Beyond the Standard Model

Lecture in SS 2018 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. Details about the Higgs can be found in the lecture on 'Electroweak Physics' from SS 2016.

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, little Higgs 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. It will not be discussed here, but deserves its own lecture. An introduction is given by the corresponding lecture notes on supersymmetry from SS 2016. 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

- Andersen et al., “Discovering technicolor”, 1104.1255, 2011
- Bambi et al. “Introduction to particle cosmology”, Springer, 2016
- Bedford, “An introduction to string theory”, 1107.3967, 2011
- Böhm, Denner, and Joos, “Gauge theories”, Teubner, 2001
- Cheng, “Introduction to extra dimensions”, 1003.1162, 2010
- Dolgov, “Cosmology and physics beyond the standard model”, Cosmology and Gravitation, American Institute of Physics, 2007
- Han, Lykken, and Zhang, “Kaluza-Klein states from large extra dimensions”, hep-ph/98113504, 2000
- Hill and Simmons, “Strong dynamics and electroweak symmetry breaking”, hep-ph/0203079, 2003
- Lane, “Two lectures on technicolor”, hep-ph/0202255, 2002
- Maas, “Brout-Englert-Higgs physics: From foundations to phenomenology”, 1712.04721
- Morrissey, Plehn, and Tait “New physics at the LHC”, 0912.3259, 2009
- Piai “Lectures on walking technicolor, holography, and gauge/gravity dualities”, 1004.0176, 2010
- Polchinski, “String theory”, Cambridge University Press, 1998
- Rovelli, “A dialog on quantum gravity”, hep-th/0310077, 2003
- Zee, “Quantum field theory in a nutshell”, Princeton University Press, 2010

However, the topic is rapidly developing, and will be even more so as soon as something is found at the LHC. 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 should be covered.

Note that this lecture has necessarily some overlap with the astroparticle physics lecture, as both 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 flood in on an almost daily 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[1]. 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 chapter 7 on more recent theoretical developments, where also some other less conventional ideas are collected.

[1]I prepare the lectures notes usually a few months in advance.

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

[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 lectures on electroweak physics (SS 2016) and hadron physics (SS 2017). A phenomenological introduction, also to some of the topics to be discussed here in more detail, can be found in the lecture on particle physics from WS 2015/2016.

distinguished by their mass. The second family contains the strange quark, with a mass 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 4.5 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 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

[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 4, see e. g. O'Raifeartaigh "Group structure of gauge theories", Cambridge, 1986.

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

condensate the particles directly interacting with the Higgs boson acquire a mass, i. e. 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 4. 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 7.5.

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

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

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- 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, 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

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. A running quantity is defined via some process. E. g., a three-point coupling g can be defined to be the three-point correlation function at a symmetric momentum configuration, e. g.

$$\Gamma^{A\bar{\psi}\psi}(\mu^2,\mu^2,\mu^2) = ig(\mu^2) \tag{2.7}$$

defines the electric coupling. In fact, sending $\mu \to 0$, the so-called Thompson limit, defines the quantity commonly known as the electric charge even beyond the tree-level expression (2.2). The derivative for a coupling defined such is called the β function, defined as

$$\frac{dg}{d\ln\mu} = \beta(g) = -\beta_0 \frac{g^3}{16\pi^2} + \mathcal{O}(g^5), \tag{2.8}$$

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

$$\alpha(q^2) = \frac{g(q^2)^2}{4\pi} = \frac{\alpha(\mu^2)}{1 + \frac{\alpha(\mu^2)}{4\pi}\beta_0 \ln \frac{q^2}{\mu^2}} \equiv \frac{4\pi}{\beta_0 \ln \frac{q^2}{\Lambda^2}}, \tag{2.9}$$

which introduces the scale Λ of a theory as a boundary condition of the ordinary differential equation (2.8). The value of Λ can be determined, e. g., by evaluating perturbatively to this order the right-hand-side of (2.7). 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

$$\beta_0 = \frac{11}{3}C_A - \frac{2}{3}N_f - \frac{1}{6}N_H \tag{2.10}$$

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 (2.9) 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, the latter being referred to as asymptotic safety and to be discussed in more detail in section 7.2.

Similar equations like (2.10) 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.

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.

For the purpose of illustration, consider the vector and axial-vector non-singlet currents j^a_μ and J^a_μ, respectively, of a gauge theory, given by

$$\begin{aligned} j^a_\mu &= \bar{\psi}_i \gamma_\mu \tau^a_{ij} \psi_j \\ J^a_\mu &= \bar{\psi}_i \gamma_\mu \gamma_5 \tau^a_{ij} \psi_j . \end{aligned}$$

These currents are, of course, not conventionally conserved. This follows from the expression

$$\partial^\mu j^a_\mu = (\partial^\mu \bar{\psi})\gamma_\mu \tau^a \psi + \bar{\psi}\gamma_\mu \tau^a \partial^\mu \psi.$$

This can be recast using the Dirac equation to yield

$$\begin{aligned}\partial^\mu j^a_\mu &= -i\bar{\psi}(g\tau^b\gamma_\mu A^\mu_b - m)\tau^a\psi - i\bar{\psi}\tau^a(-g\tau^b\gamma_\mu A^\mu_b + m)\psi \\ &= ig\bar{\psi}\left[\tau^a,\tau^b\right]\gamma_\mu A^\mu_b\psi = -gf^{abc}A^\mu_b\bar{\psi}\gamma_\mu\tau_c\psi = -gf^{ab}{}_c A^\mu_b j^c_\mu.\end{aligned}$$

The result is already indicating that the current is only covariantly conserved, i. e.,

$$D^{ab}_\mu j^\mu_b = 0, \tag{2.11}$$

and similarly for J^a_μ follows

$$D^{ab}_\mu J^\mu_b = 2im\bar{\psi}\gamma_5\tau^a\psi = 2mip^a, \tag{2.12}$$

where p is the pseudo-scalar density. The latter equation implies already that only for massless fermions there will be no gauge anomaly. However, this is not a problem, as only zero-mass fermions are admitted to the standard model anyway, and all apparent fermion masses are generated by the Brout-Englert-Higgs effect. But for the standard model this is still modified. Due to the parity violation, it is necessary to consider a current for left-handed and right-handed fermions separately, giving four instead of two covariant conservation equations, where the corresponding left-handed and right-handed covariant derivatives for the left-handed and right-handed currents appear.

So far, this was the conservation at the classical level, which already requires the fermions to be massless. At the quantum level, this result is expressed by Ward-identities. In particular, take Ward identities for correlation functions of the form

$$T^{ijk}_{\mu\nu\rho} = \langle T j^i_\mu j^j_\nu j^k_\rho \rangle,$$

where i, j, and k can take the values V, A, and P, which require to replace the j by j^a, J^a, and p^a, respectively, and the Lorenz index is dropped in the last case. Calculating the corresponding Ward identities for a local chiral transformation

$$\begin{aligned}\psi' &= e^{i\beta(x)\gamma_5}\psi(x) \\ \bar{\psi}' &= \bar{\psi}e^{i\beta(x)\gamma_5}\end{aligned}$$

yields the expressions

$$\begin{aligned}\partial^\mu_x T^{VVA}_{\mu\nu\rho}(x,y,z) &= \partial^\nu_y T^{VVA}_{\mu\nu\rho}(x,y,z) = 0 \\ \partial^\rho_z T^{VVA}_{\mu\nu\rho}(x,y,z) &= 2mT^{VVP}_{\mu\nu}(x,y,z),\end{aligned} \tag{2.13}$$

directly implementing the relations (2.11) and (2.12). In these expressions it has been assumed that the theory is anomaly-free, i. e., if the Jacobian is not invariant, then additional terms from the measure appear.

To check this, rotate first to Euclidean space-time, by replacing $t \to it$ and correspondingly in all covariant quantities the time component by i-times the time component and in all contravariant quantities the time components by $-i$-times the time components. Then expand the fermion fields in orthonormal eigenfunctions ψ_n of the Dirac operator,

$$\begin{aligned}\psi(x) &= \sum_n a_n \psi_n \\ \bar{\psi}(x) &= \sum_n \psi_n^\dagger(x)\bar{b}_n,\end{aligned}$$

which satisfy

$$i\gamma_\mu D^\mu \psi_n = \lambda_n \psi_n \tag{2.14}$$

$$-i\gamma_\mu D^\mu \psi_n^\dagger = \lambda_n \psi_n^\dagger. \tag{2.15}$$

This permits to rewrite the path integral as an infinite product of integrations over the coefficients,

$$\mathcal{D}\psi\mathcal{D}\bar{\psi} = \Pi_m da_m d\bar{b}_m, \tag{2.16}$$

keeping in mind that these differentials are Grassmannian.

Now, a local chiral transformation $\beta(x)$

$$\psi \to e^{i\beta(x)\gamma_5}\psi,$$

then corresponds to a linear transformation of the coefficients

$$a_m \to C_{mn} a_n = a_n',$$

which yields the Jacobian

$$\Pi_m da_m' d\bar{b}_m' = \frac{1}{(\det C)^2}\Pi_m da_m d\bar{b}_m,$$

or, formally,

$$\mathcal{D}\psi'\mathcal{D}\bar{\psi}' = \frac{1}{(\det C)^2}\mathcal{D}\psi\mathcal{D}\bar{\psi}.$$

This determinant can be rewritten as

$$\frac{1}{(\det C)^2} = e^{-2\mathrm{tr}\ln C} = e^{-2\mathrm{tr}\delta C}, \tag{2.17}$$

where in the last equality it was assumed that β is infinitesimal, and thus $C = 1 + \delta C$ is close to one. In this case, δC can be evaluated starting from

$$a'_m \psi_m = (1 + i\beta\gamma_5) a_n \psi_n$$

which can be reduced using the orthonormality of the eigenstates of the Dirac equation to

$$a'_m = \int d^4x \psi_m^\dagger (1 + i\beta\gamma_5) \psi_n a_n = (1 + \delta c_{mn}) a_n. \tag{2.18}$$

Inserting this result into (2.17) yields for the Jacobian of the infinitesimal transformation

$$J = \exp\left(-2i \int d^4x \beta \psi_m^\dagger \gamma_5 \psi_m\right),$$

where the trace has been evaluated.

