sigma}{d\sigma'}.$$

because of the transformation properties of the γ_{ab}. Hence, the length element $dl = f d\sigma$ is invariant under this reparametrization. Therefore, it can be considered as an invariant length-element, since it is not changing under a change of the eigenlength of the string. In fact, this can be used to define the σ coordinate, by setting it equal to $\int dl$ along the world sheet,

$$\sigma = \int_0^\sigma dl.$$

As a consequence, f can no longer depend on σ, since dl is σ-independent. Secondly, it is then possible to make a Weyl-transformation to rescale $\det\gamma$ such that it becomes -1, yielding (9.17), and fixing the Weyl invariance. Since f is Weyl-invariant by construction, this implies that $\partial_\sigma\gamma_{\sigma\sigma}$ trivially vanishes, yielding (9.16). Thus, in this coordinate system the gauge condition are fulfilled, and therefore are a permitted choice.

Since γ is by construction symmetric, these gauge condition permit to rewrite it in a simpler way. It then takes the form

$$\begin{aligned}\gamma &= \begin{pmatrix}\gamma^{\tau\tau} & \gamma^{\tau\sigma}\\ \gamma^{\sigma\tau} & \gamma^{\sigma\sigma}\end{pmatrix}\\ &= \begin{pmatrix}-\gamma_{\sigma\sigma}(\tau) & \gamma_{\tau\sigma}(\tau,\sigma)\\ \gamma_{\tau\sigma}(\tau,\sigma) & \gamma_{\sigma\sigma}^{-1}(\tau)\left(1-\gamma_{\tau\sigma}^2(\tau,\sigma)\right)\end{pmatrix},\end{aligned}$$

thereby eliminating two of the four variables in γ_{ab}, and also reducing their dependence on the world sheet parameters.

It is furthermore useful to define the average and variation of the X^- coordinate for the following as

$$\begin{aligned} Z^-(\tau) &= \frac{1}{L}\int_0^L d\sigma X^-(\tau,\sigma) \\ Y^-(\tau,\sigma) &= X^-(\tau,\sigma) - Z^-(\tau). \end{aligned}$$

This is the starting point to rewrite the action in a more useful form.

Start by rewriting the Lagrangian as

$$\begin{aligned} L_P &= -\frac{1}{4\pi\alpha'}\int_0^L d\sigma \overbrace{(-\gamma)^{\frac{1}{2}}}^{=1} \gamma^{ab}\partial_a X^\mu \partial_b X_\mu \\ &= -\frac{1}{4\pi\alpha'}\int_0^L d\sigma \left(\frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma X^\mu\partial_\sigma X_\mu + \gamma_{\tau\sigma}\partial_\tau X^\mu \partial_\sigma X_\mu + \gamma_{\tau\sigma}\partial_\sigma X^\mu\partial_\tau X_\mu - \gamma_{\sigma\sigma}\partial_\tau X^\mu\partial_\tau X_\mu\right). \end{aligned}$$

Now, it is useful to investigate the expressions piece-by-piece. Start with

$$\frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma X^\mu\partial_\sigma X_\mu = \frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma X^i\partial_\sigma X^i$$

where it has been used that

$$\partial_\sigma X^+ = \partial_\sigma\tau = 0$$

follows trivially from the gauge conditions. Next, use furthermore that

$$\gamma_{\tau\sigma}(\overbrace{\partial_\tau X^+\partial_\sigma X_+}^{=-\partial_\tau\tau\partial_\sigma X^-} + \partial_\tau X^- \overbrace{\partial_\sigma X_-}^{=-\partial_\sigma X^+} + \partial_\tau X^i\partial_\sigma X^i) = \gamma_{\tau\sigma}\left(-\partial_\sigma X^- + \partial_\sigma X^i\partial_\tau X^i\right)$$

and

$$-\gamma_{\sigma\sigma}\left(\partial_\tau X^+\partial_\tau X_+ + \partial_\tau X^-\partial_\tau X_- + \partial_\tau X^i\partial_\tau X^i\right) = -\gamma_{\sigma\sigma}\left(-2\partial_\tau X^- + \partial_\tau X^i\partial_\tau X^i\right).$$

Reinserting everything into the Lagrangian yields

$$\begin{aligned} L_P = & -\frac{1}{4\pi\alpha'}\int_0^L d\sigma\Big(\frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma X^i\partial_\sigma X^i - 2\gamma_{\tau\sigma}\left(-\partial_\sigma X^- + \partial_\tau X^i\partial_\sigma X^i\right) \\ & -\gamma_{\sigma\sigma}\left(-2\partial_\tau X^- + \partial_\tau X^i\partial_\tau X^i\right)\Big). \end{aligned}$$

Employing now the relations for the average and variation this yields

$$\begin{aligned} L_P = & -\frac{1}{4\pi\alpha'}\Big(\gamma_{\sigma\sigma}2L\partial_\tau Z^- + \int_0^L d\sigma\Big(\gamma_{\sigma\sigma}\left(-\partial_\tau X^i\partial_\tau X^i\right) \\ & +2\gamma_{\tau\sigma}\left(\partial_\sigma Y^- - \partial_\tau X^i\partial_\sigma X^i\right) + \frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma X^i\partial_\sigma X^i\Big)\Big) \end{aligned}$$

In the resulting expression there is no τ-derivative of Y^- appearing, which is thus a non-dynamical field, behaving like a Lagrange factor for $\gamma_{\tau\sigma}$, which therefore is fixed to

$$\partial_\sigma \gamma_{\tau\sigma} = 0, \tag{9.18}$$

and thus does not depend on σ.

Returning to the boundary condition of this open[3] string yields

$$\left(\partial_\sigma X^\mu\right)(\tau, 0) = \left(\partial_\sigma X^\mu\right)(\tau, L) = 0, \tag{9.19}$$

because otherwise the fields would not be continuously differentiable at the boundaries, which is imposed like for wave-functions. These are von Neumann conditions in the terms of the large extra dimensions. This is also obtained by varying the Polyakov action. First, vary with respect to the fields to obtain

$$-\frac{1}{2\pi\alpha'}\int_{-\infty}^{\infty} d\tau \, (-\gamma)^{\frac{1}{2}} \, \partial_\sigma X^\mu \delta X^\mu|_{\sigma=0}^{\sigma=L}.$$

Since this has to vanish for arbitrary variations of the fields, this implies the boundary condition (9.19).

On the other hand, when varying the original action with respect to the fields, this yields

$$\begin{aligned} \delta S_P &= S_P + \frac{1}{4\pi\alpha'}\int d\tau d\sigma \gamma_{ab}\partial_a \left(X^\mu + \delta X^\mu\right)\partial_b \left(X_\mu + \delta X_\mu\right) \\ &= \frac{1}{4\pi\alpha'}\int d\tau d\sigma \gamma_{ab}\left(\partial_a X^\mu \partial_b \delta X_\mu + \partial_\alpha \delta X^\mu \partial_b X_\mu\right). \end{aligned}$$

Since variation and differentiation are independent, they can be exchanged,

$$\partial_a \delta X^\mu = \delta \partial_a X^\mu.$$

Doing a partial integration, keeping an appearing boundary term yields

$$\begin{aligned} \delta S_P &= \frac{1}{4\pi\alpha'}\int d\tau d\sigma \gamma_{ab}\left(\partial_a \partial_b X^\mu \delta X_\mu + \partial_b \partial_a X_\mu \delta X^\mu\right) \\ &\quad -\frac{1}{4\pi\alpha'}\int d\tau \gamma_{ab}\left(\partial_a X^\mu \delta X_\mu + \partial_b X_\mu \delta X^\mu\right)|_0^L \end{aligned} \tag{9.20}$$

Note that in the boundary term as a shorthand notation one of the indices is uncontracted. This is of course always the σ-index for which the total integration has been performed.

[3]No cyclicity of any function on σ has been imposed, which would be one possibility to implement a closed string.

However, the last expression must vanish under variation, implying once more the von Neumann condition (9.19)

$$\gamma_{ab}\left(\partial_a X^\mu + \partial_b X^\mu\right) = 0.$$

Incidentally, this also implies for $\mu = +$ and $a = \tau$ and $b = \sigma$ that $\gamma_{\tau\sigma}$ vanishes on the boundary.

Since for $\mu = -$ the fields are non-dynamical, this implies that $\partial_\sigma X^- = 0$ and that therefore X^- only depends on τ.

