4 The Standard Model as an effective field theory

The Standard Model has some remarkable properties. Among these, the renormalizable terms respect a variety of symmetries, all of which are observed to hold to a high degree in nature:

- baryon number symmetry;

$$Q \to e^{i\alpha/3}Q, \quad \bar{u} \to e^{-i\alpha/3}\bar{u}, \quad \bar{d} \to e^{-i\alpha/3}\bar{d}; \tag{4.1}$$

- three separate lepton number symmetries,

$$L_f \to e^{-i\alpha_f}L_f, \quad \bar{e}_f \to e^{i\alpha_f}\bar{e}_f. \tag{4.2}$$

It is not necessary to *impose* these symmetries. They are simply consequences of gauge invariance and the fact that there are only so many renormalizable terms that one can write down. These symmetries are said to be "accidental", since they do not seem to result from any deep underlying principle.

This is already a triumph. As we will see when we consider possible extensions of the Standard Model, this did not have to be the case. But this success raises the question: why should we impose the requirement of renormalizability?

4.1 Integrating out massive fields

In the early days of quantum field theory, renormalizability was sometimes presented as a sacred principle. There was a view that field theories were fundamental and should make sense in and of themselves. Much effort was devoted to understanding whether the theories still existed in the limit where the cutoff was taken to infinity.

But there was an alternative paradigm for understanding field theories, provided by Fermi's original theory of weak interactions. In this theory, weak interactions are described by a Lagrangian of the form

$$\mathcal{L}_{\text{weak}} = \frac{G_f}{\sqrt{2}}J^\mu J_\mu. \tag{4.3}$$

Here the currents J^μ are bilinear in the fermions; they include terms like $Q\sigma^\mu T^a Q^*$. This theory, like the Standard Model, was very successful. It took some time to actually determine the form of the currents but, for more than 40 years, all experiments in weak interactions could be summarized in a Lagrangian of this form. Only as the energies of bosons in e^+e^- experiments approached the Z boson mass were deviations observed.

https://doi.org/10.1017/9781009290883.009 Published online by Cambridge University Press

The four-fermion theory is non-renormalizable. Taken seriously as a fundamental theory, it predicts violations of unitarity at TeV energy scales. But, from the beginning, the theory was viewed as an *effective* field theory, valid only at low energies. When Fermi first proposed the theory he assumed that the weak forces were caused by the exchange of particles – what we now know as the W and Z bosons.

4.1.1 Integrating out the W and Z bosons

Within the Standard Model we can derive the Fermi theory and also understand the deviations. A traditional approach is to examine the Feynman diagram of Fig. 4.1. This can be understood as a contribution to a scattering amplitude, but it is best understood here as a contribution to the effective action of the quarks and leptons. The currents of the Fermi theory are just the gauge currents which describe the coupling at each vertex. The propagator, in the limit of very small momentum transfer, is just a constant. In coordinate space this corresponds to a space–time δ-function; the interaction is local. The effect is just to give the four-fermion Lagrangian. One can consider the effects of small finite momentum by expanding the propagator in powers of q^2. This will give four-fermion operators with derivatives. These are suppressed by powers of M_W and their effects are very tiny at low energies. Still, in principle, they are there and in fact the measurement of such terms at energies that are a significant fraction of M_Z provided the first hints of the existence of the Z boson.

This effective action can also be derived in the path integral approach. Here we literally integrate out the heavy fields, the W and Z. In other words, for fixed values of the light fields, which we denote by ϕ, we perform a path integral over W and Z, expressing the result as an effective action for the ϕ fields (see Appendix C):

$$\int [d\phi] e^{iS_{\text{eff}}} = \int [d\phi] \int [dW_\mu][dW^*_\mu]\Delta_{FP} \times \exp\left[i \int d^4x (W^\dagger_\mu(\partial^2 + M_W^2)W^\mu + J^\mu W^\dagger_\mu + J^{\mu\dagger} W_\mu) \right]. \tag{4.4}$$

Here, for simplicity, we have omitted the Z particle. We have chosen the Feynman–'t Hooft gauge. The currents J^μ and $J^{\mu\dagger}$ are the usual weak currents. They are constructed out of