Unfortunately, the expression, as it stands, is ill-defined. It is necessary to regularize it. A useful possibility to make the expression well-defined, the so-called heat-kernel regularization, is by replacing the trace over the eigenstates as

$$\psi_m^\dagger \gamma_5 \psi_m \to \lim_{\tau \to 0} \frac{1}{(4\pi\tau)^2} \psi_m^\dagger \gamma_5 e^{-\lambda_m^2 \tau} \psi_m,$$

where the limit has to be performed at the end of the calculation only. Expanding the Gaussian and using the relations (2.14-2.15), this expression can be rewritten as

$$\lim_{\tau \to 0} \frac{1}{(4\pi\tau)^2} \psi_m^\dagger \gamma_5 e^{-\lambda_m^2 \tau} \psi_m = \lim_{\tau \to 0} \frac{1}{(4\pi\tau)^2} \mathrm{tr}\left(\gamma_5 e^{-\tau(\gamma_\mu D^\mu)^\dagger \gamma_\nu D^\nu}\right).$$

The exponential can be rewritten as

$$(\gamma_\mu D^\mu)^\dagger \gamma_\nu D^\nu = -D_\mu D^\mu + \frac{i}{4} [\gamma^\mu, \gamma^\nu] F_{\mu\nu}^a \tau_a.$$

For $\tau \to 0$, essentially only the third term of the expansion of the exponential, canceling the pre-factor τ^{-2}, will contribute, since $\mathrm{tr}\gamma_5 = 0$ and $\mathrm{tr}\gamma_5\gamma_\mu\gamma_\nu = 0$. Thus, the remainder is just

$$J = \exp\left(-\frac{i}{32\pi^2} \int d^4x \beta \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}^a F_{a\rho\sigma}\right).$$

Hence, the Jacobian is non-trivial, and needs to be taken into account. In particular, it will appear in Ward identities.

Entering this into the Ward identity (2.13) yields after some calculations an additional term

$$k^\rho T_{\mu\nu\rho}^{V^a V^b A^c}(p,q,k) = 2m T_{\mu\nu}^{V^a V^b P^c}(p,q,k) + \frac{\mathrm{tr}\left\{\tau_L^a, \tau_L^b\right\} \tau_L^c - \mathrm{tr}\left\{\tau_R^a, \tau_R^b\right\} \tau_R^c}{2} \frac{1}{3\pi^2} \epsilon_{\mu\nu\rho\sigma} p^\rho q^\sigma,$$

were L and R indicate the representation of the left-handed and right-handed fermions. As a consequence, the classical gauge symmetry is broken by the anomaly, and results will depend on the choice of gauge. This can be directly understood form this expression: the left-hand side should vanish as a vacuum expectation value, which should be zero for any colored object. On the right-hand side, the first term will indeed do so, if the fermion mass is zero, but this is not obvious for the second term.

There are now two possibilities how to obtain an anomaly-free theory. Either, the theory is anomaly-free, if each of the remaining 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 SU($N\geq 3$) and 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

$$\mathrm{tr}\tau^a\{\tau^b,\tau^c\} \tag{2.19}$$

$$\mathrm{tr}Q\{\tau^a,\tau^b\} \tag{2.20}$$

$$\mathrm{tr}\tau^a Q^2 \tag{2.21}$$

$$\mathrm{tr}Q^3. \tag{2.22}$$

The contribution (2.19) vanishes, as this is a pure SU(2) expression. The term (2.22) is not making a difference between left and right, and is therefore also vanishing. It turns out that (2.20) and (2.21) lead to the same result, so it is sufficient to investigate (2.21). 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

[6] Actually, unitarily equivalent is sufficient.

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 standard model is part of a larger theory, which imposes this structure.

There is actually a further possible anomaly, the so-called Witten anomaly, which comes from the parity violation of the standard model. It will not be detailed here, but 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. The basic ingredient will be the Einstein-Hilbert Lagrangian of general relativity, and the metric $g_{\mu\nu}$, 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. This is already the basic object also of general relativity. Both classically and quantum, it describes 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 $g^{\mu\nu}$,

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

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

$$\delta g^{\mu\nu} = -g^{\mu\lambda}g^{\nu\rho}\delta g_{\lambda\rho} \tag{2.23}$$

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

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

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

$$\delta\omega = \frac{1}{2}\omega g^{\mu\nu}\delta g_{\mu\nu} = -\frac{1}{2}\omega g_{\mu\nu}\delta g^{\mu\nu} \qquad (2.24)$$

as a consequence of (2.23).

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

$$\begin{aligned} T'_{\mu...\nu}(x') &= \frac{\partial x_\mu}{\partial x'_\alpha}...\frac{\partial x_\nu}{\partial x'_\beta}T_{\alpha...\beta}(x) \\ T'^{\mu...\nu}(x') &= \frac{\partial x'^\mu}{\partial x^\alpha}...\frac{\partial x'^\nu}{\partial x^\beta}T^{\alpha...\beta}(x) \end{aligned} \qquad (2.25)$$

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

$$\begin{aligned} D_\mu A_\nu &= \partial_\mu A_\nu - \Gamma^\lambda_{\mu\nu}A_\lambda \qquad (2.26) \\ \Gamma^\lambda_{\mu\nu} &= \frac{1}{2}g^{\lambda\sigma}(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}), \end{aligned}$$

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

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + g_{\mu\nu}\Lambda = -\kappa T_{\mu\nu}, \tag{2.27}$$

which can be derived as the Euler-Lagrange equation from the Lagrangian[7]

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

where the first two terms are the Einstein-Hilbert Lagrangian $\mathcal{L}_{EH}$, $\mathcal{L}_M$ is the matter Lagrangian yielding the covariantly conserved energy momentum tensor $T_{\mu\nu}$, defined as

$$T_{\mu\nu} = \left(-\eta_{\mu\nu}\mathcal{L} + 2\frac{\delta\mathcal{L}}{\delta g^{\mu\nu}}(g_{\mu\nu} = \eta_{\mu\nu}) \right), \tag{2.28}$$

$\kappa = 16\pi G_N$ is the Newton's constant, and Λ gives the cosmological constant (with arbitrary sign). The volume factor ω will be absorbed at times in the integral measure of the action.

For the purpose of quantization it will be useful to rewrite the first term of the Lagrangian, the Einstein-Hilbert contribution $\mathcal{L}_{EH}$, as

$$\begin{aligned}\mathcal{L}_{EH} &= \frac{1}{2\kappa}\omega g^{\mu\nu}(\Gamma^\Lambda_{\sigma\lambda}\Gamma^\sigma_{\mu\nu} - \Gamma^\lambda_{\mu\sigma}\Gamma^\sigma_{\nu\lambda}) + \partial_\mu V^\mu \\ \partial V^\mu &= \frac{1}{2\kappa} h(g^{\mu\lambda}g^{\sigma\tau} - g^{\mu\sigma}g^{\lambda\tau})\partial_\mu g_{\sigma\tau}.\end{aligned} \tag{2.29}$$

The second term is a total derivative, and therefore can quite often be dropped.

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

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

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 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 4, 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.

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, 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. 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 (2.29): 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 7.2.

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 which should be mentioned, 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. 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,

[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.

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 (2.27). 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 (2.27) together are independently in general relativity, and thus indeed constant. However, the known (quantum) effects contributing to such a constant provide a much too large value for Λ, about 10^{40} times too large. These include quantities like the chiral and gluon condensates. 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.

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 an 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

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

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.

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.

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[3].

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.

[3]The scale is inversely proportional to the couplings, as they have inverse mass dimension.

Chapter 4

Grand unified theories

The first example of a 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.

4.1 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 will be discussed below. 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. There is a theorem, the Coleman-Mandula theorem, that any group, which includes both the Poincare group and an internal gauge group like the SU(5), can only be a simple product group, and therefore not really unified. It requires, e. g., the addition of supersymmetry or another structural change to circumvent this theorem. This theorem, and how it can be circumvented, is discussed in much more detail in the lecture on supersymmetry, and will not be detailed here further.