To obtain some further useful results, the variation can be repeated after the gauge has been fixed. This yields

$$\begin{aligned}\delta S_P &= S_P + \frac{1}{4\pi\alpha'}\int d\tau d\sigma\Big(\gamma_{\sigma\sigma}\left(2\partial_\tau\left(X^- + \delta X^-\right) - \partial_\tau\left(X^i + \delta X^i\right)\partial_\tau\left(X^i + \delta X^i\right)\right)\\ &+2\gamma_{\tau\sigma}\left(\partial_\sigma\left(X^- + \delta X^-\right) - \partial_\tau\left(X^i + \delta X^i\right)\partial_\sigma\left(X^i + \delta X^i\right)\right)\\ &+\frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma\left(X^i + \delta X^i\right)\partial_\sigma\left(X^i + \delta X^i\right)\Big).\end{aligned}$$

Expanding the result and dropping $\mathcal{O}\left(\delta^2\right)$ terms and annihilating a term of type S_P just leaves

$$\begin{aligned}\delta S_P &= \frac{1}{4\pi\alpha'}\int d\tau d\sigma\Big(\gamma_{\sigma\sigma}\left(2\partial_\tau\delta X^- - 2\partial_\tau X^i\partial_\tau\delta X^i\right)\\ &+2\gamma_{\tau\sigma}\left(\partial_\sigma\delta X^- - \partial_\tau X^i\partial_\sigma\delta X^i - \partial_\tau\delta X^i\partial_\sigma X^i\right) + 2\frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma X^i\partial_\sigma\delta X^i\Big)\end{aligned}$$

which can be rewritten as

$$\begin{aligned}\delta S_P &= \frac{1}{4\pi\alpha'}\int d\tau d\sigma\Big(\left(2\gamma_{\tau\sigma}\partial_\sigma\delta X^-\right) + \left(-2\gamma_{\tau\sigma}\partial_\tau X^i\partial_\sigma\delta X^i + 2\frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma\delta X^i\partial_\sigma X\right)\\ &+\left(\gamma_{\sigma\sigma}2\partial_\tau\delta X^- - 2\gamma_{\sigma\sigma}\partial_\tau X^i\partial_\tau\delta X^i\right) - \left(\gamma_{\tau\sigma}\partial_\tau\delta X^i\partial_\sigma X^i\right)\Big).\end{aligned}$$

After partial integration of the first term this yields once more that $\partial_\sigma\gamma_{\tau\sigma}$ still vanishes at the end of the string.

The second term in parentheses yields after partial integration

$$-\partial_\sigma\left(2\gamma_{\tau\sigma}\partial_\tau X^i\right) + \partial_\sigma\left(\frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma X^i\right) = -2\gamma_{\tau\sigma}\partial_\sigma\partial_\tau X^i + \frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma^2 X^i.$$

Again, this boundary term has to vanish. The second does this, if the derivative of the X^i with respect to σ does so at the boundary, again yielding (9.19). Since this is not the case for the τ-derivative, this again requires $\gamma_{\tau\sigma} = 0$ at the boundary of the string. Hence, this implies that both the function and its first derivative vanishes on the boundary. Because of the equation of motion for $\gamma_{\tau\sigma}$ (9.18), this implies

$$\gamma_{\tau\sigma} \equiv 0,$$

and it can be dropped everywhere.

This eliminates one degree of freedom, leaving only

$$Z^{-}(\tau)\,,\ \gamma_{\sigma\sigma}(\tau)\,,\ X^{i}(\tau,\sigma)\,,$$

which is a rather short list. Furthermore, this simplifies the Polyakov Lagrangian to

$$L_P=-\frac{L}{2\pi\alpha'}\gamma_{\sigma\sigma}\partial_\tau Z^-+\frac{1}{4\pi\alpha'}\int_0^L d\sigma\left(\gamma_{\sigma\sigma}\partial_\tau X^i\partial_\tau X^i-\frac{1}{\gamma_{\sigma\sigma}}\partial_\sigma X^i\partial_\sigma X^i\right),$$

which will now serve as the starting point for quantization. It should be noted that the gauge-fixing was the reason for eliminating the degrees of freedom, reducing the set to a one more manageable for the following.

The first step for quantization is then the calculation of the canonical momenta

$$\begin{aligned}P_- &= -P^+=\frac{\partial L_P}{\partial\left(\partial_\tau Z^-\right)}=-\frac{L}{2\pi\alpha'}\gamma_{\sigma\sigma} &(9.21)\\ \Pi^i &= \frac{\delta L_P}{\delta\left(\partial_\tau X^i\right)}=\frac{1}{2\pi\alpha'}\gamma_{\sigma\sigma}\partial_\tau X^i=\frac{P^+}{L}\partial_\tau X^i.\end{aligned}$$

From this the Hamiltonian is immediately constructed to be

$$\begin{aligned}H &= P_-\partial_\tau Z^-+\int_0^L d\sigma\Pi^i\partial_\tau X^i-L\\ &= \frac{L}{4\pi\alpha' P^+}\int_0^L d\sigma\left(2\pi\alpha'\Pi^i\Pi^i+\frac{1}{2\pi\alpha'}\partial_\sigma X^i\partial_\sigma X^i\right) &(9.22)\end{aligned}$$

This is the Hamiltonian for $D-2$ free fields X^i and the conserved quantity P^+, as can be seen form the equations of motion

$$\begin{aligned}\partial_\tau Z^- &= \frac{\partial H}{\partial P^-}=\frac{H}{P^+}\\ \partial_\tau P^+ &= -\frac{\partial H}{\partial Z^-}=0\\ \partial_\tau X^i &= \frac{\delta H}{\delta\Pi^i}=2\pi\alpha' c\Pi^i &(9.23)\\ \partial_\tau\Pi^i &= -\frac{\delta H}{\delta X^i}=\frac{c}{2\pi\alpha'}\partial_\sigma^2X^i, &(9.24)\end{aligned}$$

where a partial integration has been performed in (9.24) and c is defined as

$$c:=\frac{L}{2\pi\alpha' P^+}.$$

Inserting (9.23) in (9.24) yields the wave equation for X^i

$$\partial_\tau^2X^i=c^2\partial_\sigma^2X^i,$$

where c takes the role of the wave speed. Thus, the transverse degrees of freedom form waves along the string.

Since P^+ and L are constants of motion, so is c. Thus, given the boundary conditions for the open string, the equations of motions can be solved, yielding

$$\begin{aligned}\hat{X}^i(\tau,\sigma) &= \hat{Z}^i + \frac{\hat{P}^i}{P^+}\tau + i(2\alpha')^{\frac{1}{2}} \sum_{n=-\infty,n\neq 0}^{n=\infty} \frac{\alpha_n^i}{n} e^{-\frac{\pi i n c\tau}{L}} \cos\frac{\pi n\sigma}{L} && (9.25)\\ \alpha_{-n}^i &= \alpha_n^{i+}. && (9.26)\end{aligned}$$

The relation (9.26) applies since the X^i are real functions. For the purpose at hand also the center-of-mass variables

$$\begin{aligned}\hat{Z}^i(\tau) &= \frac{1}{L}\int_0^L d\sigma \hat{X}^i(\tau,\sigma)\\ \hat{P}^i(\tau) &= \int_0^L d\sigma \Pi^i(\tau,\sigma) = \frac{P^+}{L}\int_0^L d\sigma\partial_\tau X^i(\tau,\sigma)\end{aligned}$$

have been introduced. Thus, the center-of-mass of the string follows a free, linear trajectory in space, which overlays the transverse motions of the oscillations transverse to the string. Herein $\hat{Z}^i$ and $\hat{P}^i$ in (9.25) have to be taken at $\tau = 0$, and will become Schrödinger operators in the quantization procedure to come now.

The quantization procedure is started as usually with imposing equal-time canonical commutation relations

$$\begin{aligned}\left[Z^-, P^+\right] &= -i\\ \left[X^i(\sigma), \Pi^j(\sigma')\right] &= i\delta^{ij}\delta(\sigma-\sigma')\end{aligned}$$

Performing a Fourier expansion this is equivalent to the relations

$$\begin{aligned}\left[\hat{X}^i, \hat{P}^j\right] &= i\delta^{ij}\\ \left[\alpha_m^i, \alpha_n^j\right] &= m\delta^{ij}\delta_{m,-n} && (9.27)\end{aligned}$$

Here, a non-standard, though useful, normalization of (9.27) has been performed.

The natural consequence is now that every transverse component behaves as a harmonic oscillator with a non-standard normalization. The corresponding creation and annihilation operators are then given for $m > 0$

$$\begin{aligned}\alpha_m^i &= \hbar\sqrt{m}a\\ \alpha_{-m}^i &= \hbar\sqrt{m}a^\dagger\\ -1 &= \left[a^\dagger, a\right] && (9.28)\end{aligned}$$

where m gives the oscillator level for direction i. So far, so standard.

Defining now the momentum vector $k = (k^+, k^i)$ the state $|0,k\rangle$ of lowest excitation has the properties

$$\begin{aligned} P^+ |0,k\rangle &= k^+ |0,k\rangle \\ P^i |0,k\rangle &= k^i |0,k\rangle \\ \alpha^i_m |0,k\rangle &= 0 \text{ for } m > 0 \end{aligned} \tag{9.29}$$

Therefore k is the center-of-mass momentum. Higher excited states are then denoted by $|N,k\rangle$ and can be constructed as

$$|N,k\rangle = \left(\prod_{i=2}^{D-1}\prod_{n=1}^{\infty}\frac{\left(\alpha^i_{-n}\right)^{N_{in}}}{\sqrt{(n^{N_{in}}N_{in})!}}\right)|0,k\rangle\,,$$

just as ordinary oscillator states. Therefore, N_{in} are the occupation numbers for each direction and level. In particular, these can be interpreted as internal degrees of freedom, while the motion of the center-of-mass corresponds to a particle like behavior of the whole string. As will be discussed below, from this point of view every state corresponds to a certain particle with a certain spin.