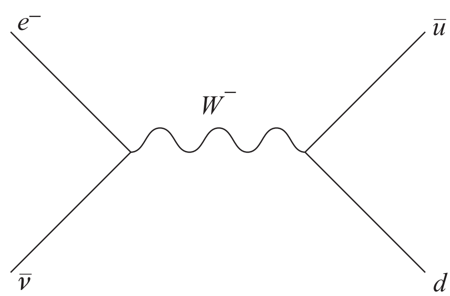

Fig. 4.1 Exchange of the massive W boson gives rise to the four-fermion interaction.

https://doi.org/10.1017/9781009290883.009 Published online by Cambridge University Press

the various light fields, the quarks and leptons, which we have grouped, generically, into the set of fields ϕ. Written in this way, this is the most basic field theory path integral, and we are familiar with the result:

$$e^{iS_{\rm eff}} = \exp\left[\int d^4xd^4y J^\mu(x)\Delta(x,y)J_\mu(y)\right]. \tag{4.5}$$

Here $\Delta(x,y)$ denotes the propagator for a scalar of mass M_W. In the limit $M \to \infty$ this is just a δ-function (one can compute this or see it directly from the path integral; if we neglect the derivative terms in the action, the propagator is just a constant in momentum space):

$$\Delta(x,y) = \frac{i}{M_W^2}\delta(x-y). \tag{4.6}$$

So

$$S_{\rm eff} = \frac{1}{M_W^2}J^{\mu\dagger}J_\mu. \tag{4.7}$$

The lesson is that, up to the late 1970s, one could view QED + QCD + the Fermi theory as a perfectly acceptable theory of particle interactions. The theory had to be understood, however, as an effective theory, valid only up to an energy scale of order 100 GeV or so. Sufficiently precise experiments would require the inclusion of operators of dimension higher than four. The natural scale for these operators would be the weak scale. The Fermi theory is ultraviolet divergent. These divergences would be cut off at scales of order the W boson mass.

4.1.2 The simplest Higgs boson, obtained from integrating out other physics at higher energies

It is possible that the Higgs boson is precisely the doublet of the minimal Standard Model, and that upcoming experiments will simply verify that its couplings to quarks, leptons, gauge bosons and itself are exactly those expected. But they might show deviations and, in any case, at least at the LHC these measurements will probably be good only to the 5%−10% level, leaving some room for possible deviations.

If there is new physics at scales of order a few TeV or less, these might affect the properties of the Higgs. One simple possibility is that there is a second Higgs doublet. In other words, there might be two Higgs doublets, ϕ_1 and ϕ_2, with a potential $V(\phi_1,\phi_2)$ and Yukawa couplings to the quarks and leptons. There are strong restrictions on these couplings from low-energy physics (and especially from phenomena like K–$\bar{K}$ mixing). These are satisfied, for example, if one Higgs doublet couples only to up quarks and the other only to down quarks. We will see, for example, in Chapter 11 that in supersymmetric theories these conditions are automatically satisfied, at least at tree level. But there are now further restrictions from the success of the Standard Model in accounting for the properties of the observed Higgs.

To see how these constraints might be satisfied and to see the connection with notions from effective field theory we will focus on the mass matrix for the Higgs fields. Take the quadratic terms to have the form

$$V_m = \mu_1^2|\phi_1|^2 + \mu_2^2|\phi_2|^2 + m^2\phi_1\phi_2. \quad (4.8)$$

Suppose that the mass-squared matrix has one positive and one negative eigenvalue. Take ϕ to correspond to the negative eigenvalue and H to correspond to the positive eigenvalue:

$$V = -\mu^2|\phi|^2 + m^2|H|^2 + \text{quartic}. \quad (4.9)$$

If $m^2 \gg \mu^2$ then we can integrate out H to obtain a potential for ϕ. This limit is referred to as the *decoupling limit* of the two-Higgs-doublet model; if there is a second Higgs doublet, either this, or so-called "alignment", must hold for consistency with the present experimental constraints.