4.2 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

$$\begin{aligned}
A_\mu = A_\mu^a \tau_a &= G_\mu^i \mathrm{diag}(\lambda^i, 0, 0) + W_\mu^i \mathrm{diag}(0,0,0,\sigma^i) \\
&- \frac{1}{\sqrt{15}} B_\mu \mathrm{diag}(-2,-2,-2,3,3) + \sqrt{2}(X_\mu^c x^c + Y_\mu^c y^c + X_\mu^{c\dagger} \xi^c + Y_\mu^{c\dagger} \chi^c) \\
&= \sqrt{2} \begin{pmatrix} & \frac{1}{\sqrt{2}} G_\mu^i \lambda^i & & X_\mu^{1\dagger} & Y_\mu^{1\dagger} \\ & & & X_\mu^{2\dagger} & Y_\mu^{2\dagger} \\ & & & X_\mu^{3\dagger} & Y_\mu^{3\dagger} \\ X_\mu^1 & X_\mu^2 & X_\mu^3 & \multicolumn{2}{c}{\frac{1}{\sqrt{2}} W_\mu^i \sigma_i} \\ Y_\mu^1 & Y_\mu^2 & Y_\mu^3 & & \end{pmatrix} - \frac{1}{\sqrt{15}} \begin{pmatrix} -2 & & & & \\ & -2 & & & \\ & & -2 & & \\ & & & 3 & \\ & & & & 3 \end{pmatrix} B_\mu
\end{aligned} \tag{4.1}$$

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

[1]The mixing of the weak fields and the photon has not been performed, therefore the weak isospin bosons are electrically neutral.

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 quark and 4 leptons, counting two quark and lepton flavors 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 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.

4.3 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

$$D_\mu = \partial_\mu - ig\tau_a A_\mu^a, \tag{4.2}$$

where g is the gauge coupling, A^a_μ the gauge boson field, and τ^a are the generators of the group, e. g. SU(5) yielding (4.1), 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 (4.2) has to be compared to the covariant derivatives of matter fields in the standard model, which is given by

$$D_\mu = \partial_\mu - ig_s\tau^s_a G^a_\mu - ig_i\tau^w_a W^a_\mu + ig_h\frac{y}{2}B_\mu.$$

The strong interaction is parametrized by g_s, the strong coupling constant, G^a_μ 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^a_μ 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 (4.1). 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

$$\frac{1}{4}\text{tr}(yy) = \frac{10}{3} \stackrel{!}{=} c^2\text{tr}(\tau^h\tau^h) = c^2 T_R$$

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

[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.

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

$$e = \frac{g_i g_h}{\sqrt{g_i^2 + g_h^2}} = \sqrt{\frac{5}{3}} g \tag{4.3}$$

and the Weinberg angle θ_W by

$$\sin^2 \theta_W = \frac{g_h}{\sqrt{g_h^2 + g_i^2}} = \frac{3}{8}. \tag{4.4}$$

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.

In one-loop approximation a generic coupling constant g changes with energy μ as

$$\left(\frac{g(\mu)^2}{4\pi}\right)^{-1} = \alpha(\mu)^{-1} = \alpha(\Lambda_{\text{GUT}})^{-1} - \frac{\beta_0}{2\pi} \ln \frac{\Lambda_{\text{GUT}}}{\mu}, \tag{4.5}$$

where Λ_{GUT} is the energy scale were the gauge symmetry of the GUT is manifest. The only missing ingredient are then the β functions, which are specific for the respective coupling, and can be calculated, e. g., in perturbation theory. For a generic manifest Yang-Mills theory, like QCD, coupled to fermions and scalars it is given by

$$\beta_0 = \frac{11}{3} C_A - \frac{2}{3} T_f - \frac{1}{3} T_s,$$

where C_A is the adjoint Casimir of the gauge group, e. g. 5 for SU(5), and T_f and T_s depend on the number and representation of fermions and scalars, respectively, with masses below μ. If below the GUT scale only the known particles (including a light Higgs) exist, then the relevant coefficients β_0 are given by

$$\begin{aligned} \beta_0^s &= \quad 11 - \tfrac{4}{3} N_g &&= 7 \\ \beta_0^i &= \tfrac{22}{3} - \tfrac{4}{3} N_g - \tfrac{1}{6} N_H &&= \frac{19}{6} \\ \beta_0^h &= 0 - \tfrac{20}{9} N_g - \tfrac{1}{6} N_H &&= -\frac{41}{6}, \end{aligned}$$

for the three interactions, where $N_g = 3$ is the number of fermion generations and $N_H = 1$ the number of complex Higgs doublets. The sign of this leading coefficient β_0 is very significant. Simply by putting these numbers into the equation (4.5) shows that for a positive β_0 the coupling decreases with increasing energy. Thus, such theories become weaker and weaker interacting with increasing energies, they are asymptotically free. In contrast the opposite occurs for theories with negative β_0. These become stronger interacting with increasing energy, and eventually perturbation theory will fail for them. However, for energies up to the GUT scale perturbation theory appears to be sufficient for the hypercharge coupling. As an aside, a unification of the couplings would be impossible if the electric interaction would be stronger than either or both of the strong and weak gauge coupling, the later only appearing small by the effective masses of the W and Z bosons.

Using (4.4), it follows that

$$\sin^2\theta_W = \frac{3}{8} - \frac{\alpha_i^{-1}+\alpha_h^{-1}}{2\pi}\frac{109}{24}\ln\frac{\Lambda_{\text{GUT}}}{\mu}.$$

To eliminate the unknown scale Λ_{GUT} another evolution equation can be used. Particularly convenient is the combination

$$\alpha_i^{-1}+\alpha_h^{-1}-\frac{8}{3}\alpha_s = \frac{67}{6\pi}\ln\frac{\Lambda_{\text{GUT}}}{\mu},$$

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 $\Lambda_{\text{GUT}} \approx 8\times 10^{14}$ GeV, $\alpha_s(\Lambda_{\text{GUT}}) \approx 1/42$, 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.

4.4 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

$$\mathcal{L} = \frac{4G_{\mathrm{GUT}}}{\sqrt{2}}(\bar{u}\gamma^{\mu}u\bar{e}\gamma_{\mu}d)$$

for quark fields u and d and the electron field e. This vertex permits that a d 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.

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 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 mesons, and could oscillate by mixing into mesonic states of the standard model, and would therefore decay definitely at least 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, 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 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 $\Lambda_{\text{ETC}} > \Lambda_T$, 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, $\text{SU}(\geq 6)_F\times\text{SU}(N_T)\times\text{SU}(3)\times\text{SU}(2)\times\text{U}(1)$. 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

$$g^2\bar{T}\gamma_\mu T D^{\mu\nu}_{\text{ETC}}(p^2)\bar{q}\gamma_\nu q \overset{p^2\ll\Lambda_{\text{ETC}}}{\to} \frac{g^2}{M^2_{\text{ETC}}}\bar{T}\gamma_\mu T\bar{q}\gamma^\mu q \overset{p^2\ll\Lambda_T}{\to} \frac{g^2\langle\bar{T}T\rangle}{M^2_{\text{ETC}}}\bar{q}q,$$

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

$$m_q \approx \beta\frac{N_T\Lambda^3_{\text{TC}}}{\Lambda^2_{\text{ETC}}} \tag{5.1}$$

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

$$\frac{\delta m_K^2}{m_K^2} \to \frac{\delta m_K^2}{m_K^2} + \gamma \frac{f_K^2 m_K^2}{\Lambda_{\text{ETC}}^2}.$$

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 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[3]

$$t = \int_g^{g(t)} \frac{dg'}{\beta(g')},$$

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

$$\beta_T = -\frac{g_T^3}{16\pi^2} \left(\frac{11}{3} N_T - \frac{8}{3} N_f \right) + \mathcal{O}(g^5)$$

[3]A solution of this to first order in g^2 was given by (4.5).

requiring about $N_f \approx 11N_T/8$. 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 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[4]. 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

[4]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.

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 have been obtained. E. g., for SU(3), the theory is QCD-like up to about 8-9 fundamental flavors, is walking up to 11-13 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 $\mathrm{SU}(3)_{1+2}\times\mathrm{SU}(3)_2$. 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 $\mathrm{SU}(3)_3$ 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

$$(\bar{t}_L\bar{\chi}_L)\begin{pmatrix}0 & g_t v\\ M_{t\chi} & M_{\chi\chi}\end{pmatrix}\begin{pmatrix}t_R\\ \chi_R\end{pmatrix}+\mathrm{h.c.},$$

where $M_{t\chi}$ and $M_{\chi\chi}$ are free parameters. The obtained masses for the mass eigenstates are thus

$$m_i^2=\frac{1}{2}\left(M_{t\chi}^2+M_{\chi\chi}^2+(g_t v)^2\pm\sqrt{\left(M_{t\chi}^2+M_{\chi\chi}^2+(g_t v)^2\right)^2-4g_t^2v^2M_{t\chi}^2}\right) \tag{5.6}$$

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

$$\begin{aligned} m_1^2 &= \frac{g_t^2v^2M_{t\chi}^2}{M_{t\chi}^2+M_{\chi\chi}^2+(g_t v)^2}+\mathcal{O}\left(\frac{g_t^4v^4M_{t\chi}^4}{M_{\chi\chi}^6}\right)\\ m_2^2 &= M_{t\chi}^2+M_{\chi\chi}^2+(g_t v)^2+\mathcal{O}\left(\frac{g_t^2v^2M_{t\chi}^2}{M_{\chi\chi}^4}\right)\end{aligned}$$

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 $SU(3)_{1+2} \times SU(3)_3$ to $SU(3)_c$ 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.

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 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.

[1] N has to be larger than two for compatibility with experiment.

6.4 Flavons

A serious obstacle in technicolor theories 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.

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.

[2] In fact, the SU(5) GUT of section 4.2 treats left-hand fermions and right-hand fermions both vectorial, but in different multiplets of the gauge symmetry.

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 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.

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 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.

6.8 Inflaton

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 and strong phase transition leads to an inflationary period, but in both cases far to short and too ineffective. Thus, having a third phase transition in an additional sector could solve the problem. For this an additional scalar field with suitable potential and symmetry can be introduced, the inflaton[3].