The total set of states (9.25) forms the Hilbert space of a single string, H_1. In particular, $|0,0\rangle$ is not the vacuum, but merely a momentum-zero string with no internal excitations, except zero-point oscillations: A quantum-mechanical string always quivers. The vacuum is devoid of a string, its Hilbert-space H_0 is denoted by the single state $|\text{vac}\rangle$. However, none of the operators so far can mediate between H_0 and H_1, but only act inside H_1. Since there are no interactions, an N-string Hilbert space can be build just as a product space of H_1s as

$$h_n = |\text{vac}\rangle \oplus H_1 \oplus ... \oplus H_n.$$

where the states are implicitly symmetrized, yielding a Fock space, since the string states are bosonic, given that there creation and annihilation operators fulfill bosonic canonical commutation relations, (9.28).

Since the states are just free states, it is straightforward to construct the number-state version of the Hamiltonian. For this purpose, it is necessary to calculate the explicit form of the canonical momentum operators Π^i first as

$$\begin{aligned} \Pi^i &= \frac{P^+}{L}\left(\partial_\tau X^i\right) \\ &= \frac{P^+}{L}\left(\frac{\hat{P}^i}{P^+} + \frac{\pi c}{L}\left(2\alpha'\right)^{\frac{1}{2}}\sum_{n=-\infty,n\neq 0}^{n=+\infty}\alpha^i_n e^{-\frac{\pi i n c\tau}{L}}\cos\frac{\pi n\sigma}{L}\right). \end{aligned}$$

In addition, also $\partial_\sigma X^i$ is required, and is given by

$$\partial_\sigma X^i = -\frac{i\pi}{L}(2\alpha')^{\frac{1}{2}} \sum_{n=-\infty,n\neq 0}^{n=+\infty} \alpha_n^i e^{-\frac{\pi inc\tau}{L}} \sin\frac{\pi n\sigma}{L}.$$

Putting everything together yields the Hamiltonian

$$\begin{aligned}
& \frac{L}{4\pi\alpha' P^+}\int_0^L d\sigma \left(2\pi\alpha\Pi^i\Pi^i + \frac{1}{2\pi\alpha'}\partial_\sigma X^i \partial_\sigma X^i\right) \\
= & \frac{L}{4\pi\alpha' P^+}\Big(2\pi\alpha' P^i P^i + \int_0^L d\sigma \\
& \Big(\frac{\pi}{4\alpha' L P^+}\sum_{n=-\infty n\neq 0}^{n=+\infty} \alpha_n^i e^{-\frac{\pi inc\tau}{L}} \cos\frac{\pi n\sigma}{L} \sum_{m=-\infty,m\neq 0}^{m=+\infty} \alpha_m^i e^{-\frac{\pi mnc\tau}{L}} \cos\frac{\pi m\sigma}{L} \\
& -\frac{\pi}{4\alpha' L P^+}\sum_{n=-\infty n\neq 0}^{n=+\infty} \alpha_n^i e^{-\frac{\pi inc\tau}{L}} \sin\frac{\pi n\sigma}{L} \sum_{m=-\infty,m\neq 0}^{m=+\infty} \alpha_m^i e^{-\frac{\pi imc\tau}{L}} \sin\frac{\pi m\sigma}{L}\Big)\Big),
\end{aligned}$$

where in the integration σ was replaced by $\pi\sigma/L$. Since sine and cosine are orthogonal, the integrations can be performed explicitly. Those over cos yield $\pi\delta_{n-m}$, while those over sin yield $-\pi\delta_{n-m}$. This leads to

$$H = \frac{P^i P^i}{2P^+} + \frac{1}{2P^+\alpha'}\sum_{n=-\infty n\neq 0}^{n=+\infty} \alpha_n^i \alpha_{-n}^i,$$

and finally by rearranging to

$$H = \frac{P^i P^i}{2P^+} + \frac{1}{2P^+\alpha'}\sum_{n=1}^{\infty} \alpha_{-n}^i \alpha_n^i + A.$$

This Hamiltonian is already in normal order, and A is a (divergent) constant which appears in the process of normal ordering.

The actual value of A can be determined by explicitly verifying the Lorentz covariance of the result, since the Hamiltonian is just the energy, and thus a zero-component of a four-vector. However, in light-cone gauge this is far from trivial, and this will therefore be done here only in a rather sketchy way.

First, consider the zero-point energy. Every oscillator will have a zero-point energy of $\omega/2 = 1/(2P^+\alpha')$, while the transverse momenta P^i will be 0. In total, at zero excitation, it should be expected that

$$\langle 0,0|H|0,0\rangle = A,$$

due to the normal ordering. Due to the non-standard normalization, each oscillator actually contributes $n\omega/2$ of vacuum energy to this value. These oscillations appear for $D-2$

dimensions. Rewriting A as ωA this yields[4]

$$A=\frac{D-2}{2}\sum_{n=1}^{\infty}n,$$

which is, of course, infinite. However, in contrast to normal quantum mechanics or quantum field theory, the vacuum energy is not necessarily irrelevant, but may couple to gravity. It is therefore necessary to maintain Lorentz invariance when treating it, and it cannot be absorbed just in a redefinition of the zero-point energy, as in quantum mechanics.

To regularize the result Lorentz-invariantly, it is necessary to include a cut-off function

$$\begin{aligned} & e^{-\frac{\varepsilon|k_\sigma|}{\sqrt{\gamma_{\sigma\sigma}}}} \\ k_\sigma &= \frac{n\pi}{L}, \end{aligned}$$

in the sum, and taking the limit $\varepsilon\to 0$ only after summation. The factor of $\sqrt{\gamma_{\sigma\sigma}}^{-1}$ is required to maintain the effects of reparametrization invariance correctly. The reason for this is simple. Outside light-cone gauge, the string length is not fixed, but can be changed by a reparametrization. Therefore, k_σ, which depends on the length of the string, changes under such transformations. Including the function of $\gamma_{\sigma\sigma}$ exactly cancels this effect.

Inserting this expression into the sum permits to evaluate it exactly, yielding

$$\begin{aligned} A &= \frac{D-2}{2}\sum_{n=1}^{\infty}ne^{-\frac{\varepsilon|k_\sigma|}{\sqrt{\gamma_{\sigma\sigma}}}} \\ &= \frac{D-2}{2}\left(\frac{2LP^+\alpha'}{\varepsilon^2\pi}-\frac{1}{12}+O(\varepsilon)\right), \end{aligned}$$

where the second line of (9.21) has been used. The first term is proportional to L and can therefore be absorbed in the action by an additional term proportional to

$$-\int d\sigma\,(-\gamma)^{\frac{1}{2}}=-L.$$

This is a constant, and therefore is not changing the action. In fact, the value of the action has to be regularized itself by a similar expression, and also regularized by $e^{-\varepsilon}$. Thus, by appropriately selecting the pre-factors, both terms cancel. Since also the last term vanishes in the limit of $\varepsilon\to 0$, the only thing remaining is

$$A=\frac{2-D}{24}, \tag{9.30}$$

which is known as the Casimir energy, and can be traced back to the fact that the string is only of finite length. Thus, the string has indeed a non-zero vacuum energy. In contrast

[4]With standard normalization, n would be replaced by 1, changing nothing qualitatively.

to the first contribution, this constant, non-divergent term cannot be naturally absorbed by a counter term in the action without spoiling Lorentz invariance.

Having now obtained the Hamiltonian and the state space, it is about time to determine the properties of the physical state space. In particular, the question is whether the string excitations can be interpreted as particle states, the original motivation to study it. For that purpose the primary object is of course whether the states satisfy the energy-momentum relation of a point particle, and if yes, what are their masses.

The corresponding operator for the rest mass is just given by the mass-shell equation, where it is to be used that $P^- = H$ to yield

$$m^2 = 2P^+H - P^iP^i, \tag{9.31}$$

as a result of the light-cone equation

$$m^2 = P^+P^- + P^-P^+ - P^iP^i.$$

Inserting the result (9.30) into (9.31) for the lowest-energy state yields

$$\begin{aligned} m^2 &= 2P^+\left(\frac{P^iP^i}{2P^+} + \frac{1}{2P^+\alpha'}(N+A)\right) - P^iP^i \\ &= \frac{1}{\alpha'}\left(N + \frac{2-D}{24}\right). \end{aligned}$$

That is quite an important result, as it implies that the mass is only dependent on the state sum N defined as

$$N = \sum_{i=2}^{D-1}\sum_{n=1}^{\infty} nN_{in}$$

and the space-time dimensionality D. Thus, mass becomes an intrinsic property rather than an external parameter as in the standard model. The importance of the Regge slope is now also clear, as it links as constant of proportionality the number of a state and its rest mass.