At tree level the potential for $|\phi|^2$ includes a negative quadratic term and a positive quartic. There are also sixth- and higher-order terms, suppressed by powers of m^2. Loop corrections involving the heavy field provide further modifications. The Yukawa couplings are also of Standard Model type. Again, at tree level, if

$$\phi_1 = \cos\alpha\phi - \sin_\alpha \phi_2 H, \quad \phi_2 = \sin\alpha\phi + \cos\alpha H \quad (4.10)$$

then ϕ_1 and ϕ_2 have the Yukawa couplings

$$\mathcal{L}_y = y_1\phi_1 Q\bar{u} + y_2\phi_2 Q\bar{d}, \quad (4.11)$$

where y_1 and y_2 are matrices in the space of generations. It follows that the Yukawa couplings to the up quarks are $y_u \cos\alpha$ and those to the down quarks are $y_d \sin\alpha$.

4.1.3 What might the Standard Model come from?

As successful as the Standard Model is, and despite the fact that it is renormalizable, it is likely that, like the four-fermion theory, it is the low-energy limit of some underlying, more fundamental, theory. In the second half of this book our model for this theory will be string theory. Consistent theories of strings, for reasons which are somewhat mysterious, are theories which describe general relativity and gauge interactions. Unlike field theory, string theory is finite. It does not require a cutoff for its definition. In principle, all physical questions have well-defined answers within the theory. If this is the correct picture for the origin of the laws of nature at extremely short distances, then the Standard Model is just its low-energy limit. When we study string theory we will understand in some detail how such a structure can emerge. For now, the main lesson we should take concerns the requirement of renormalizability: the Standard Model should be viewed as an effective theory, valid up to some energy scale Λ. Renormalizability is not a constraint we impose upon the theory; rather, we *should* include operators of dimension five or higher, with coefficients scaled by inverse powers of Λ. The value of Λ is an experimental question. From the success of the Standard Model, as we will see, we know that the cutoff is large. From string theory we might imagine that $\Lambda \approx M_{\rm p} = 1.2 \times 10^{18}$ GeV. But, as we will now describe, we have experimental evidence that there is new physics which we must include at scales well below $M_{\rm p}$. We will also see that there are theoretical reasons to believe that there should be new physics at TeV energy scales.

4.2 Lepton and baryon number violation; neutrino mass

We have remarked that, at the level of renormalizable operators, baryon number and lepton number are conserved in the Standard Model. Viewed as an effective theory, however, we should include higher-dimension operators with dimensionful couplings. We would expect such operators to arise, as in the case of the four-fermion theory, as a result of new phenomena and interactions at very high energy scales. The coefficients of these operators would be determined by this dynamics.

There would seem, at first, to be a vast array of possibilities for operators which might be included in the Standard Model Lagrangian. But we can organize the possible terms in two ways. First, if $M_{\rm bsm}$ is the scale of some new physics, operators of progressively higher dimension will be suppressed by progressively larger powers of $M_{\rm bsm}$. Second, the most interesting and readily detectable operators are those which violate the symmetries of the renormalizable Lagrangian. This is already familiar in the weak interaction theory. In the Standard Model the symmetries are precisely baryon number and lepton number.

The existence of the neutrino mass is now well established, and several parameters governing these masses are known. As we will see, if the only degrees of freedom involved are the three known two-component neutrinos, the structure of the leading lepton-number-violating operators is known. Several combinations of parameters are determined by the current data, and measuring the remaining ones is a central component of the international (and especially the US) high-energy physics program for the next few decades. Determining whether there are additional degrees of freedom is another major component.

4.2.1 Dimension five: lepton number violation and neutrino mass

To proceed systematically, we should write down operators of dimension five, six and so on. At the level of dimension five, we can write several terms which violate lepton number:

$$\mathcal{L} = \frac{1}{M_{\rm bsm}} \gamma_{f,f'} \phi\phi L_f L_f' + \text{c.c.} \tag{4.12}$$