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.

[3]There are many similar scenarios, and the name quintessence is also attached to them.

Chapter 7

New paradigms

So far, most extensions have been discussed in terms of either perturbation theory or methods borrowed from QCD. However, in recent times it turns out that both paradigms are too narrow to capture all possible features of quantum field theories, and a number of new approaches have been established. These will be briefly introduced here, together with their implications. These are largely issues under development, and though having already interesting implications still need further development.

7.1 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 and lower edge of the conformal window, which relates again strong-coupling 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.

7.2 Asymptotic safety

There is one interesting feature of the β-function (2.8). 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. The results support the idea that this may indeed be possible for theories like canonical quantum gravity or the standard model. While not completely established yet, this is a promising development.

Especially for canonical quantum gravity, i. e. a canonical quantization of classical general relativity as laid out in section 2.2, this seems to be possible and actually robust. While this is not sufficient to establish this as the quantum gravity - for this it needs to measure quantum effects of gravity to see whether this is the correct theory at all - it diminishes currently the need for the more exotic concepts discussed in section 7.3, 7.4, and chapter 9.

In addition, it has been seen that adding ordinary matter to the setup, and thus having a unified quantum theory of all known interactions, seems not to destroy this feature. In fact, it rather seems possible that this may, by interaction with gravity, also tame ultraviolet problems of the standard model, e. g. the Landau poles in the Higgs sector and of QED.

In the same context also attempts exist to use a discretization of gravity in terms of elementary geometric objects, so-called (causal) dynamical triangulation, to quantize ordinary gravity. Such a discretization is akin to the lattice discretization of ordinary quantum field theories, and accessible to similar methods. While a general method, and thus able to cope with any behavior of gravity, the results also hint towards an asymptotic safety scenario.

An interesting result in this scenario is that the effective dimensionality of the theory is at short distances two-dimensional. This is consistent. In two dimensions gravity is not having any dynamical degrees of freedom[1], and thus the theory is trivially stable. Note, however that this is not related to the actual dimensionality, i. e. the number of components of a vector, but rather to how distances are measured.

7.3 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

[1]That is actually already the case in three dimensions, in contrast to Maxwell or Yang-Mills theory, which reach this status only in two dimensions.

the non-commutativity becomes manifest and the infrared, where the coordinates should again effectively commute. This problem is known as IR-UV mixing.

7.4 Loop quantum gravity

In contrast to asymptotic safety in section 7.2, 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.

7.5 The FMS mechanism

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.

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.

7.5.1 Scalar QCD

The first step is to address the situation without the BEH effect. Then the theory is essentially scalar QCD, i. e. QCD with the fermionic quarks replaced by scalars, and only two colors. Such a theory is naively expected to behave like QCD, and this is indeed the case. Thus, then confinement occurs, and the only degrees of freedom observable are bound states, i. e. the analogue of hadrons.

7.5.2 Elitzur's theorem

To approach the electroweak sector, the first step is to realize that the electroweak symmetry breaking is really just a hiding of the symmetry by a gauge choice. It can be shown that the actual gauge symmetry can never be broken. This is known as Elitzur's theorem.

The argument[2] proceeds as follows, and is best seen by first considering a simpler example. Take as a theory a theory of two-space-time dimension in cylindrical coordinates

[2] In the original argument of Elitzur is a loophole, as some types of non-analyticities are not considered. The following is a more modern view which comes to the same conclusion. It is also necessary to generally do not use in the derivations any sources which are not invariant under the symmetry transformation, as these would explicit break the symmetry, and therefore potentially make a smooth and analytic approach

r and θ with the (Euclidean) action $S = r^2$.Then the partition function is given by

$$Z = \int_{-\pi}^{\pi} d\cos\theta \int r dr e^{-r^2}$$

This partition function has a global symmetry for any rotation $\theta \to \theta + \delta$. This invariance manifests in the fact that the expectation value of every even function of θ vanishes.

To break this symmetry would require that there is a particular direction singled out, i. e. some angle θ_0 should be special, as this would break this symmetry. I. e., that for some vector-valued quantity, which has a direction,

$$\int_{-\pi}^{\pi} d\cos\theta \int r dr (a(\vec{r}^2)\vec{e}_r + b(\vec{r}^2)\vec{e}_\theta)e^{-r^2} \neq \vec{0}$$

where the coefficient functions must be only depending on $\vec{r}^2$, as otherwise the observable itself would break the symmetry explicitly. However, the whole integral has no possibility to single out such a preferred direction, as both the integral and the integral measure are invariant.

Thus, the only possibility would be to modify either the action or the integral measure. The former would be done by an explicit symmetry breaking term, the other e. g. by a gauge condition, which singles out a subrange of θ values.

The situation in gauge theories is similar. The vectors $\vec{e}_\theta$ and $\vec{e}_r$ correspond to the gauge fields, and the dependence only on $\vec{r}^2$ is equivalent to being a gauge-invariant object. Gauge orbits are then given by the variation in θ at fixed r, and gauge transformations move around this orbit. Gauge-fixing then is the same as restricting the integral on the angle θ, and therefore making non-vanishing integrals of functions of the unit vectors, and thus gauge fields, possible.

7.5.3 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.

of the source to zero impossible.

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[3] 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

$$\langle v_2 \rangle = \left\langle \left| \int d^d x \phi(x) \right|^2 \right\rangle \tag{7.1}$$

would have the desired property. Note that a quantity like

$$\langle v^2 \rangle = \left\langle \int d^d x \, |\phi(x)|^2 \right\rangle ,$$

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.

[3]Or actually by all orbits up to a measure zero set.

However, in a gauge theory this is not enough[4]. To show that this really distinguishes between the BEH case and any alternatives, the observable must also be gauge-invariant under local gauge transformations, and (7.1) 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[5]. But for not asymptotically free theories, especially those suffering from the triviality problem of section 3.1, no such statement exists[6].

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

[4] Note that for global symmetries similar considerations apply, and without explicit symmetry breaking a quantity like (7.1) 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 (7.1) should be rather seen as an indication for a metastability against external explicit symmetry breaking, rather than a real breaking of symmetries.

[5] Though in practice it is usually impossible to make reliable statements on how large this neighborhood is.

[6] 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.

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 Hooft gauges[7]. 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[8]. 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

[7]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.

[8]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.

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 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.

7.5.4 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[9], 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[10].

In the same way, it is possible to construct a non-perturbative partner state for the gauge bosons,

$$\begin{aligned} \mathcal{O}^a_{1^-\mu}(x) &= \mathrm{tr}\tau^a X^\dagger D_\mu X \\ X &= \begin{pmatrix} \phi_1 & -\phi_2^* \\ \phi_2 & \phi_1^* \end{pmatrix}, \end{aligned} \tag{7.2}$$

which is a custodial triplet, and a gives the corresponding index. Using that the vacuum expectation value is constant, this yields

$$\langle \mathcal{O}^{a\dagger}_{1^-\mu}(y)\mathcal{O}^a_{1^-\mu}(x)\rangle \sim v^4 g^2 \langle W^i_\mu(y) W_i^\mu(x)\rangle + \mathcal{O}\left(\frac{\eta}{v}\right)$$

[9]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.

[10]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.

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.

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[11].

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.

7.5.5 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.

[11] 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.

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.

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[12] 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[13] 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[14].

This completes the standard model.

7.5.6 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.

[12]This is somewhat symbolically, and requires a more precise formulation to avoid a path dependence.

[13]There is no full proof yet, but the evidence is overwhelmingly substantial.

[14]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.

Indeed, in some toy theories, like toy-GUTs with $\mathrm{SU}(N>2)$ 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 (7.2). Because there 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 4.2, 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, 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 (2.27) 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

$$S_{EH} = -\frac{1}{2G_{N4+n}^{1+n/2}} \int d^{4+n}z\sqrt{|g^{4+n}|}R^{4+n} \qquad (8.2)$$

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

$$M_P = M_s(2\pi R M_s)^{\frac{n}{2}} = M_s\sqrt{V_n M_s^n},$$

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

$$\mathcal{L}_{EH} = -\frac{1}{2}H_{MN}\partial^2 H^{MN} + \frac{1}{2}H_N^N\partial^2 H_M^M - H^{MN}\partial_M\partial_N H_L^L + H^{MN}\partial_M\partial_L H_N^L - \frac{1}{M_s^{1+\frac{n}{2}}}H^{MN}T_{MN}.$$

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

$$
g_{MN} = \eta_{MN} + 16\pi G^{4+n}_N H_{MN},
$$

[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

$$\begin{aligned}
& \sum_{\vec{n}} \int d^4 x \,(1 + 8\pi G_N h + 32\pi G_N \phi) \\
& \times \left(\mathcal{L} + (\eta_{\mu\nu} - 16\pi G_N h_{\mu\nu} - 16\pi G_N \eta_{\mu\nu}\phi)\frac{\delta\mathcal{L}}{\delta g_{\mu\nu}}(g_{\mu\nu} = \eta_{\mu\nu})\right) \\
\approx & \sum_{\vec{n}} \int d^4 x \left(\mathcal{L} + \frac{\delta\mathcal{L}}{\delta g_{\mu\nu}}\left(2\eta_{\mu\nu} - 16\pi G_N(h_{\mu\nu} + \phi\eta_{\mu\nu})\right) + \mathcal{L}\left(8\pi G_N h_{\mu\nu} + 32\pi G_N \phi\eta_{\mu\nu}\right)\right) \\
= & \sum_{\vec{n}} \int d^4 x \left(\mathcal{L} + \eta_{\mu\nu}\frac{\delta\mathcal{L}}{\delta g_{\mu\nu}} + 16\pi G_N \left((h_{\mu\nu} + \phi\eta_{\mu\nu})\left(\eta_{\mu\nu}\mathcal{L} - 2\frac{\delta\mathcal{L}}{\delta g_{\mu\nu}}\right)\right)\right),
\end{aligned}$$