The lowest state is of course $N = 0$, hence $|0, k\rangle$, and this yields

$$m^2 = \frac{2-D}{24\alpha'}$$

Since for any phenomenologically relevant string theory $D > 2$ the rest mass of the lowest state is imaginary, $m^2 < 0$. Thus it is a tachyon. That is of course unfortunate, since interpreting this as a particle is very problematic. E. g., constructing a theory of such a non-interacting scalar tachyon yields a potential energy proportional to $m^2\phi^2/2$. Hence, the vacuum state is unstable. Of course, this would be the lowest approximation, and it could still be that the bosonic string theory is nonetheless stable, but this is unknown so far. Fortunately, in particular in supersymmetric string theories tachyons usually do not appear, so they provide a possibility to circumvent this problem without having to deal with it explicitly.

The first non-tachyonic state is obtained for the state $\alpha_{-1}^i|0,k\rangle$ with $N=1$. Its mass reads

$$m^2 = \frac{26-D}{24\alpha'}. \tag{9.32}$$

Since there are $D-2$ ways to obtain $N=1$, this state is $D-2$-times degenerate. To be still Lorentz invariant, these transverse modes must form a representation of SO$(D-2)$ for a massless particle and SO$(D-1)$ for a massive particle. The former follows because there is no rest-frame for a massless particle, and the minimum momentum is at least $P^\mu = \left(E, E, \vec{0}\right)$, thus having less symmetry then the one for a massive particle in the rest frame being $P^\mu = (m, \vec{0})$.

As a consequence, in $D = 4$ massive bosonic particles have integer spin $j > 0$ as representations of SO(3) with $2j+1$-fold degeneracies. Massless particles, however, are denoted by their helicity forming a representation of the group SO(2), having only one state with positive helicity. Because of CPT symmetry the number of states is actually doubled, since a state with positive helicity can be transformed by CPT into one with negative helicity. Put it in another view, the lowest non-trivial representation of SO(3) is 3-dimensional, a spin-1 state with three magnetic quantum numbers. For SO(2), the lowest non-trivial representation has actually only two possible magnetic quantum numbers, either 1 or -1. However, CPT guarantees that if one exists, then so does the other.

Going back to D dimensions there are thus $D-1$ states for massive bosonic particles, but only $D-2$ for massless ones. Since the degeneracy for the $N=1$ states is $D-2$, this implies that their mass must be zero. From this immediately follows that the theory is only Lorentz-invariant in $D = 26$ dimensions, since otherwise (9.32) would not yield zero. This implies also $A = -1$, due to (9.30).

Hence, this indirect inference yields that the consistency of the string theory with Lorentz and CPT invariance requires a certain number of dimensions, different to quantum field theories, which at least in principle can be formulated in any number of space-time dimensions. Note that this is actually a quantum effect, since only quantization yields the mass-dimension relation (9.32).

A more formal argument will be given below, when it can be done simultaneously for both the open and the closed string, which will be analyzed now.

9.7.4 Closed string spectrum

A closed string is obtained when instead of open boundaries periodic boundaries are imposed. In this case the light-cone gauge conditions become.

$$\begin{aligned}X^\mu(\tau,L) &= X^\mu(\tau,0)\\ \partial_\sigma X^\mu(\tau,L) &= \partial_\sigma X^\mu(\tau,0)\\ \gamma_{ab}(\tau,L) &= \gamma_{ab}(\tau,0)\end{aligned}$$

Similarly, it is then possible to quantize the closed string as the open string. However, this provides another ambiguity, since the zero position of σ can now be anywhere along the string. Consequently, a shift of the zero point is another symmetry of the system as

$$\sigma' = \sigma + s(\tau).$$

To fix it requires another gauge condition, which is conveniently chosen as

$$\gamma_{\tau\sigma}(\tau,0) = 0$$

This implies that lines of constant τ are orthogonal to lines of constant σ at $\sigma = 0$. This reduces the problem to translations about one string length as

$$\sigma' = \sigma + s(\tau) \quad \mathrm{mod}\ L. \tag{9.33}$$

Nonetheless, this is sufficient to start.

Up to the formulation of the Hamiltonian then everything is as for the open string case. Of course, the solutions to the equations of motion are now different, respecting the new boundary conditions. They read

$$X^i(\tau,\sigma) = X^i + \frac{P^i}{P^+}\tau + i\left(\frac{\alpha'}{2}\right)^{\frac{1}{2}} \sum_{n=-\infty n\neq 0}^{\infty}\left(\frac{\alpha_n^i}{n}e^{-\frac{2\pi i n(\sigma+c\tau)}{L}} + \frac{\beta_n^i}{n}e^{\frac{2\pi i n(\sigma-c\tau)}{L}}\right),$$

in analogy to point quantum mechanics of a particle in a periodic box. As a consequence, there are now two independent sets of Fourier coefficients, α and β. These corresponds to oppositely directed waves along the string with α being those running in the left direction and β to the right direction.

Nonetheless, quantization proceeds as usual with the canonical quantization conditions

$$\begin{aligned}\left[Z^-,P^-\right] &= -i\\ \left[X^i,P^i\right] &= i\delta^{ij}\\ \left[\alpha_m^i,\alpha_n^j\right] &= m\delta^{ij}\delta_{m,-n}\\ \left[\beta_m^i,\beta_n^j\right] &= m\delta^{ij}\delta_{m,-n}.\end{aligned}$$

Thus, the system is again that of a set of free oscillators with a superimposed center-of-mass motion. The eigenstates are thus

$$|N,R,k\rangle = \left(\prod_{i=2}^{D-1}\prod_{n=1}^{\infty}\prod_{r=1}^{\infty}\frac{\left(\alpha_{-n}^{i}\right)^{N_{in}}\left(\beta_{-n}^{i}\right)^{R_{in}}}{\left(n^{N_{in}}N_{in}!r^{R_{in}}R_{in}!\right)^{\frac{1}{2}}}\right)|0,0,k\rangle\,.$$

Herein N counts the number of left-moving states and R the number of right-moving states. It is then possible to obtain again the Hamiltonian in number-operator form, and to obtain the mass-shell equation as

$$m^2 = 2P^+H - P^iP^i = \frac{2}{\alpha'}\left(N+R+A+B\right),$$

and in the same way as previously also

$$A = B = \frac{2-D}{24}$$

is obtained.

However, in this case the values of N and R are restricted, since all physical states have to be invariant under the residual gauge freedom (9.33). To see this, the operator for translations on the string is useful. To obtain it, the simplest starting point is the energy-momentum tensor on the world-sheet. It is given by

$$\begin{aligned}
T^{ab} &= -4\pi\left(-\gamma\right)^{-\frac{1}{2}}\frac{\delta L}{\delta\gamma_{ab}} &(9.34)\\
&= -\frac{4\pi}{\left(-\gamma\right)^{\frac{1}{2}}}\frac{\delta}{\delta\gamma_{ab}}\left(-\frac{1}{4\pi\alpha'}\left(-\gamma\right)^{\frac{1}{2}}\gamma^{cd}\partial_cX^\mu\partial_dX_\mu\right)\\
&= \frac{1}{\alpha'\left(-\gamma\right)^{\frac{1}{2}}}\left(\frac{\delta\left(-\gamma\right)^{\frac{1}{2}}}{\delta\gamma_{ab}}\gamma^{cd}\partial_cX^\mu\partial_dX_\mu + \left(-\gamma\right)^{\frac{1}{2}}\frac{\delta\gamma^{cd}}{\delta\gamma_{ab}}\partial_cX^\mu\partial_dX_\mu\right)\\
&= \frac{1}{\alpha'\left(-\gamma\right)^{\frac{1}{2}}}\left(-\frac{1}{2\left(-\gamma\right)^{\frac{1}{2}}}\frac{\delta\gamma}{\delta\gamma_{ab}}\gamma^{cd}\partial_cX^\mu\partial_dX_\mu + \left(-\gamma\right)^{\frac{1}{2}}\partial_aX^\mu\partial_bX_\mu\right)\\
&= \frac{1}{\alpha'\left(-\gamma\right)^{\frac{1}{2}}}\left(-\frac{1}{2\left(-\gamma\right)^{\frac{1}{2}}}\gamma\gamma^{ab}\gamma^{cd}\partial_cX^\mu\partial_dX_\mu + \left(-\gamma\right)^{\frac{1}{2}}\partial_aX^\mu\partial_bX_\mu\right)\\
&= \frac{1}{\alpha'}\left(\partial_aX^\mu\partial_bX_\mu - \frac{1}{2}\gamma^{ab}\gamma^{cd}\partial_cX^\mu\partial_dX_\mu\right)\\
&= \frac{1}{\alpha'}\left(\partial^aX^\mu\partial^bX_\mu - \frac{1}{2}\gamma^{ab}\partial_cX^\mu\partial^cX_\mu\right).
\end{aligned}$$

To argue that this indeed is an energy-momentum tensor[5], it is necessary to show that it has the necessary properties of an energy-momentum tensor, in particular it has to be conserved and traceless, and its $\tau\tau$-component must equal the Hamilton operator.