Here ϕ again denotes the Higgs doublet and the indices are contracted suitably. With non-zero ϕ these terms give rise to neutrino masses. This type of mass term is usually called a *Majorana mass*. In nature these masses are quite small. For example, if $M_{\rm bsm} = 10^{16}$ GeV, which we will see is a plausible scale, then the neutrino masses would be of order 10^{-3} eV. In typical astrophysical and experimental situations, neutrinos are produced with energies of order MeV or larger, so it is difficult to measure these masses by studying the energy–momentum dispersion relation (very sensitive measurements of the end-point spectra beta decay are sensitive to electronvolt-scale neutrino masses). More promising are oscillation experiments, in which these operators give rise to transitions between one type of neutrino and another, which are similar to the phenomenon of K meson oscillations. Roughly speaking, in the β-decay of a d quark, say, one produces the neutrino partner of the electron. However, the mass (energy) eigenstate is a linear combination of the three types of

https://doi.org/10.1017/9781009290883.009 Published online by Cambridge University Press

neutrino (as we will see, typically it is principally a combination of two). So, experiments or observations downstream from the production point will measure processes in which neutrinos produce muons or taus. The oscillation periods are of order $E/\Delta m^2$. For MeV neutrinos and $\Delta m \sim 10^{-3}$ eV, this corresponds to distances of order kilometers, which is of interest for neutrinos in the atmosphere or those observed near nuclear reactors; for lighter neutrinos, effects at solar system scales become of interest.

Evidence that neutrinos do have non-zero masses and mixings comes from the study of neutrinos coming from the Sun (the solar neutrinos) and neutrinos produced in the upper atmosphere by cosmic rays (which produce pions that subsequently decay to muons and ν_μs, whose decays in turn produce electrons, ν_μs, and $\nu_{\rm e}$s). Accelerator and reactor experiments have provided dramatic and beautiful evidence in support of this picture. It developed as a result of heroic experimental and theoretical work over more than four decades. The pioneering experiments were those of Ray Davis who, along with John Bahcall, conceived of neutrinos as a tool for the study of the interior of the Sun. His observation of neutrinos at rates lower than those expected in the standard solar model prompted the study of the mixing hypothesis and a range of other experiments. Later, studies of neutrinos from cosmic rays failed to yield the predicted fractions of ν_μs and ν_es. Dedicated studies of neutrinos from nuclear reactors and accelerators have provided further support for the mixing hypothesis and precise measurements of several parameters.

The masses and mixings of the neutrinos can be characterized by a unitary matrix, similar to the CKM matrix for the quarks, known as the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix. It can be parameterized as follows:

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}c_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \times \mathrm{diag}(1, e^{i\alpha_{21}/2}, e^{i\alpha_{31}/2}). \tag{4.13}$$

From the range of experiments described above, we know that

$$\begin{aligned} (\delta m^2)_{21} &= 7.54^{+0.26}_{-0.22} \times 10^{-5}\ \mathrm{eV}^2, \\ \delta m^2 = (\Delta m^2)_{31} - \Delta m^2_{12} &= 2.43 \pm 0.06 \times 10^{-3}\ \mathrm{eV}^2, \end{aligned} \tag{4.14}$$

where the second line holds if $m_1 < m_2$. With the same hierarchy, i.e. ordering of the masses, one has:

$$\begin{gathered} \sin^2\theta_{12} = 0.308 \pm 0.017, \quad \sin^2\theta_{23} = 0.437^{-.033}_{-0.023}, \quad \sin^2\theta_{13} = 0.0234^{+0.0020}_{-0.0019}, \\ \frac{\delta}{\pi} = 1.39^{+0.38}_{-0.27}. \end{gathered} \tag{4.15}$$

More detail can be found in the references cited at the end of this chapter.

It is conceivable that these masses are not described by the Lagrangian of Eq. (4.12). Instead, the masses might be Dirac, by which one means that there might be additional degrees of freedom; by analogy to the $\bar{e}$ fields we could label these by $\bar{\nu}$, and they would have very tiny Yukawa couplings to the normal neutrinos. This would truly represent a breakdown of the Standard Model: even at low energies, we would be missing basic

https://doi.org/10.1017/9781009290883.009 Published online by Cambridge University Press

degrees of freedom. But this does not seem likely. If there are singlet neutrinos N, nothing would prevent them from gaining a Majorana mass m_N, so that

$$\mathcal{L}_{\rm Maj} = m_N NN. \tag{4.16}$$

As for the leptons and quarks, there would also be a coupling of ν to the field N. There would now be a *mass matrix* for the neutrinos, involving both N and ν. For simplicity, consider the case of just one generation. Then this matrix would have the form

$$m_\nu = \begin{pmatrix} m_N & yv \\ yv & 0 \end{pmatrix}. \tag{4.17}$$

Such a matrix has one large eigenvalue, of order m_n, and one small eigen value, of order y^2v^2/M_N. This provides a natural way to understand the smallness of the neutrino mass; it is referred to as the *seesaw mechanism*. Alternatively, we could consider of integrating out the right-handed neutrino and generating the operator of Eq. (4.12).