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]

$$
\begin{aligned}
\omega_{\mu\nu} &= h_{\mu\nu} - \frac{in_i R}{2\pi\vec{n}^2}\left(\partial_\mu A_{\nu i} + \partial_\nu A_{\mu i}\right) - \left(P^T_{ij} + 3P^L_{ij}\right)\left(\frac{2}{3}\frac{\partial_\mu\partial_\nu}{m^2_{\vec{n}}} - \frac{\eta_{\mu\nu}}{3}\right)\phi_{ij} \\
B_{\mu i} &= P^T_{ij}\left(A_{\mu j} - \frac{in_k R}{\pi\vec{n}^2}\partial_\mu\phi_{jk}\right) \\
\Phi_{ij} &= \sqrt{2}\left(P^T_{ik}P^T_{jl} + \frac{1}{1-n}\left(1-\sqrt{\frac{2+n}{3}}\right)P^T_{ij}P^T_{kl}\right)\phi_{kl} \qquad (8.9) \\
P^T_{ij} &= \delta_{ij} - \frac{n_i n_j}{\vec{n}^2} \\
P^L_{ij} &= \frac{n_i}{n_j}\vec{n}^2
\end{aligned}
$$

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

$$\zeta_\alpha(x_\mu, x_i) = \sum_{\vec{n}} \zeta^{\vec{n}}_\alpha(x_\mu)\exp\left(\frac{2\pi i n_i x^i}{R}\right).$$

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

$$\phi^{\vec{n}}_{ij} \to \phi^{\vec{n}}_{ij} + n_i\zeta^{\vec{n}}_j + n_j\zeta^{\vec{n}}_i.$$

As a consequence, zero modes do not change since for these $n_i = 0$. For non-zero $\vec{n}$, the relation

$$n_i P^T_{ij} = n_i\left(\delta_{ij} - \frac{n_i n_j}{\vec{n}^2}\right) = n_j - n_j\frac{\vec{n}^2}{\vec{n}^2} = 0$$

[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

$$\mathcal{M} = -\frac{n-1}{n+2}\frac{4\pi C_4}{3} m_{f_1} m_{f_2} \bar{f}_2 f_2 \bar{f}_1 f_1. \qquad (8.12)$$

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

$$S = \int d^4x dy \bar{\psi}(i\Gamma^\mu \partial_\mu + i\Gamma^5 \partial_y + \phi(y))\psi$$

where the Γ_μ denote the 4×4 five-dimensional version of the Dirac γ matrices[11],

$$\begin{aligned}
\Gamma_{0\ldots 3} &= \gamma_{0\ldots 3} = \begin{pmatrix} 0 & \sigma_{0\ldots 3} \\ \bar{\sigma}_{0\ldots 3} & 0 \end{pmatrix} \\
\Gamma_5 &= \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} \\
\sigma &= (1, \vec{\sigma}) \\
\bar{\sigma} &= (1, -\vec{\sigma}).
\end{aligned}$$

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

$$\begin{aligned} &\int d^4x dy (h(x)\psi_L^T(x,y) C_5 \psi_R(x,y) + \text{h.c.}) \\ \stackrel{j=0}{=} &\int d^4x h(x)\psi_L(0,x)\psi_R(0,x) \int dy f_L(0,y) f_R(0,y) = e^{-\frac{\mu^2 r^2}{2}} \int d^4x h(x)\psi_L(0,x)\psi_R(0,x). \end{aligned}$$

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

$$m^n_{KK}=x^G_n ke^{-\pi kR},$$

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 $\partial_5 A_\mu=0$ 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

$$m_{KK}^n \approx \pi\left(n+\frac{1}{2}\left(\left|c_{L,R}\mp\frac{1}{2}\right|-1\right)-\frac{(-1)^n}{4}\right)ke^{-\pi kR} \quad (8.14)$$

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 4.2. 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

String theory

9.1 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

$$\frac{A_{2g}}{A_{\text{free}}} \sim (\hbar G_N)^2 \sum_{\text{Intermediate states}} \int_0^E dE' E'^3 \sim \frac{1}{M_P^4} \int dE' E'^3 \to \infty \text{ for } E \to \infty \tag{9.1}$$

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 (9.1) 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.

9.2 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.

9.2.1 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

$$S_{pp}=-m\int d\tau\sqrt{-\partial_\tau X^\mu\partial^\tau X_\mu}. \tag{9.2}$$

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

$$u^\mu=\frac{\dot{X}^\mu}{\sqrt{-\dot{X}_\mu\dot{X}^\mu}} \tag{9.3}$$

yields the equation of motion after imposing the vanishing of the action under the variation and a partial integration as

$$m\dot{u}^\mu = 0 \tag{9.4}$$

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 (9.2) 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 (9.2) 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

$$\eta(\tau) := (-\gamma_{\tau\tau}(\tau))^{\frac{1}{2}},$$

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

$$\begin{aligned} X(\tau) &= X(\tau'(\tau)) \\ \eta'(\tau') &= \eta(\tau)\frac{d\tau}{d\tau'} = \frac{1}{\dot{\tau}'}\eta(\tau) \end{aligned} \tag{9.5}$$

This makes the expression invariant under diffeomorphisms: The transformation of η (9.5) 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

$$\begin{aligned} 0 = \frac{d}{d\tau}\frac{\partial L}{\partial \dot{\eta}} - \frac{\partial L}{\partial \eta} &= \frac{\dot{X}^\mu \dot{X}_\mu}{\eta^2} + m^2 \\ &\Longrightarrow \eta^2 = -\frac{\dot{X}^\mu \dot{X}_\mu}{m^2}. \end{aligned}$$

Thus knowledge of X determines η completely, since no derivatives of η appear. Inserting this expression into 9.5 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}.

9.2.2 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

$$X^\mu(\sigma,\tau) = X^\mu\left(\sigma'(\sigma,\tau), \tau'(\sigma,\tau)\right) \tag{9.6}$$

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 (9.2) 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

$$S_{NG} = \int_M d\tau d\sigma \mathcal{L}_{\rm NG}$$

in which M is the world-sheet of the string and $\mathcal{L}_{\rm NG}$ is the Nambu-Goto Lagrangian

$$\mathcal{L}_{\rm NG} = -\frac{1}{2\pi\alpha'}\sqrt{-\det h_{ab}} = -\frac{1}{2\pi\alpha'}\sqrt{\partial_\tau X_\mu \partial_\sigma X^\mu \partial_\sigma X_\rho \partial_\tau X^\rho - \partial_\tau X_\mu \partial_\tau X^\mu \partial_\sigma X^\rho \partial_\sigma X_\rho},$$

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

$$\begin{aligned} X'^{\mu} &= \Lambda^{\mu}_{\ \nu} X^{\nu} + a^{\mu} \\ \partial_a \Lambda^{\mu}_{\ \nu} X^{\nu} \partial_b \Lambda_{\mu}^{\ \gamma} X_{\gamma} &= \overbrace{\Lambda^{\mu}_{\ \nu} \Lambda_{\mu}^{\ \gamma}}^{\delta^{\gamma}_{\nu}} \partial_a X^{\nu} \partial_b X_{\gamma} = \partial_a X^{\mu} \partial_b X_{\mu}. \end{aligned}$$

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,

$$S_P = -\frac{1}{4\pi\alpha'} \int_M d\tau d\sigma \, (-\gamma)^{\frac{1}{2}} \gamma^{ab} h_{ab} \tag{9.7}$$

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

$$h_{ab} = \frac{1}{2}\gamma_{ab}\gamma_{cd}h^{cd} \tag{9.8}$$

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

$$S_P = -\frac{1}{4\pi\alpha'}\int d\tau d\sigma \gamma^{ab}\gamma_{ab}\left(-h\right)^{\frac{1}{2}} = -\frac{1}{2\pi\alpha'}\int d\tau d\sigma \left(-h\right)^{\frac{1}{2}} = S_{NG}$$

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

$$\begin{aligned} X'^{\mu}\left(\tau,\sigma\right) &= X^{\mu}\left(\tau,\sigma\right) \\ h'_{ab} &= h_{ab} \\ \gamma'_{ab} &= e^{2\omega(\tau,\sigma)}\gamma_{ab}. \end{aligned}$$

for arbitrary functions $\omega\left(\tau,\sigma\right)$. 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,

$$\frac{\gamma'_{ab}}{\left(-\gamma'\right)^{\frac{1}{2}}} = \frac{\gamma_{ab}e^{2\omega}}{\left(-\gamma'\right)^{\frac{1}{2}}} = \frac{\gamma_{ab}e^{2\omega}}{\left(-\gamma e^{4\omega}\right)^{\frac{1}{2}}} = \frac{\gamma_{ab}}{\left(-\gamma\right)^{\frac{1}{2}}}.$$

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

$$\left(-\gamma'\right)^{\frac{1}{2}}\gamma'^{ab} = \left(-\gamma e^{4\omega}\right)^{\frac{1}{2}}\gamma^{ab}e^{-2\omega} = \left(-\gamma\right)^{\frac{1}{2}}\gamma^{ab}.$$

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.