[5]In quantum field theory this is already a non-obvious fact, lest in string theory.

Start with its conservation. The elements of the energy-momentum tensors appear to be not invariant under diffeomorphisms, since the appearing expressions for γ^{ab} are not, since it seems there are no compensating factor of $\det\gamma$. However, the expression in terms of the Lagrangian is, so there must be a hidden invariance. This is in fact only possible if the energy-momentum tensor is a constant, which would imply its conservation.

Using (9.12), this can be obtained explicitly

$$\begin{aligned}\partial_a T^{ab} &= \frac{1}{\alpha'}\partial_a\left(\partial^a X^\mu\partial^b X_\mu - \frac{1}{2}\gamma^{ab}\partial_c X^\mu\partial^c X_\mu\right)\\ &= \frac{1}{\alpha'}\left(\partial_a h^{ab} - \partial_a\left(\frac{1}{2}\gamma^{ab}\gamma^{cd}h_{cd}\right)\right)\\ &= \frac{1}{\alpha'}\left(\partial_a h^{ab} - \partial_a h^{ab}\right) = 0,\end{aligned}$$

and thus the energy-momentum tensor is conserved.

The next condition is the one of tracelessness. Calculating the trace $T^a_{\ a}$ explicitly yields

$$\begin{aligned}\gamma_{ab}\frac{\delta L}{\delta\gamma_{ab}} &= \frac{1}{(-\gamma)^{\frac{1}{2}}}\left(\gamma_{ab}\partial^a X^\mu\partial^b X_\mu - \partial^c X^\mu\partial_c X_\mu\right)\\ &= \frac{1}{(-\gamma)^{\frac{1}{2}}}\left(\partial^a X^\mu\partial_a X_\mu - \partial^c X^\mu\partial_c X_\mu\right) = 0 \qquad (9.35)\\ &= \gamma_{ab}\frac{T^{ab}}{(-\gamma)^{\frac{1}{2}}},\end{aligned}$$

where it has been used that $\gamma^{ab}\gamma_{ab} = 2$. Finally, this yields

$$T^a_{\ a}\frac{1}{(-\gamma)^{\frac{1}{2}}} = 0,$$

confirming that the energy-momentum tensor is indeed traceless. Incidentally, this shows that the classical energy-momentum tensor vanishes when the equations of motions are fulfilled, by virtue of (9.35) and the fact that the Lagrange function is not depending on the τ-derivatives of γ_{ab}.

Using (9.12), this could also be shown more directly as

$$\begin{aligned}T^a_{\ a} &= \frac{1}{\alpha'}\left(\partial^a X^\mu\partial_a X_\mu - \frac{1}{2}\gamma^a_{\ a}\partial^c X^\mu\partial_c X_\mu\right)\\ &= \frac{1}{\alpha'}\left(\partial^a X^\mu\partial_a X_\mu - \frac{1}{2}\gamma^a_{\ a}\gamma_{cd}\partial^c X^\mu\partial^d X_\mu\right)\\ &= \frac{1}{\alpha'}\left(\partial^a X^\mu\partial_a X_\mu - \partial^a X^\mu\partial_a X_\mu\right) = 0,\end{aligned}$$

and thus the same result.

Finally, the $\tau\tau$ component should be the Hamiltonian. To show this, it is simpler to go backwards. By reexpressing the Hamiltonian (9.22) as a function of $\partial_\sigma X^\mu$ and $\partial_\tau X^\mu$ it becomes

$$\begin{aligned} H &= \frac{L}{4\pi\alpha' P^+}\int_0^L d\sigma\left(2\pi\alpha'\Pi^i\Pi^i + \frac{1}{2\pi\alpha'}\partial_\sigma X^i\partial_\sigma X^i\right)\\ &= \frac{L}{4\pi\alpha' P^+}\int_0^L d\sigma\left(2\pi\alpha'\frac{P^+}{L}\frac{P^+}{L}\partial_\tau X^i\partial_\tau X^i + \frac{1}{2\pi\alpha'}\partial_\sigma X^i\partial_\sigma X^i\right). \end{aligned}$$

Using now (9.21) changes this to

$$\begin{aligned} H &= \frac{1}{4\pi\alpha'}\int_0^L d\sigma\left(2\pi\alpha'\frac{P^+}{L}\partial_\tau X^i\partial_\tau X^i + \frac{1}{2\pi\alpha'}\frac{L}{P^+}\partial_\sigma X^i\partial_\sigma X^i\right)\\ &= \frac{1}{4\pi\alpha'}\int_0^L d\sigma\left(2\pi\alpha'\frac{\gamma_{\sigma\sigma}}{2\pi\alpha'}\partial_\tau X^i\partial_\tau X^i + \frac{1}{2\pi\alpha'}\frac{2\pi\alpha'}{\gamma_{\sigma\sigma}}\partial_\sigma X^i\partial_\sigma X^i\right)\\ &= \frac{1}{4\pi\alpha'}\int_0^L d\sigma\left(\gamma_{\sigma\sigma}\partial_\tau X^\mu\partial_\tau X_\mu + \frac{1}{\gamma_{\sigma\sigma}}\partial_\sigma X^\mu\partial_\sigma X_\mu\right). \end{aligned}$$

The expansion of i to μ in the last line was permitted because this is only an addition of zero in the second term and also a zero in the first term by virtue of the boundary conditions after exchange of integration and differentiation.

To bring the $\tau\tau$ component of the energy-momentum tensor into the same form it can be expressed as

$$\begin{aligned} T^{\tau\tau} &= \frac{1}{\alpha'}\left(\partial^\tau X^\mu\partial^\tau X_\mu - \gamma^{\tau\tau}\partial_\tau X^\mu\partial^\tau X_\mu - \gamma^{\tau\tau}\partial_\sigma X^\mu\partial^\sigma X_\mu\right)\\ &= \frac{1}{\alpha'}\Big(\partial^\tau X^\mu\partial^\tau X_\mu - \frac{1}{2}\left(\gamma^{\tau\tau}\gamma_{\tau\tau}\right)\partial_\tau X^\mu\partial_\tau X_\mu\\ &\quad -\frac{1}{2}\gamma^{\tau\tau}\gamma_{\tau\sigma}\partial_\tau X^\mu\partial_\sigma X_\mu - \frac{1}{2}\gamma^{\tau\tau}\gamma_{\tau\sigma}\partial_\sigma X^\mu\partial_\tau X_\mu - \frac{1}{2}\gamma^{\tau\tau}\gamma_{\sigma\sigma}\partial_\sigma X^\mu\partial_\sigma X_\mu\Big). \end{aligned}$$

Because of the gauge condition $\gamma^{\tau\tau}$ and $-\gamma^{\sigma\sigma}$ are related, and yielding that the square of $\gamma^{\tau\tau}$ is -1, because otherwise the gauge condition for the determinant would be violated, given that $\gamma_{\tau\sigma}$ vanishes. This yields

$$\begin{aligned} T^{\tau\tau} &= \frac{1}{\alpha'}\left(\frac{1}{2}\partial_\tau X^\mu\partial_\tau X_\mu + \frac{1}{2}\partial_\sigma X^\mu\partial_\sigma X_\mu\right)\\ &= \frac{1}{2\alpha'}\left(\gamma^{\sigma\sigma 2}\partial_\tau X^\mu\partial_\tau X_\mu + \partial_\sigma X^\mu\partial_\sigma X_\mu\right)\\ &= \frac{1}{2\alpha'}\gamma^{\sigma\sigma}\left(\gamma^{\sigma\sigma}\partial_\tau X^\mu\partial_\tau X_\mu + \frac{1}{\gamma^{\sigma\sigma}}\partial_\sigma X^\mu\partial_\sigma X_\mu\right), \end{aligned}$$

which concludes

$$H = -\frac{1}{2\pi}\int_0^L d\sigma\gamma^{\sigma\sigma}T^{\tau\tau}$$

where the factor $\gamma^{\sigma\sigma}$ is actually part of the measure to make the expression diffeomorphism invariant, and thus shows the correct relation between the Hamiltonian and the energy-momentum tensor.

Hence, it is permitted to use this expression for the energy-momentum tensor to obtain the operator of linear translation. It is given by the $\sigma\tau$ component, as in case of classical mechanics. Since $\gamma_{\tau\sigma} = 0$ this component is given by

$$T^{\sigma\tau} = \frac{1}{\alpha'}\left(\partial^\sigma X^\mu \partial^\tau X_\mu\right) = 2\pi c \Pi^i \partial_\sigma X^i,$$

since the $\pm$ have a vanishing σ component each. Integrating yields the operator as

$$\begin{aligned} P &= -\int_0^L d\sigma \Pi^i \partial_\sigma X^i \\ &= \frac{2\pi}{L}\left(\sum_{n=1}^{\infty}\left(\alpha_{-n}^i \alpha_n^i - \beta_{-n}^i \beta_n^i\right) + A - B\right) \\ &= \frac{2\pi}{L}\left(N - R\right). \end{aligned}$$

The residual gauge freedom is essentially giving that the coordinates hop around the string by an integer times L, permitting to turn left-moving into right-moving modes. This can be restricted by enforcing

$$N = R. \tag{9.36}$$

Thus, the expectation value of translations along the string is zero, and any physical state has a localized coordinate system on the string. With other words, the number of left and right moving modes must be the same.