It seems more plausible that the observed neutrino mass is Majorana than Dirac, but this is a question that hopefully will be settled in time by experiments searching for *neutrino-less double beta decay*, $n+n \to p+p+e^- + e^-$. If it is Majorana, this suggests that there is another scale in physics that is well below the Planck scale. For, even if the new Yukawa couplings are of order one, the neutrino mass is of order

$$m_\nu = 10^{-5}\ {\rm eV}(M_{\rm p}/\Lambda), \tag{4.18}$$

where Λ is another scale that is well below the Planck scale and M_P is the Planck mass. If the Yukawas are small, as are many of the quark Yukawa couplings, the scale can be much smaller.

4.2.2 Other symmetry-breaking dimension-five operators

There is another class of symmetry-violating dimension-five operators which can appear in the effective Lagrangian. These are electric and magnetic dipole moment operators. For example, the operator

$$\mathcal{L}_{\mu e} = \frac{e}{M_{\rm bsm}} F_{\mu\nu}\bar{\mu}\sigma^{\mu\nu} e \tag{4.19}$$

(we are using a four-component notation) would lead to the decay of the muon to an electron and a photon. Here $M_{\rm bsm}$ denotes the scale relating to Beyond the standard model physics. There are stringent experimental limits on such muon-number-violating processes, for example:

$$\text{branching ratio}(\mu \to e\gamma) < 1.2 \times 10^{-11}. \tag{4.20}$$

Other operators of this type include those which would generate lepton-number-violating τ decays, on which the limits are far less stringent.

In the Standard Model, CP is an approximate symmetry. We have explained that three generations of quarks are required to violate CP within the Standard Model. So, amplitudes which violate CP must involve all three generations and are typically highly suppressed. From an effective-Lagrangian viewpoint, if we integrate out the W and Z bosons then the

https://doi.org/10.1017/9781009290883.009 Published online by Cambridge University Press

operators which violate CP are of dimension six and typically have coefficients suppressed by quark masses and mixing angles, as well as loop factors. As a result, new physics at relatively modest scales has the potential for dramatic effects. Electric dipole moment operators for quarks or leptons would arise from operators of the form

$$\mathcal{L}_{\mathrm{d}} = \frac{em_q}{M_{\mathrm{bsm}}^2} \tilde{F}_{\mu\nu} \bar{q} \sigma^{\mu\nu} q + \text{ c.c.}, \tag{4.21}$$

where

$$\tilde{F}_{\mu\nu} = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} F^{\rho\sigma}. \tag{4.22}$$

Here $\epsilon_{\mu\nu\rho\sigma}$ is the completely antisymmetric tensor with four indices; $\epsilon_{0123} = 1$. The presence of the ϵ symbol is the signal of CP violation, as the reader can check. In the non-relativistic limit, this is $\vec{\sigma} \cdot \vec{E}$. These would lead, for example, to a neutron electric dipole moment of order

$$d_{\mathrm{n}} = \frac{e}{M_{\mathrm{bsm}}}. \tag{4.23}$$

Searches for such dipole moments set a limits $d_{\mathrm{n}} < 10^{-25} e$ cm. So, unless there is some source of suppression, M_{bsm} in CP-violating processes is larger than about 10^2 TeV.

4.2.3 Irrelevant operators and high-precision experiments

There are a number of dimension-five operators on which it is possible to set somewhat less stringent limits, and in one case there is a possible discrepancy. Corrections to the muon magnetic moment could arise from

$$\mathcal{L}_{g-2} = \frac{e}{M_{\mathrm{bsm}}} F_{\mu\nu} \bar{\mu} \sigma^{\mu\nu} \mu + \text{c.c.}, \tag{4.24}$$

where $F_{\mu\nu}$ is the electromagnetic field (in terms of the fundamental $SU(2)$ and $U(1)$ fields, one can write similar gauge-invariant combinations which reduce to this at low energies). The muon magnetic moment has been measured to extremely high precision, and its Standard Model contribution is calculated with comparable precision; as of the time of writing there is a 2.6σ discrepancy between the two. Whether this reflects new physics is uncertain. We will encounter one candidate for this physics when we discuss supersymmetry.