9.3 Quantized theory

9.3.1 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

$$\begin{aligned} -a^+b^- - a^-b^+ + a^ib^i &= -\frac{1}{2}\left(a^0+a^1\right)\left(b^0-b^1\right) - \frac{1}{2}\left(a^0-a^1\right)\left(b^0+b^1\right) + a^ib^i \\ &= -a^0b^0 + a^1b^1 + a^ib^i = a^\mu b_\mu. \end{aligned}$$

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^-.

9.3.2 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 9.2.1, the gauge condition to fix the diffeomorphism invariance becomes

$$X^+(\tau) = \tau.$$

The action is given by equation (9.5), 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

$$H = P^- = -P_+. \tag{9.9}$$

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 $|k_-, k_i\rangle$.

9.3.3 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

$$\begin{aligned} -\infty &\leq \tau \leq +\infty \\ 0 &\leq \sigma \leq L. \end{aligned} \tag{9.10}$$

Thus, L is the length of the string. Again, it is chosen that

$$\tau = X^+. \tag{9.11}$$

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

$$\partial_\sigma \gamma_{\sigma\sigma} = 0 \tag{9.12}$$

$$\det \gamma_{ab} = -1 \tag{9.13}$$

The conditions (9.11-9.13) 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 (9.13), and fixing the Weyl invariance. Since f is Weyl-invariant by construction, this implies that $\partial_\sigma\gamma_{\sigma\sigma}$ trivially vanishes, yielding (9.12). 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

$$\begin{aligned} L_P &= -\frac{1}{4\pi\alpha'}\int_0^L d\sigma\Big(\frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma X^i\partial_\sigma X^i - 2\gamma_{\tau\sigma}\left(-\partial_\sigma X^- + \partial_\tau X^i\partial_\sigma X^i\right) \\ &\quad -\gamma_{\sigma\sigma}\left(-2\partial_\tau X^- + \partial_\tau X^i\partial_\tau X^i\right)\Big). \end{aligned}$$

Employing now the relations for the average and variation this yields

$$\begin{aligned} L_P &= -\frac{1}{4\pi\alpha'}\Big(\gamma_{\sigma\sigma}2L\partial_\tau Z^- + \int_0^L d\sigma\Big(\gamma_{\sigma\sigma}\left(-\partial_\tau X^i\partial_\tau X^i\right) \\ &\quad +2\gamma_{\tau\sigma}\left(\partial_\sigma Y^- - \partial_\tau X^i\partial_\sigma X^i\right) + \frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma X^i\partial_\sigma X^i\Big)\Big) \end{aligned}$$

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

$$\partial_\sigma\gamma_{\tau\sigma} = 0, \tag{9.14}$$

and thus does not depend on σ.

Returning to the boundary condition of this open[1] string yields

$$\left(\partial_\sigma X^\mu\right)(\tau,0) = \left(\partial_\sigma X^\mu\right)(\tau,L) = 0, \tag{9.15}$$

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 (9.15).

On the other hand, when varying the original action with respect to the fields, this yields

$$\begin{aligned} \delta S_P &= S_P + \frac{1}{4\pi\alpha'}\int d\tau d\sigma\gamma_{ab}\partial_a\left(X^\mu+\delta X^\mu\right)\partial_b\left(X_\mu+\delta X_\mu\right) \\ &= \frac{1}{4\pi\alpha'}\int d\tau d\sigma\gamma_{ab}\left(\partial_a X^\mu\partial_b\delta X_\mu + \partial_\alpha\delta X^\mu\partial_b X_\mu\right). \end{aligned}$$

[1]No cyclicity of any function on σ has been imposed, which would be one possibility to implement a closed string.

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

$$\begin{aligned} \delta S_P &= \frac{1}{4\pi\alpha'} \int d\tau d\sigma \gamma_{ab} \left(\partial_a \partial_b X^\mu \delta X_\mu + \partial_b \partial_a X_\mu \delta X^\mu\right) \\ &\quad - \frac{1}{4\pi\alpha'} \int d\tau \gamma_{ab} \left(\partial_a X^\mu \delta X_\mu + \partial_b X_\mu \delta X^\mu\right)|_0^L \end{aligned} \tag{9.16}$$

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. However, the last expression must vanish under variation, implying once more the von Neumann condition (9.15)

$$\gamma_{ab} \left(\partial_a X^\mu + \partial_b X^\mu\right) = 0.$$

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

$$\begin{aligned} \delta S_P &= S_P + \frac{1}{4\pi\alpha'} \int d\tau d\sigma \Big(\gamma_{\sigma\sigma} \left(2\partial_\tau \left(X^- + \delta X^-\right) - \partial_\tau \left(X^i + \delta X^i\right) \partial_\tau \left(X^i + \delta X^i\right)\right) \\ &\quad + 2\gamma_{\tau\sigma} \left(\partial_\sigma \left(X^- + \delta X^-\right) - \partial_\tau \left(X^i + \delta X^i\right) \partial_\sigma \left(X^i + \delta X^i\right)\right) \\ &\quad + \frac{1 - \gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}} \partial_\sigma \left(X^i + \delta X^i\right) \partial_\sigma \left(X^i + \delta X^i\right) \Big). \end{aligned}$$

Expanding the result and dropping $\mathcal{O}\left(\delta^2\right)$ terms and annihilating a term of type S_P just leaves

$$\begin{aligned} \delta S_P &= \frac{1}{4\pi\alpha'} \int d\tau d\sigma \Big(\gamma_{\sigma\sigma} \left(2\partial_\tau \delta X^- - 2\partial_\tau X^i \partial_\tau \delta X^i\right) \\ &\quad + 2\gamma_{\tau\sigma} \left(\partial_\sigma \delta X^- - \partial_\tau X^i \partial_\sigma \delta X^i - \partial_\tau \delta X^i \partial_\sigma X^i\right) + 2\frac{1 - \gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}} \partial_\sigma X^i \partial_\sigma \delta X^i \Big) \end{aligned}$$

which can be rewritten as

$$\begin{aligned} \delta S_P &= \frac{1}{4\pi\alpha'} \int d\tau d\sigma \Big(\left(2\gamma_{\tau\sigma} \partial_\sigma \delta X^-\right) + \Big(-2\gamma_{\tau\sigma} \partial_\tau X^i \partial_\sigma \delta X^i + 2\frac{1 - \gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}} \partial_\sigma \delta X^i \partial_\sigma X \Big) \\ &\quad + \left(\gamma_{\sigma\sigma} 2\partial_\tau \delta X^- - 2\gamma_{\sigma\sigma} \partial_\tau X^i \partial_\tau \delta X^i\right) - \left(\gamma_{\tau\sigma} \partial_\tau \delta X^i \partial_\sigma X^i\right) \Big). \end{aligned}$$

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

$$-\partial_\sigma\left(2\gamma_{\tau\sigma}\partial_\tau X^i\right)+\partial_\sigma\left(\frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma X^i\right)=-2\gamma_{\tau\sigma}\partial_\sigma\partial_\tau X^i+\frac{1-\gamma_{\tau\sigma}^2}{\gamma_{\sigma\sigma}}\partial_\sigma^2X^i.$$

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 (9.15). 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 $\gamma_{\tau\sigma}$ (9.14), 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

$$\begin{aligned}P_- &= -P^+=\frac{\partial L_P}{\partial\left(\partial_\tau Z^-\right)}=-\frac{L}{2\pi\alpha'}\gamma_{\sigma\sigma} \\ \Pi^i &= \frac{\delta L_P}{\delta\left(\partial_\tau X^i\right)}=\frac{1}{2\pi\alpha'}\gamma_{\sigma\sigma}\partial_\tau X^i=\frac{P^+}{L}\partial_\tau X^i.\end{aligned} \tag{9.17}$$

From this the Hamiltonian is immediately constructed to be

$$\begin{aligned}H &= P_-\partial_\tau Z^-+\int_0^L d\sigma\Pi^i\partial_\tau X^i-L \\ &= \frac{L}{4\pi\alpha' P^+}\int_0^L d\sigma\left(2\pi\alpha'\Pi^i\Pi^i+\frac{1}{2\pi\alpha'}\partial_\sigma X^i\partial_\sigma X^i\right)\end{aligned} \tag{9.18}$$

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

$$\begin{aligned} \partial_\tau Z^- &= \frac{\partial H}{\partial P^-} = \frac{H}{P^+} \\ \partial_\tau P^+ &= -\frac{\partial H}{\partial Z^-} = 0 \\ \partial_\tau X^i &= \frac{\delta H}{\delta \Pi^i} = 2\pi\alpha' c \Pi^i \qquad (9.19) \\ \partial_\tau \Pi^i &= -\frac{\delta H}{\delta X^i} = \frac{c}{2\pi\alpha'}\partial_\sigma^2 X^i, \qquad (9.20) \end{aligned}$$

where a partial integration has been performed in (9.20) and c is defined as

$$c := \frac{L}{2\pi\alpha' P^+}.$$

Inserting (9.19) in (9.20) yields the wave equation for X^i

$$\partial_\tau^2 X^i = c^2 \partial_\sigma^2 X^i,$$

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

$$\begin{aligned} \hat{X}^i(\tau,\sigma) &= \hat{Z}^i + \frac{\hat{P}^i}{P^+}\tau + i(2\alpha')^{\frac{1}{2}} \sum_{n=-\infty, n\neq 0}^{n=\infty} \frac{\alpha_n^i}{n} e^{-\frac{\pi i n c \tau}{L}} \cos\frac{\pi n \sigma}{L} \qquad (9.21) \\ \alpha_{-n}^i &= \alpha_n^{i+}. \qquad (9.22) \end{aligned}$$