The lowest state is given again by

$$m^2 = \frac{2}{\alpha'} 2 \frac{2-D}{24} = \frac{2-D}{6\alpha'},$$

and is therefore again a tachyon. The lowest excited state is given by $|1, 1, k\rangle$

$$m^2 = \frac{26-D}{6\alpha'}.$$

However, in contrast to the previous case, it is not constructed by a single creation operator with just one space-time index, but by two as

$$|1, 1, k\rangle = \alpha_{-1}^i \beta_{-1}^j |0, 0, k\rangle,$$

and therefore is a tensor state e_{ij}. As in the case of large extra dimensions, this state can be separated as

$$e^{ij} = \frac{1}{2}\left(e^{ij} + e^{ji} - \frac{2}{D-2}\delta^{ij} e^{kk}\right) + \frac{1}{2}\left(e^{ij} - e^{ji}\right) + \frac{1}{D-2}\delta^{ij} e^{kk}.$$

The first term is traceless symmetric, the second antisymmetric and the third scalar. Furthermore, the occupation numbers N_{in} and R_{in} can vary freely as long as $N = R$ is fulfilled. Therefore, the number of states is substantially increased with respect to the open string spectrum at the same N. Whether it is necessarily massless, and thus again $D = 26$, is not a trivial question, but will turn out to be correct. This time, the helicity of the state will be useful to show this will yield a graviton, an axion, and a dilaton.

To verify the assignment of spin, a little more formal investigation is useful. Note that it is always possible to obtain a spin algebra from creation and annihilation operators, when summing over oscillators, called the Schwinger representation. In case of the open string, the corresponding operators are given by

$$S^{ij} = -i\sum_{n=1}^{\infty}\frac{1}{n}\left(\alpha^i_{-n}\alpha^j_n - \alpha^j_{-n}\alpha^i_n\right) \tag{9.37}$$

and the ones for the closed string are completely analogous, just requiring that it is now necessary to sum over both, left-moving and right-moving modes. The two indices already indicate that these will be the corresponding n-dimensional generalization of the spin.

That (9.37) are indeed spin operators can be shown by explicitly calculating the corresponding algebra. Start by evaluating the commutator as

$$\begin{aligned}
\left[S^{ij}, S^{kl}\right] &= -\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{nm}\left[\alpha^i_{-n}\alpha^j_n - \alpha^j_{-n}\alpha^i_n, \alpha^k_{-m}\alpha^l_m - \alpha^l_{-m}\alpha^k_m\right]\\
&= -\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{nm}(\left[\alpha^i_{-n}\alpha^j_n, \alpha^k_{-m}\alpha^l_m - \alpha^l_{-m}\alpha^k_m\right]\\
&\quad - \left[\alpha^j_{-n}\alpha^i_n, \alpha^k_{-m}\alpha^l_m - \alpha^l_{-m}\alpha^k_m\right])\\
&= -\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{nm}(\left[\alpha^i_{-n}\alpha^j_n, a^k_{-m}\alpha^l_m\right] - \left[\alpha^i_{-n}\alpha^j_n, \alpha^l_{-m}\alpha^k_m\right]\\
&\quad - \left[\alpha^j_{-n}\alpha^i_n, \alpha^k_{-m}\alpha^l_m\right] + \left[\alpha^j_{-n}\alpha^i_n, \alpha^l_{-m}\alpha^k_m\right]).
\end{aligned}$$

It is simpler to evaluate each of the four terms individually. For this the relation

$$[ab, c] = a\,[b, c] + [a, c]\,b$$

for double commutators is quite useful, as well as the quantization conditions (9.27) are necessary. In the following the summation is kept implicit. This yields for the first term

$$\begin{aligned}
\left[\alpha^i_{-n}\alpha^j_n, \alpha^k_{-m}\alpha^l_m\right] &= \alpha^i_{-n}\left[\alpha^i_n, \alpha^k_{-m}\alpha^l_m\right] + \left[\alpha^i_{-n}, \alpha^k_{-m}\alpha^l_m\right]\alpha^j_n \\
&= \alpha^i_{-n}\alpha^k_{-m}\left[\alpha^j_n, \alpha^l_m\right] + \alpha^i_{-n}\left[\alpha^j_n, \alpha^k_{-m}\right]\alpha^l_m \\
&\quad +\alpha^k_{-m}\left[\alpha^i_{-n}, \alpha^l_m\right]\alpha^i_n + \left[\alpha^i_{-n}, \alpha^k_{-m}\right]\alpha^l_m\alpha^j_n \\
&= \alpha^i_{-n}\alpha^k_{-m}n\delta^{jl}\delta_{n,-m} + \alpha^i_{-n}\alpha^l_m n\delta^{jk}\delta_{n,m} \\
&\quad -\alpha^k_{-m}\alpha^j_n n\delta^{il}\delta_{-n,-m} - \alpha^l_m\alpha^j_n n\delta^{ik}\delta_{-n,m} \\
&= n(\alpha^i_{-n}\alpha^k_n\delta^{jl} + \alpha^i_{-n}\alpha^l_n\delta^{jk} - \alpha^k_{-n}\alpha^j_n\delta^{il} - \alpha^l_{-n}\alpha^j_n\delta^{ik}), \qquad (9.38)
\end{aligned}$$

for the second term

$$\begin{aligned}
\left[\alpha^i_{-n}\alpha^j_n, \alpha^l_{-m}\alpha^k_m\right] &= \alpha^i_{-n}\left[\alpha^j_n, \alpha^l_{-m}\alpha^k_m\right] + \left[\alpha^i_{-n}, \alpha^l_{-m}\alpha^k_m\right]\alpha^j_n \\
&= \alpha^i_{-n}\alpha^l_{-m}\left[\alpha^j_n, \alpha^k_m\right] + \alpha^i_{-n}\left[\alpha^j_n, \alpha^l_{-m}\right]\alpha^k_m \\
&\quad +\alpha^l_{-m}\left[\alpha^i_{-n}, \alpha^k_m\right]\alpha^j_n + \left[\alpha^i_{-n}, \alpha^l_{-m}\right]\alpha^k_m\alpha^j_n \\
&= \alpha^i_{-n}\alpha^l_{-m}n\delta^{jk}\delta_{n,-m} + \alpha^i_{-n}\alpha^k_m n\delta^{jl}\delta_{n,m} \\
&\quad -\alpha^l_{-m}\alpha^j_n n\delta^{ik}\delta_{-n,-m} - \alpha^k_m\alpha^j_n n\delta^{il}\delta_{-n,m} \\
&= n(\alpha^i_{-n}\alpha^l_n\delta^{jk} + \alpha^i_{-n}\alpha^k_n\delta^{jl} - \alpha^l_{-n}\alpha^j_n\delta^{ik} - \alpha^k_{-n}\alpha^j_n\delta^{il}), \qquad (9.39)
\end{aligned}$$

for the third term

$$\begin{aligned}
\left[\alpha^j_{-n}\alpha^i_n, \alpha^k_{-m}\alpha^l_m\right] &= \alpha^j_{-n}\left[\alpha^i_n, \alpha^k_{-m}\alpha^l_m\right] + \left[\alpha^j_{-n}, \alpha^k_{-m}\alpha^l_m\right]\alpha^i_n \\
&= \alpha^j_{-n}\alpha^k_{-m}\left[\alpha^i_n, \alpha^l_m\right] + \alpha^j_{-n}\left[\alpha^i_n, \alpha^k_{-m}\right]\alpha^l_m \\
&\quad +\alpha^k_{-m}\left[\alpha^j_{-n}, \alpha^l_m\right]\alpha^i_n + \left[\alpha^j_{-n}, \alpha^k_{-m}\right]\alpha^l_m\alpha^i_n \\
&= \alpha^j_{-n}\alpha^k_{-m}n\delta^{il}\delta_{n,-m} + \alpha^j_{-n}\alpha^l_m n\delta^{ik}\delta_{n,m} \\
&\quad -\alpha^k_{-m}\alpha^i_n n\delta^{jl}\delta_{-n,-m} - \alpha^l_m\alpha^i_n n\delta^{jk}\delta_{-n,m} \\
&= n(\alpha^j_{-n}\alpha^k_n\delta^{il} + \alpha^j_{-n}\alpha^l_n\delta^{ik} - \alpha^k_{-n}\alpha^i_n\delta^{jl} - \alpha^l_{-n}\alpha^i_n\delta^{jk}), \qquad (9.40)
\end{aligned}$$