There are other operators on which we can set TeV-scale limits. The success of QCD in describing jet physics allows one to constrain four-quark operators which would give rise to a hard component in the scattering amplitude. Such operators might arise, for example, if quarks were composite. Constraints on flavor-changing processes provide tight constraints on a variety of operators. Operators such as

$$\mathcal{L}_{\mathrm{fc}} = \frac{1}{M_{\mathrm{bsm}}^2} s \sigma^{\mu} d^* s \sigma_{\mu} d^* \tag{4.25}$$

(where we have switched to a two-component notation) would contribute to $K\bar{K}$ mixing and other processes. This would constrain M_{bsm} to be larger than 100 TeV or so. Any new physics at the TeV scale must explain why such an operator is so severely suppressed.

https://doi.org/10.1017/9781009290883.009 Published online by Cambridge University Press

4.2.4 Dimension-six operators: proton decay

Proceeding to dimension six we can write down numerous terms which violate baryon number, as well as additional lepton-number-violating interactions:

$$\mathcal{L}_{\rm bv} = \frac{1}{M_{\rm bsm}^2} Q\sigma^{\mu}\bar{u}^* L\sigma_{\mu}\bar{d}^* + \cdots . \quad (4.26)$$

This can lead to processes such as $p \to \pi e$. Experiments deep underground set limits of order 10^{33} years on this process. Correspondingly, the scale $M_{\rm bsm}$ must be larger than 10^{15} GeV.

So, viewing the Standard Model as an effective-field theory, we see that there are many possible non-renormalizable operators which might appear but most have scales which are tightly constrained by experiment. One might hope – or despair – that the Standard Model will provide a complete description of nature up to scales many orders of magnitude larger than we can hope to probe in experiment.

However, there are a number of reasons to think that the Standard Model is incomplete, and at least one which suggests that it will be significantly modified at scales not far above the weak scale.

4.3 Challenges for the Standard Model

On the one hand, the Standard Model is tremendously successful. With the discovery of the Higgs particle, it can be said to describe the physics of strong, weak and electromagnetic interactions with great precision to energies of order 100 GeV or distances as small as 10^{-17} cm. It explains why baryon number and the separate lepton numbers are conserved, with only one assumption: there is no interesting new physics up to some high-energy scale. As of the end of the 8 TeV run at the LHC, there are almost no discrepancies between theory and experiment.

On the other hand, the Standard Model cannot be a complete theory. The existence of neutrino mass requires at least additional states (if these masses are Dirac), and more likely some new physics at a high-energy scale which accounts for the Majorana neutrino masses. This scale is probably not larger than 10^{16} GeV, well below the Planck scale. The existence of gravity means that there is certainly something missing from the theory. The plethora of parameters – there are 19, counting those of the minimal Higgs sector and the θ parameter (see the next subsection) – suggests that there is a deeper structure. More directly, features of the big bang cosmology which are now well established cannot be accommodated within the Standard Model.

4.3.1 The strong CP problem

In the Standard Model there is a puzzle even at the level of dimension-four operators. Consider

$$\mathcal{L}_{\theta} = \theta F\tilde{F}, \quad (4.27)$$

https://doi.org/10.1017/9781009290883.009 Published online by Cambridge University Press

where θ is a dimensionless parameter and

$$\tilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}. \tag{4.28}$$

We usually ignore such operators because classically they are inconsequential; they are total derivatives and do not modify the equations of motion. In a $U(1)$ theory, for example,

$$F\tilde{F} = 2\epsilon^{\mu\nu\rho\sigma}\partial_\mu A_\nu \partial_\rho A_\sigma = 2\partial_\mu(\epsilon^{\mu\nu\rho\sigma}A_\nu\partial_\rho A_\sigma). \tag{4.29}$$

In the next chapter we will see that this has a non-Abelian generalization, but that, despite constituting a total divergence, these terms have real effects at the quantum level. In QCD they turn out to be highly constrained. From the limits on the neutron electric dipole moment, we will show in Chapter 5 that $\theta < 10^{-9}$. This is the first real puzzle we have encountered. Why is it such a small dimensionless number? Answering this question, as we will see in Chapter 5, may point to new physics, likely at some very high energy scale.