The relation (9.22) 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 (9.21) 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

$$\begin{aligned}\left[\hat{X}^{i},\hat{P}^{j}\right] &= i\delta^{ij}\\ \left[\alpha_{m}^{i},\alpha_{n}^{j}\right] &= m\delta^{ij}\delta_{m,-n}\end{aligned} \tag{9.23}$$

Here, a non-standard, though useful, normalization of (9.23) 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 $m>0$

$$\begin{aligned}\alpha_{m}^{i} &= \hbar\sqrt{m}a\\ \alpha_{-m}^{i} &= \hbar\sqrt{m}a^{\dagger}\\ -1 &= \left[a^{\dagger},a\right]\end{aligned} \tag{9.24}$$

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

$$\begin{aligned}P^{+}|0,k\rangle &= k^{+}|0,k\rangle\\ P^{i}|0,k\rangle &= k^{i}|0,k\rangle\\ \alpha_{m}^{i}|0,k\rangle &= 0 \text{ for } m>0\end{aligned} \tag{9.25}$$

Therefore k is the center-of-mass momentum. Higher excited states are then denoted by $|N,k\rangle$ and can be constructed as

$$|N,k\rangle=\left(\prod_{i=2}^{D-1}\prod_{n=1}^{\infty}\frac{\left(\alpha_{-n}^{i}\right)^{N_{in}}}{\sqrt{(n^{N_{in}}N_{in})!}}\right)|0,k\rangle\,,$$

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 (9.21) 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, (9.24).

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

$$\begin{aligned}
\Pi^i &= \frac{P^+}{L}\left(\partial_\tau X^i\right) \\
&= \frac{P^+}{L}\left(\frac{\hat{P}^i}{P^+} + \frac{\pi c}{L}\left(2\alpha'\right)^{\frac{1}{2}} \sum_{n=-\infty,n\neq 0}^{n=+\infty} \alpha_n^i e^{-\frac{\pi i n c\tau}{L}} \cos\frac{\pi n\sigma}{L}\right).
\end{aligned}$$

In addition, also $\partial_\sigma X^i$ is required, and is given by

$$\partial_\sigma X^i = -\frac{i\pi}{L}\left(2\alpha'\right)^{\frac{1}{2}} \sum_{n=-\infty,n\neq 0}^{n=+\infty} \alpha_n^i e^{-\frac{\pi i n c\tau}{L}} \sin\frac{\pi n\sigma}{L}.$$

Putting everything together yields the Hamiltonian

$$\begin{aligned}
& \frac{L}{4\pi\alpha' P^+}\int_0^L d\sigma\left(2\pi\alpha\Pi^i\Pi^i + \frac{1}{2\pi\alpha'}\partial_\sigma X^i\partial_\sigma X^i\right) \\
=\; & \frac{L}{4\pi\alpha' P^+}\Big(2\pi\alpha' P^i P^i + \int_0^L d\sigma \\
& \Big(\frac{\pi}{4\alpha' L P^+}\sum_{n=-\infty n\neq 0}^{n=+\infty} \alpha_n^i e^{-\frac{\pi i n c\tau}{L}} \cos\frac{\pi n\sigma}{L} \sum_{m=-\infty,m\neq 0}^{m=+\infty} \alpha_m^i e^{-\frac{\pi m n c\tau}{L}} \cos\frac{\pi m\sigma}{L} \\
& -\frac{\pi}{4\alpha' L P^+}\sum_{n=-\infty n\neq 0}^{n=+\infty} \alpha_n^i e^{-\frac{\pi i n c\tau}{L}} \sin\frac{\pi n\sigma}{L} \sum_{m=-\infty,m\neq 0}^{m=+\infty} \alpha_m^i e^{-\frac{\pi i m c\tau}{L}} \sin\frac{\pi m\sigma}{L}\Big)\Big),
\end{aligned}$$

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[2]

$$A = \frac{D-2}{2} \sum_{n=1}^{\infty} n,$$

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

$$e^{-\frac{\varepsilon|k_\sigma|}{\sqrt{\gamma_{\sigma\sigma}}}}$$
$$k_\sigma = \frac{n\pi}{L},$$

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

[2]With standard normalization, n would be replaced by 1, changing nothing qualitatively.

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

$$\begin{aligned} A &= \frac{D-2}{2}\sum_{n=1}^{\infty} n e^{-\frac{\varepsilon|k_\sigma|}{\sqrt{\gamma_{\sigma\sigma}}}} \\ &= \frac{D-2}{2}\left(\frac{2LP^+\alpha'}{\varepsilon^2\pi}-\frac{1}{12}+O(\varepsilon)\right), \end{aligned}$$

where the second line of (9.17) 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

$$A = \frac{2-D}{24}, \tag{9.26}$$

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 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

$$m^2 = 2P^+H - P^iP^i, \tag{9.27}$$

as a result of the light-cone equation

$$m^2 = P^+P^- + P^-P^+ - P^iP^i.$$

Inserting the result (9.26) into (9.27) 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 $\alpha^i_{-1}|0,k\rangle$ with $N=1$. Its mass reads

$$m^2=\frac{26-D}{24\alpha'}. \tag{9.28}$$

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 (9.28) would not yield zero. This implies also $A=-1$, due to (9.26).

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 (9.28).

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.

9.3.4 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

$$\sigma' = \sigma + s(\tau) \quad \mathrm{mod}\ L. \tag{9.29}$$

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

$$m^2 = 2P^+H - P^iP^i = \frac{2}{\alpha'}(N + R + A + B),$$

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 (9.29). 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

$$
\begin{aligned}
T^{ab} &= -4\pi\left(-\gamma\right)^{-\frac{1}{2}}\frac{\delta L}{\delta\gamma_{ab}} \qquad (9.30)\\
&= -\frac{4\pi}{\left(-\gamma\right)^{\frac{1}{2}}}\frac{\delta}{\delta\gamma_{ab}}\left(-\frac{1}{4\pi\alpha'}\left(-\gamma\right)^{\frac{1}{2}}\gamma^{cd}\partial_c X^\mu\partial_d X_\mu\right)\\
&= \frac{1}{\alpha'\left(-\gamma\right)^{\frac{1}{2}}}\left(\frac{\delta\left(-\gamma\right)^{\frac{1}{2}}}{\delta\gamma_{ab}}\gamma^{cd}\partial_c X^\mu\partial_d X_\mu+\left(-\gamma\right)^{\frac{1}{2}}\frac{\delta\gamma^{cd}}{\delta\gamma_{ab}}\partial_c X^\mu\partial_d X_\mu\right)\\
&= \frac{1}{\alpha'\left(-\gamma\right)^{\frac{1}{2}}}\left(-\frac{1}{2\left(-\gamma\right)^{\frac{1}{2}}}\frac{\delta\gamma}{\delta\gamma_{ab}}\gamma^{cd}\partial_c X^\mu\partial_d X_\mu+\left(-\gamma\right)^{\frac{1}{2}}\partial_a X^\mu\partial_b X_\mu\right)\\
&= \frac{1}{\alpha'\left(-\gamma\right)^{\frac{1}{2}}}\left(-\frac{1}{2\left(-\gamma\right)^{\frac{1}{2}}}\gamma\gamma^{ab}\gamma^{cd}\partial_c X^\mu\partial_d X_\mu+\left(-\gamma\right)^{\frac{1}{2}}\partial_a X^\mu\partial_b X_\mu\right)\\
&= \frac{1}{\alpha'}\left(\partial_a X^\mu\partial_b X_\mu-\frac{1}{2}\gamma^{ab}\gamma^{cd}\partial_c X^\mu\partial_d X_\mu\right)\\
&= \frac{1}{\alpha'}\left(\partial^a X^\mu\partial^b X_\mu-\frac{1}{2}\gamma^{ab}\partial_c X^\mu\partial^c X_\mu\right).
\end{aligned}
$$

To argue that this indeed is an energy-momentum tensor[3], 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.

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 (9.8), 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.

[3] In quantum field theory this is already a non-obvious fact, lest in string theory.

The next condition is the one of tracelessness. Calculating the trace T^a_a explicitly yields

$$\begin{aligned}\gamma_{ab}\frac{\delta L}{\delta\gamma_{ab}} &= \frac{1}{(-\gamma)^{\frac{1}{2}}}\left(\gamma_{ab}\partial^a X^\mu\partial^b X_\mu - \partial^c X^\mu\partial_c X_\mu\right)\\ &= \frac{1}{(-\gamma)^{\frac{1}{2}}}\left(\partial^a X^\mu\partial_a X_\mu - \partial^c X^\mu\partial_c X_\mu\right) = 0 \qquad (9.31)\\ &= \gamma_{ab}\frac{T^{ab}}{(-\gamma)^{\frac{1}{2}}},\end{aligned}$$

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 (9.31) and the fact that the Lagrange function is not depending on the τ-derivatives of γ_{ab}.