and finally the fourth

$$\begin{aligned}
\left[\alpha^j_{-n}\alpha^i_n, \alpha^l_{-m}\alpha^k_m\right] &= \alpha^j_{-n}\left[\alpha^i_n, \alpha^l_{-m}\alpha^k_m\right] + \left[\alpha^j_{-n}, \alpha^l_{-m}\alpha^k_m\right]\alpha^i_n \\
&= \alpha^j_{-n}\alpha^l_{-m}\left[\alpha^i_n, \alpha^k_m\right] + \alpha^j_{-n}\left[\alpha^i_n, \alpha^l_{-m}\right]\alpha^k_m \\
&\quad +\alpha^l_{-m}\left[\alpha^j_{-n}, \alpha^k_m\right]\alpha^i_n + \left[\alpha^j_{-n}, \alpha^l_{-m}\right]\alpha^k_m\alpha^i_n \\
&= \alpha^i_{-n}\alpha^l_{-m}n\delta^{ik}\delta_{n,-m} + \alpha^j_{-n}\alpha^k_m\delta^{ij}\delta_{n,m} \\
&\quad -\alpha^l_{-m}\alpha^i_n n\delta^{jk}\delta_{-n,-m} - \alpha^k_m\alpha^i_n n\delta^{jl}\delta_{-n,m} \\
&= n(\alpha^j_{-n}\alpha^l_n\delta^{ik} + \alpha^j_{-n}\alpha^k_n\delta^{ij} - \alpha^l_{-n}\alpha^i_n\delta^{ik} - \alpha^k_{-n}\alpha^i_n\delta^{jl}). \qquad (9.41)
\end{aligned}$$

Combining (9.38-9.41) permits to drop the summation over m. In addition, for every δ each term appears twice, reducing the total expression to

$$\begin{aligned}\left[S^{ij},S^{kl}\right] &= -2\sum_{n=1}^{\infty}\frac{1}{n}(\alpha^i_{-n}\alpha^k_n\delta^{jl}+\alpha^i_{-n}\alpha^l_n\delta^{jk}+\alpha^j_{-n}\alpha^k_n\delta^{il}+\alpha^j_{-n}\alpha^l_n\delta^{ik}\\ &\quad -\alpha^k_{-n}\alpha^j_n\delta^{il}-\alpha^l_{-n}\alpha^j_n\delta^{ik}-\alpha^k_{-n}\alpha^i_n\delta^{jl}-\alpha^l_{-n}\alpha^i_n\delta^{jk}).\end{aligned}$$

Reordering, expanding -1 to i^2, and combing terms with the same δ permits to reconstruct spin operators. Finally, the result becomes

$$\left[S^{ij},S^{kl}\right]=2i\left(\delta^{jl}S^{ik}+\delta^{jk}S^{il}+\delta^{il}S^{jk}+\delta^{ik}S^{il}\right).$$

Thus, indeed the operators satisfy a spin algebra. If all indices are different then the commutator vanishes. Since furthermore all diagonal elements of the S^{ij} vanish only elements with the same indices remain. For example this leaves

$$\left[S^{12},S^{23}\right]=2iS^{13}.$$

The commutator hence contains always the two unequal indices in the same order. Since the spin operator is antisymmetric by definition also the correct exchange property for the arguments of the commutator is obtained, completing the construction.

To see how the helicity emerges investigate first the 23 component of the spin operator, being the one relevant in a four-dimensional sub-space. The helicity of the lowest excitation of the open string is then given by

$$\langle 1,k|\,S^{23}\,|1,k\rangle=-i\sum_{n=1}^{\infty}\frac{1}{n}\langle 1,k|\left(\alpha^2_{-n}\alpha^3_n-\alpha^3_{-n}\alpha^2_n\right)|1,k\rangle=\langle 1,k|\,i\,|1,k\rangle=i.$$

Thus the value is 1. For the lowest excitation of the closed string, the value is found analogously to be two. Thus the lowest excitation of the open string is a vector particle while the one of the closed string is rather a graviton, in accordance with the previous considerations.

Comparing all results a number of interesting observations are obtained. Since vector particles always harbor a gauge symmetry, the open string already furnishes a gauge theory. Since it is non-interacting, this gauge theory has to be non-interacting as well, leaving only a U(1) gauge theory. A more detailed calculation would confirm this. Therefore, it is admissible to call the state $|1,k\rangle$ a photon.

Similarly, a spin 2 particle couples to a conserved tensor current. Since the only one available is the energy-momentum tensor, the symmetric contribution of the lowest excitation of the closed string can be interpreted as a graviton. The antisymmetric particle can be given the meaning of an axion, as it is equivalent to a 2-form gauge boson. Finally, the scalar particle is then the dilaton, as in the case of large extra dimensions.

Calculating the helicity gives already the correct result for the photon and the graviton. Indeed, for the axion and the dilaton a value of zero is obtained, as they would have also in a generic quantum field theory of these particles.

It should be noted that it can be shown that a string theory turns out to be only consistent if it at least contains the closed string, with the open string being an optional addition. Thus, the graviton is there in any string theory.

9.7.4.1 Dualities

It could be easily imagined that there are many different string theories, like there are many different field theories. However, the number of consistently quantizable string theories is very limited, and only five are known today. Furthermore, it can be shown that these string theories are dual to each other.

To get an idea of the concept of dualities, note the following. The Polyakov action (9.11) can also be viewed with a different interpretation: Promoting the world-sheet indices to space-time indices and taking the indices μ to label internal degrees of freedom, then the Polyakov action just describes D massless scalar fields X_μ (with internal symmetry group SO($D-1$,1)) in two space-time dimensions with a non-trivial metric γ, which is dynamically coupled to the fields. This is an example of a duality of two theories. This also demonstrates why two-dimensional field theories have played a pivotal role in understanding string theories. Another such relation is the AdS/CFT correspondence, which state that certain classical (super)gravity theories on a so-called anti de Sitter space, a special case of a curved space-time, are dual to (super)conformal field theories.

A more typical example for a string theory is the following. Start with the closed string, now also periodic in the eigentime. The condition (9.36) implies that any solution for the open string has the form

$$X^\mu = X_L^\mu + X_R^\mu.$$

It is convenient to write the left-moving and right-moving solutions for the following as

$$\begin{aligned}
X_L^\mu &= \frac{x^\mu}{2} + \frac{L^2 p^\mu}{2}(\tau-\sigma) + \frac{iL}{2}\sum_{n\neq 0}\frac{\alpha_n^\mu}{n}e^{-2in(\tau-\sigma)} \\
X_R^\mu &= \frac{x^\mu}{2} + \frac{L^2 p^\mu}{2}(\tau+\sigma) + \frac{iL}{2}\sum_{n\neq 0}\frac{\beta_n^\mu}{n}e^{-2in(\tau+\sigma)},
\end{aligned}$$

where x^μ and p^μ are the position and momentum of the center of mass, and the α and β are the Fourier coefficients of the excited modes.

Now compactify one of the dimensions on a circle of radius R. This will only affect the zero-modes, so for the following the sum of excited states is dropped. Also, only the functions in the direction of the compactified dimension, say number 25, are affected. Assume that the string is warped W times around the compact dimensions. The left-moving and right-moving solutions for the ground-state take then the form

$$\begin{aligned} X_L^{25} &= \frac{x^{25}+c}{2} + (\alpha p^{25} + WR)(\tau+\sigma) \\ X_R^{25} &= \frac{x^{25}-c}{2} + (\alpha p^{25} - WR)(\tau-\sigma) \\ X^{25} &= x^{25} + \frac{2\alpha K}{R}\tau + 2WR\sigma. \end{aligned}$$

where c is an arbitrary constant which just turns the zero-point of the world-sheet coordinates around the compactified dimensions. Also, the center of mass momentum is then quantized as in the large extra dimension theories, and given by

$$p^{25} = \frac{K}{R}$$

where K is describing the Kaluza-Klein mode, and is thus enumbered by an integer.

Now a duality transformation can be performed by mapping $W \to K$ and $R \to \alpha/R$. Then the zero-mode takes the form

$$X^{25} = x^{25} + 2WR\tau + 2\frac{2\alpha K}{R}\sigma.$$

However, this is exactly the expression which would be obtained if a string would wind K times around a compact dimension of size α/R for the Wth Kaluza-Klein mode. Since these parameters only appear in the zero mode, the remaining part of the solution is the same. Hence, these two theories have the same solutions under this mapping of the parameters, they are dual to each other. Such relations are called duality relations. In general, when the exact solutions are not known like in the present case, it is much harder to establish the duality of two theories. In particular, a duality in a classical theory could be broken by quantum effects. Therefore, most dualities so far have only been conjectured on the basis that no counter-example for them is known.

Since all known, consistent quantum string theories are dual to each other, the idea that there is a common underlying structure, the mentioned M-theory, is very appealing, though unproven.