4.3.2 The hierarchy problem and the question of naturalness

The second very puzzling feature in the Standard Model is the Higgs field. The fact that the model seems to be described by a single Higgs scalar is itself puzzling. We could have included several doublets or perhaps tried to explain the breaking of the gauge symmetry through some more complicated dynamics, as we will discuss in Chapter 8. But there is a more serious question associated with fundamental scalar fields, raised long ago by Ken Wilson. This problem is often referred to as the *hierarchy problem* or the *naturalness problem*.

Consider, first, the one-loop corrections to the electron mass in QED. These are logarithmically divergent. In other words,

$$\delta m = am_0 \frac{\alpha}{4\pi} \ln \Lambda. \tag{4.30}$$

We can understand this result in simple terms. In the limit $m_0 \to 0$ the theory has an additional symmetry, a chiral symmetry, under which e and $\bar{e}$ transform by independent phases. This symmetry forbids a mass term, so the result must be linear in the (bare) mass. So, on dimensional grounds, any divergence is at most logarithmic. This actually resolves a puzzle of *classical* electrodynamics. Lorentz modeled the electron as a uniformly charged sphere of radius a. As $a \to 0$ the electrostatic energy diverges. In modern terms, we would say that we know a is smaller than 10^{-17} cm, corresponding to a self-energy far larger than the electron mass itself. But we see that in the quantum theory the cutoff occurs at a scale of order the electron mass, and there is no large self-energy correction.

For scalars, however, there is no such symmetry and corrections to masses are quadratically divergent. One can see this easily for the Higgs self-coupling, which gives rise to a mass correction of the form

$$\delta m^2 = \lambda \int \frac{d^4k}{(2\pi)^4(k^2+m^2)}, \tag{4.31}$$

https://doi.org/10.1017/9781009290883.009 Published online by Cambridge University Press

with similar corrections from the top quark loop correction, gauge loops, and others. If we view the Standard Model as an effective-field theory, these integrals should be cut off at a scale where new physics enters. We have argued that this might occur at, say, 10^{14} GeV. But in this case the correction to the Higgs mass would be gigantic compared with the Higgs mass itself. Given that $y_t^2 > \lambda$, we would expect even larger effects from top quark loops.

It is hard to see how this puzzle can be resolved without introducing new physics at a scale not much larger than 1 TeV. Exploring candidates for this new physics will be one of the major subjects of this book. After discussing another fine tuning problem in our current understanding of the laws of nature, we will elevate these concerns to a principle that we might wish to impose on our theories: the principle of naturalness.

4.3.3 The universe: the baryon density, dark matter and dark energy

As we will discuss in Chapter 18, we have good evidence that the energy density of the universe occurs largely in unfamiliar forms: about 27% in non-baryonic pressureless matter (dark matter) and about 68% in some form having with negative pressure (dark energy), with only the remaining 5% comprising ordinary baryons. The dark energy is likely to be a cosmological constant (of which more later).

As we will discuss, particularly in Chapter 19, we might hope to understand the dark matter in terms of some type(s) of new particle. A particle with mass of order 1 TeV (give or take factors of 10) and roughly weak-interaction cross sections would be produced in suitable quantities in the early universe. Beyond the hierarchy problem, this might be another pointer to new physics in the TeV energy range. Alternatively the axion, a much lighter and more weakly interacting particle proposed to solve the strong CP problem, might play this role and would lead to different types of experimental signals.

The baryon density, as we will also see, cannot arise from the Standard Model itself. We will consider a number of possible new physics mechanisms by which it might arise. Without strong assumptions about the history of the universe, it is difficult to pin down the relevant energy scale.