Using (9.8), 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 (9.18) 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 (9.17) 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

$$\begin{aligned} P &= -\int_0^L d\sigma \Pi^i \partial_\sigma X^i \\ &= \frac{2\pi}{L}\left(\sum_{n=1}^{\infty}\left(\alpha^i_{-n}\alpha^i_n - \beta^i_{-n}\beta^i_n\right) + A - B\right) \\ &= \frac{2\pi}{L}\left(N - R\right). \end{aligned}$$

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

$$N = R. \tag{9.32}$$

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

$$|1,1,k\rangle = \alpha^i_{-1}\beta^j_{-1} |0,0,k\rangle ,$$

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

$$S^{ij} = -i\sum_{n=1}^{\infty}\frac{1}{n}\left(\alpha^i_{-n}\alpha^j_n - \alpha^j_{-n}\alpha^i_n\right) \tag{9.33}$$

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 (9.33) 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 (9.23) are necessary. In the following the summation is kept implicit. This yields for the first term

$$\begin{aligned}\left[\alpha^i_{-n}\alpha^j_n,\alpha^k_{-m}\alpha^l_m\right] &= \alpha^i_{-n}\left[\alpha^i_n,\alpha^k_{-m}\alpha^l_m\right]+\left[\alpha^i_{-n},\alpha^k_{-m}\alpha^l_m\right]\alpha^j_n\\ &= \alpha^i_{-n}\alpha^k_{-m}\left[\alpha^j_n,\alpha^l_m\right]+\alpha^i_{-n}\left[\alpha^j_n,\alpha^k_{-m}\right]\alpha^l_m\\ &\quad+\alpha^k_{-m}\left[\alpha^i_{-n},\alpha^l_m\right]\alpha^i_n+\left[\alpha^i_{-n},\alpha^k_{-m}\right]\alpha^l_m\alpha^j_n\\ &= \alpha^i_{-n}\alpha^k_{-m}n\delta^{jl}\delta_{n,-m}+\alpha^i_{-n}\alpha^l_m n\delta^{jk}\delta_{n,m}\\ &\quad-\alpha^k_{-m}\alpha^j_n n\delta^{il}\delta_{-n,-m}-\alpha^l_m\alpha^j_n n\delta^{ik}\delta_{-n,m}\\ &= n(\alpha^i_{-n}\alpha^k_n\delta^{jl}+\alpha^i_{-n}\alpha^l_n\delta^{jk}-\alpha^k_{-n}\alpha^j_n\delta^{il}-\alpha^l_{-n}\alpha^j_n\delta^{ik}),\end{aligned}\tag{9.34}$$

for the second term

$$\begin{aligned}\left[\alpha^i_{-n}\alpha^j_n,\alpha^l_{-m}\alpha^k_m\right] &= \alpha^i_{-n}\left[\alpha^j_n,\alpha^l_{-m}\alpha^k_m\right]+\left[\alpha^i_{-n},\alpha^l_{-m}\alpha^k_m\right]\alpha^j_n\\ &= \alpha^i_{-n}\alpha^l_{-m}\left[\alpha^j_n,\alpha^k_m\right]+\alpha^i_{-n}\left[\alpha^j_n,\alpha^l_{-m}\right]\alpha^k_m\\ &\quad+\alpha^l_{-m}\left[\alpha^i_{-n},\alpha^k_m\right]\alpha^j_n+\left[\alpha^i_{-n},\alpha^l_{-m}\right]\alpha^k_m\alpha^j_n\\ &= \alpha^i_{-n}\alpha^l_{-m}n\delta^{jk}\delta_{n,-m}+\alpha^i_{-n}\alpha^k_m n\delta^{jl}\delta_{n,m}\\ &\quad-\alpha^l_{-m}\alpha^j_n n\delta^{ik}\delta_{-n,-m}-\alpha^k_m\alpha^j_n n\delta^{il}\delta_{-n,m}\\ &= n(\alpha^i_{-n}\alpha^l_n\delta^{jk}+\alpha^i_{-n}\alpha^k_n\delta^{jl}-\alpha^l_{-n}\alpha^j_n\delta^{ik}-\alpha^k_{-n}\alpha^j_n\delta^{il}),\end{aligned}\tag{9.35}$$

for the third term

$$\begin{aligned}
\left[\alpha^j_{-n}\alpha^i_n,\alpha^k_{-m}\alpha^l_m\right] &= \alpha^j_{-n}\left[\alpha^i_n,\alpha^k_{-m}\alpha^l_m\right]+\left[\alpha^j_{-n},\alpha^k_{-m}\alpha^l_m\right]\alpha^i_n\\
&= \alpha^j_{-n}\alpha^k_{-m}\left[\alpha^i_n,\alpha^l_m\right]+\alpha^j_{-n}\left[\alpha^i_n,\alpha^k_{-m}\right]\alpha^l_m\\
&\quad +\alpha^k_{-m}\left[\alpha^j_{-n},\alpha^l_m\right]\alpha^i_n+\left[\alpha^j_{-n},\alpha^k_{-m}\right]\alpha^l_m\alpha^i_n\\
&= \alpha^j_{-n}\alpha^k_{-m}n\delta^{il}\delta_{n,-m}+\alpha^j_{-n}\alpha^l_m n\delta^{ik}\delta_{n,m}\\
&\quad -\alpha^k_{-m}\alpha^i_n n\delta^{jl}\delta_{-n,-m}-\alpha^l_m\alpha^i_n n\delta^{jk}\delta_{-n,m}\\
&= n(\alpha^j_{-n}\alpha^k_n\delta^{il}+\alpha^j_{-n}\alpha^l_n\delta^{ik}-\alpha^k_{-n}\alpha^i_n\delta^{jl}-\alpha^l_{-n}\alpha^i_n\delta^{jk}), \qquad (9.36)
\end{aligned}$$

and finally the fourth

$$\begin{aligned}
\left[\alpha^j_{-n}\alpha^i_n,\alpha^l_{-m}\alpha^k_m\right] &= \alpha^j_{-n}\left[\alpha^i_n,\alpha^l_{-m}\alpha^k_m\right]+\left[\alpha^j_{-n},\alpha^l_{-m}\alpha^k_m\right]\alpha^i_n\\
&= \alpha^j_{-n}\alpha^l_{-m}\left[\alpha^i_n,\alpha^k_m\right]+\alpha^j_{-n}\left[\alpha^i_n,\alpha^l_{-m}\right]\alpha^k_m\\
&\quad +\alpha^l_{-m}\left[\alpha^j_{-n},\alpha^k_m\right]\alpha^i_n+\left[\alpha^j_{-n},\alpha^l_{-m}\right]\alpha^k_m\alpha^i_n\\
&= \alpha^i_{-n}\alpha^l_{-m}n\delta^{ik}\delta_{n,-m}+\alpha^j_{-n}\alpha^k_m\delta^{ij}\delta_{n,m}\\
&\quad -\alpha^l_{-m}\alpha^i_n n\delta^{jk}\delta_{-n,-m}-\alpha^k_m\alpha^i_n n\delta^{jl}\delta_{-n,m}\\
&= n(\alpha^j_{-n}\alpha^l_n\delta^{ik}+\alpha^j_{-n}\alpha^k_n\delta^{ij}-\alpha^l_{-n}\alpha^i_n\delta^{ik}-\alpha^k_{-n}\alpha^i_n\delta^{jl}). \qquad (9.37)
\end{aligned}$$

Combining (9.34-9.37) 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

$$\langle 1,k|\, S^{23}\, |1,k\rangle = -i\sum_{n=1}^{\infty}\frac{1}{n}\langle 1,k|\left(\alpha_{-n}^2\alpha_n^3-\alpha_{-n}^3\alpha_n^2\right)|1,k\rangle = \langle 1,k|\, i\, |1,k\rangle = i.$$

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.

9.3.4.1 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 (9.7) 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 (9.32) 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.

9.4 Virasaro algebra

9.4.1 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 (9.23)

$$[\alpha_m^\mu, \alpha_n^\nu] = m\delta_{m+n}\eta^{\mu\nu} \tag{9.38}$$

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

$$L_m = \frac{1}{2\pi\alpha'} \int_0^{2\pi} d\sigma e^{im\sigma} T_{\tau\tau}\bigg|_{\tau=0} = \frac{1}{2\pi\alpha'} \int_0^{2\pi} d\sigma e^{-im\sigma} T_{\sigma\sigma}\bigg|_{\tau=0} = \frac{1}{2} \sum_{n=-\infty}^{\infty} \alpha^\mu_{m-n} \alpha^\mu_n$$

for the open string and

$$\begin{aligned} L_m &= \frac{1}{2} \sum_{n=-\infty}^{\infty} \alpha^\mu_{m-n} \alpha^\mu_n \\ \tilde{L}_m &= \frac{1}{2} \sum_{n=-\infty}^{\infty} \beta^\mu_{m-n} \beta^\mu_n \end{aligned}$$

for the closed string. Note that the energy momentum tensor vanishes by being the equation of motion (9.30) 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 (9.38). 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 (9.26), 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 (9.23). However, it is more complicated if $m+n=0$. It is direct to show that for any m

$$[L_m,\alpha_n^\mu]=-n\alpha_{m+n}^\mu. \tag{9.39}$$

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.

9.4.2 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,

$$\langle p|z\rangle = 0, \tag{9.40}$$

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

$$L_{m>0}|p\rangle = 0 \tag{9.41}$$
$$(L_0 - a)|p\rangle = 0, \tag{9.42}$$

while spurious states obey besides the second condition (9.42) also (9.40) for all physical states. The correctness of this assignment follows from the fact that the conditions (9.41) and (9.42) 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 (9.42) and (9.40) 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

$$L_1|\chi_1\rangle = L_1 L_{-1}|\chi_1 >= 2L_0|\chi_1\rangle = 2(a-1)|\chi_1\rangle, \tag{9.43}$$

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