9.8 Virasaro algebra

9.8.1 The algebra

The property of being consistent only in a certain number of dimensions can be linked to an algebraic structure, the Virasaro algebra. For this, it is useful to not use a particular gauge, but rather a more general setting. For the following, this essentially boils down to use instead of the canonical commutation relations (9.27)

$$[\alpha_m^\mu, \alpha_n^\nu] = m\delta_{m+n}\eta^{\mu\nu} \tag{9.42}$$

and thus to permit quantized oscillations in all directions. Of course, this is to be expected: These oscillations are the same as in light cone gauge, as the remainder coordinate functions are completely determined by the reduced set of spatial directions due to the present symmetries, and therefore were not needed to be given explicitly.

The starting point for the construction of the algebra is then the Fourier expansion of the diagonal elements of the world-sheet energy momentum tensor. For this purpose, it is useful to set the string length to 2π, to avoid a proliferation of factors of L. Classically, for the open string, its is defined as

$$T_{aa} = \alpha' \sum_{n=-\infty}^{\infty} L_n e^{-in\xi^a},$$

(no summation over a implied) where the L_n are the expansion modes and the ξ^a are the momenta along the directions σ and τ on the world-sheet. Because of the two different movement directions on the closed string, the modes for the τ and σ directions are different,

$$\begin{aligned} T_{\tau\tau} &= 4\alpha' \sum_{n=-\infty}^{\infty} \tilde{L}_n e^{-in\xi^\tau} \\ T_{\sigma\sigma} &= 4\alpha' \sum_{n=-\infty}^{\infty} L_n e^{-in\xi^\sigma}. \end{aligned}$$

These modes can be expressed in terms of the Fourier coefficients α as

$$L_m = \frac{1}{2\pi\alpha'} \int_0^{2\pi} d\sigma e^{im\sigma} T_{\tau\tau}\bigg|_{\tau=0} = \frac{1}{2\pi\alpha'} \int_0^{2\pi} d\sigma e^{-im\sigma} T_{\sigma\sigma}\bigg|_{\tau=0} = \frac{1}{2} \sum_{n=-\infty}^{\infty} \alpha_{m-n}^\mu \alpha_n^\mu$$

for the open string and

$$\begin{aligned} L_m &= \frac{1}{2} \sum_{n=-\infty}^{\infty} \alpha_{m-n}^\mu \alpha_n^\mu \\ \tilde{L}_m &= \frac{1}{2} \sum_{n=-\infty}^{\infty} \beta_{m-n}^\mu \beta_n^\mu \end{aligned}$$

for the closed string. Note that the energy momentum tensor vanishes by being the equation of motion (9.34) of a cyclic variable. From the vanishing of the energy momentum tensor then follows $L_m = \tilde{L}_m = 0$ for all m, the so-called Virasaro constraints. Since L_m is not differing between open and closed strings, it will not be differentiated in the following between both, except for the presence or absence of the second mode $\tilde{L}_m$.

When now quantizing the system, there appears an ordering problem for L_0, as α_{-n} is not commuting with α_n, see (9.42). Thus, an ambiguity arises, and therefore the quantum version of L_0 and $\tilde{L}_0$ are defined as

$$L_0 = \frac{\alpha_0^2}{2} + \frac{1}{2}\sum_{n=1}^{\infty} \alpha_{-n}\alpha_n = a + \frac{1}{2}\sum_{n=1}^{\infty} \alpha_{-n}\alpha_n,$$

and similarly for $\tilde{L}_0$. The constant a can be determined when observing that the mass operator m^2, defined to be the Hamiltonian minus P^2, is given by

$$m^2 = \frac{2}{\alpha'}\sum_{n=1}^{\infty} \alpha_n\alpha_{-n} = -\frac{1}{\alpha'}\left(a - \sum_{n=1}^{\infty} \alpha_{-n}\alpha_n\right),$$

since the same operator ordering problem arises. Since the mass is invariant under the gauge choice the value of a can be read off (9.30), as then $-A = a = 1$.

The Virasaro algebra is now given by the algebra of the operators L_m. For $m + n \neq 0$, it can be straightforwardly, albeit tediously, shown that

$$[L_m, L_n] = (m - n)L_{m+n},$$

using the canonical commutator relations for the αs (9.27). However, it is more complicated if $m + n = 0$. It is direct to show that for any m

$$[L_m, \alpha_n^\mu] = -n\alpha_{m+n}^\mu. \tag{9.43}$$

holds. The commutator is now given by

$$\begin{aligned}
[L_m, L_n] &= \frac{1}{2}\left(\sum_{p=-\infty}^{-1}\left((m-p)\alpha_p^\mu\alpha_{m+n-p}^\mu + p\alpha_{n+p}^\mu\alpha_{m-p}^\mu\right)\right. \\
&\quad \left. + \sum_{p=0}^{\infty}\left(p\alpha_{m-p}^\mu\alpha_{n+p}^\mu + (m-p)\alpha_{m+n-p}^\mu\alpha_p^\mu\right)\right) \\
&= \frac{1}{2}\left(\sum_{p=-\infty}^{-1}(m-p)\alpha_p^\mu\alpha_{m+n-p}^\mu + \sum_{p=-\infty}^{n-1}(p-n)\alpha_p^\mu\alpha_{n+m-p}^\mu\right. \\
&\quad \left. + \sum_{p=n}^{\infty}(p-n)\alpha_{n+m-p}^\mu\alpha_p^\mu + \sum_{p=0}^{\infty}(m-p)\alpha_{m+n-p}^\mu\alpha_p^\mu\right)
\end{aligned}$$

Now it remains to bring the terms all in the same order as necessary for the definitions of the L_m. This is again a somewhat tedious exercise, and ultimately yields

$$[L_m, L_n] = (m-n)L_{m+n} + \frac{d}{12}(m^3 - m)\delta_{m+n}$$

where the last term is called the central extension of the algebra.

9.8.2 Physical states

One of the main advantages of the Virasaro algebra is to permit a simple identification of physical states, and to check that only physical states of a string theory contribute to observables. As in quantum mechanics and in quantum field theory, a state p is considered to be physical if it has a positive norm and positive semi-definite inner product with other physical states q,

$$\begin{aligned} \langle p|p\rangle &> 0 \\ |\langle p|q\rangle|^2 &\geq 0. \end{aligned}$$

There may exist other states in a theory. One such class are states with zero inner product with any physical state p, so-called spurious states,

$$\langle p|z\rangle = 0, \tag{9.44}$$

These spurious state then do not contribute to any observable. What is not permitted are states with negative norm or overlaps, so-called ghost states g, as these would spoil any probability interpretation of the theory.

Physical states can now be shown to behave as

$$L_{m>0}|p\rangle = 0 \tag{9.45}$$

$$(L_0 - a)|p\rangle = 0, \tag{9.46}$$

while spurious states obey besides the second condition (9.46) also (9.44) for all physical states. The correctness of this assignment follows from the fact that the conditions (9.45) and (9.46) can be shown to correspond to the vanishing of the quantized world-sheet energy momentum tensor, and thus imply the satisfaction of the equations of motion.

Since the adjoint of L_m is L_{-m}, spurious states can be written as

$$|z\rangle = \sum_{n>0} L_{-n}|\chi_n\rangle,$$

where the χ_n satisfy

$$(L_0 - a + n)|\chi_n\rangle = 0.$$

This implements both conditions for spurious states (9.46) and (9.44) by construction. Since for $m < -2$ the L_m can be rewritten, using the Virasaro algebra, in terms of L_{-1} and L_{-2}, this can be simplified to

$$|z\rangle = L_{-1}|\chi_1\rangle + L_{-2}|\chi_2\rangle.$$

A state can be both physical and spurious. By construction, it follows that such states have zero scalar product with any physical states including themselves, i. e., they have zero inner norm. Such states are called null states.

Such null states n can be constructed using spurious states of the form

$$|n\rangle = L_{-1}|\chi_1\rangle.$$

Such a state fulfills all conditions of being physical, except

$$L_1|\chi_1\rangle = L_1 L_{-1}|\chi_1 >= 2L_0|\chi_1\rangle = 2(a-1)|\chi_1\rangle, \tag{9.47}$$

using the Virasaro algebra. Only since $a = 1$, the state is physical. Given the definition of L_{-1}, it actually follows that $|\chi\rangle = |0, k\rangle$, i. e. the state where the string has no internal excitations. Incidentally, this implies that the tachyon is not a physical state. Furthermore, this implies that any physical state is actually an equivalence class of states

$$|p\rangle \sim |p\rangle + |n\rangle,$$

as no measurement can differentiate between the original state and the one where an arbitrary zero norm state has been added. In fact, an infinite number of such null states can be constructed. These are required to cancel in any physical process contribution from negative norm states, the ghost states, very similar to the situation in gauge theories. This is, however, beyond the present scope. In fact, in light-cone gauge such states do not arise, implying that the theory is well-defined. the reason for their appearance here is that the covariant formulation is not fully fixing the reparametrization invariance, as it explicitly contains unphysical degrees of freedom, the additional $X^{\pm}$, just like in an ordinary gauge theory.

Index