The dark energy raises puzzles which do not point in any obvious way to a particular energy scale. If the dark energy is a cosmological constant then this represents, from the perspective of our effective Lagrangian, a term of dimension zero, whose coefficient has dimensions of $(\text{mass})^4$. Dimensional analysis would suggest that it should be of order the largest possible scale to the fourth power. If this is the Planck scale then dimensional analysis fails by 120 orders of magnitude. In a sense our analysis of the effective action seems back to front. We began with a discussion of dimension-five and dimension-six operators, operators which are irrelevant, and then turned our attention to the Higgs mass, a dimension-two, relevant, operator. We still have not considered the most relevant operator of all, the unit operator.

In quantum field theory, consistently with dimensional analysis, this energy is *quartically* divergent; it is the first divergence one encounters in any quantum field theory textbook. At one loop it is given by an expression of the form

https://doi.org/10.1017/9781009290883.009 Published online by Cambridge University Press

$$\Lambda = \sum_i (-1)^F \int \frac{d^3k}{(2\pi)^3} \frac{1}{2}\sqrt{k^2 + m_i^2}, \tag{4.32}$$

where the sum is over all particle species (including spins). This is just the sum of the zero-point energies of the oscillators of each momentum. If one cuts this off, again at 10^{14} GeV, one gets a result of order

$$\Lambda = 10^{54}\ \text{GeV}^4. \tag{4.33}$$

The measured value of the dark-energy density is by contrast,

$$\Lambda = 10^{-47}\ \text{GeV}^4. \tag{4.34}$$

This wide discrepancy is probably one of the most troubling problems facing fundamental physics today.

4.4 The naturalness principle

Both the Higgs mass and the cosmological constant appear to be finely tuned; they are much smaller than the values we would have guessed from dimensional analysis, and we have seen that quantum corrections are likely to be much larger than the observed parameters themselves. In contrast, we have noted that the electron mass (and the masses of the leptons and quarks more generally), while surprisingly small, does not receive large quantum corrections.

While many physicists were uncomfortable with these tunings, it was 't Hooft who framed this question in terms of a principle, which he dubbed the *naturalness condition*. He argued that a parameter in nature should be small only if the underlying theory becomes more symmetric as the parameter tends to zero. The electron mass in QED provides an illustration of this principle: as it tends to zero, the theory, as we have described, develops a new symmetry, a $U(1)$ chiral symmetry. All the small Yukawa couplings of the Standard Model are similarly natural. We will see that the small masses (relative to the Planck scale) of the hadrons are also compatible with the principle.

Our two puzzling quantities do not satisfy this criterion. The Standard Model does not become more symmetric if one sets the Higgs mass to zero. Similarly, general relativity (as we will see) does not become more symmetric as the cosmological constant tends to zero. The small value of the θ parameter, which violates CP conservation in strong interactions, also poses puzzles. Because the Standard Model violates CP even in the absence of θ, this would seem another violation of naturalness.

These issues each suggest that there should be some new degrees of freedom, or symmetries, or both, beyond those of the Standard Model. This has motivated a broad range of proposals for new physics. These will be the subject of much of this book. But, in recent years, at least one alternative picture for how the parameters of the Standard Model might arise has gained traction. We will consider this idea, known as the *landscape*, in Chapter 30.

4.5 Summary: successes and limitations of the Standard Model

Overall, we face a tension between the striking successes of the Standard Model and its limitations. On the one hand, the model successfully accounts for almost all the phenomena observed in accelerators. On the other hand, it fails to account for some of the most basic phenomena of the universe: dark matter, dark energy and the existence of gravity itself. As a theoretical structure, it also explains successfully what might be viewed as mysterious conservation laws: baryon number and lepton number. But it has 17 parameters – 16 of which are pure numbers, with values which range "all over the map". The rest of this book explores possible solutions of these puzzles, and their implications for particle physics, astrophysics and cosmology.

Suggested reading

The texts by by Peskin and Schroeder (1995) and Schwartz (2014) provide a good introduction both to weak interactions and also to the strong interactions; it includes deep inelastic scattering, parton distributions and the like. Other excellent texts include the books by Cheng and Li (1984), Donoghue *et al.* (1992), Pokorski (2000) and Bailin and Love (1993) among many others. For summaries of data on neutrino oscillations, the Particle Data Group website provides up-to-date reviews; the text by Barger *et al.* (2012) provides a first-rate pedagogical introduction.