# Field theory and the Standard Model: A symmetry-oriented approach

*Luis Álvarez-Gaumé* [a,b] *and Miguel Á. Vázquez-Mozo* [c]

[a] Simons Center for Geometry and Physics, Stony Brook University, New York 11794-3636, USA.
E-mail: lalvarezgaume@scgp.stonybrook.edu

[b] Theory Department CERN, CH-1211 Geneva 23, Switzerland.
E-mail: Luis.Alvarez-Gaume@cern.ch

[c] Departamento de Física Fundamental, Universidad de Salamanca, Plaza de la Merced s/n, E-37008 Salamanca, Spain.
E-mail: Miguel.Vazquez-Mozo@cern.ch

The Standard Model of particle physics represents the cornerstone of our understanding of the microscopic world. In these lectures we review its contents and structure, with a particular emphasis on the central role played by symmetries and their realization. This is not intended to be an exhaustive review but a discussion of selected topics that we find interesting, with the specific aim of clarifying some subtle points and potential misunderstandings. A number of more technical topics are discussed in separated boxes interspersed throughout the text.



This article should be cited as: Field theory and the Standard Model: A symmetry-oriented approach, Luis Álvarez-Gaumé and Miguel Á Vázquez-Mozo, DOI: 10.23730/CYRSP-2025-001.1, in: Proceedings of the 2022 European School of High-Energy Physics,
CERN Yellow Reports: School Proceedings, CERN-2025-001, DOI: 10.23730/CYRSP-2025-001, p. 1.

## 1 Preliminaries

Quantum field theory (QFT) is the language in which we codify our knowledge about the fundamental laws of nature in a manner compatible with quantum mechanics, relativity, and locality. Its most significant achievement has been formulating the Standard Model (SM) of strong, weak, and electromagnetic interactions. This theory summarizes what we know about the physics of the fundamental constituents of matter. It also delineates our ignorance, providing a glimpse of the known unknowns that will motivate future research. The story of QFT and the SM has been told many times with various degrees of detail and depth (see Refs. [1–18] for a necessarily incomplete sample of books on both topics). In the pages reserved for these lecture notes, it is utterly impossible to provide a detailed account of the towering achievements accumulated since the discovery of the electron by J. J. Thomson in 1897, whose most recent milestone was the announcement in 2012 of the discovery of the Higgs boson at CERN. Generations of physicists and engineers have made possible the formulation of a theory describing the most fundamental laws of nature known so far.

High energy physics is not the only arena in which QFT has shown its powers. In the nonrelativistic regime, it leads to quantum many body theory, a mathematical framework used in condensed matter physics to study phenomena such as superconductivity, superfluidity, and metals' thermal and electronic properties [21–23]. Furthermore, in the last few decades QFT has also played a central role in understanding the formation of the large scale structure of the universe [24–26].

Exciting as all these developments are, these lectures will focus on the applications of QFT to particle physics and particularly the construction of the SM. We will highlight symmetry arguments to show how virtually all known forms of symmetry realizations play a role in it. But even within this

restricted scope, space limitations require choosing not just the material to include but also the viewpoint to adopt. In explaining some of the ideas and techniques in our study of the SM, it is useful to focus on several key concepts, many of which are related to implementing symmetries in a quantum system with infinite degrees of freedom. In doing so, we will encounter many surprises and some misconceptions to be clarified. Explaining physics can be compared to the performance of a well-known piece of music. Often the performer surprises the audience by accentuating some features of the work that only then are sufficiently appreciated. In such a vein, we will highlight some important fundamental aspects of the SM the reader may not have encountered previously, some of which also point to the limitations of the theory. Although we will not shy away from diving into calculations when needed, our aim here is less giving a detailed account of the technicalities involved than providing the reader with both essential conceptual tools and inspiration to further deepen in the study of the topics to be presented.

Having set our plan of action, we turn to physics and begin by reviewing the system of units to be used throughout the lectures. Since we are dealing with quantum relativistic systems, it is natural to work with natural units, where the speed of light and the Planck constant are both set to $c = \hbar = 1$. Doing a bit of dimensional analysis, it is easy to see that setting these two fundamental constants to 1 means that of the three fundamental dimensions $L$ (length), $T$ (time), and $M$ (mass) only one is independent. Indeed, from $[c] = LT^{-1}$ and $[\hbar] = ML^2T^{-1}$ it follows that $T = L$ and $M = L^{-1}$, meaning that time has the dimension of length and masses of $(\text{length})^{-1}$. Alternatively, we may prefer to use energy ($E$) as the fundamental dimension, as we will actually do in the following. In this case, from $[\text{energy}] = ML^2T^{-2}$ we see that both lengths and times have dimensions of $(\text{energy})^{-1}$, while masses are measured in units of energy.

Using natural units simplifies expressions by eliminating factors of $\hbar$ and $c$ and brings other advantages. The most relevant for us is that it provides a simple classification of the operators, or terms, appearing in the action or Hamiltonian defining a theory. As an example, let us consider the scalar field action

$$S = \int d^4x \left( \frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - \frac{m^2}{2}\phi^2 - \frac{\lambda_4}{4!}\phi^4 - \frac{\lambda_6}{6!}\phi^6 \right). \tag{1.1}$$

Action is measured in the same units as $\hbar$ (not by chance historically known as the quantum of action) and is therefore dimensionless in natural units. Taking into account that $[d^4x] = E^{-4}$ and $[\partial_\mu] = E$, we find from the kinetic term that $[\phi] = E$, which in turn confirms that $[m] = E$ as behooving a mass. As for the coupling constants, $\lambda_4$ is dimensionless while $[\lambda_6] = E^{-2}$.

Terms such as $\phi^6$, whose coupling constants have negative energy dimension, are called higher-dimensional operators. In the modern (Wilsonian) view of QFT to be discussed in Section 10, they are seen as induced by physical processes above some energy scale $\Lambda$, much higher than the energy at which we want to describe the physics using the corresponding action. The presence of higher-dimensional operators in the action signals that we are dealing with a theory that is not fundamental, but some effective description valid at energies $E \ll \Lambda$, that should eventually be replaced (completed) by some more fundamental theory at higher energies.

Although the action of an effective field theory (EFT) may contain an infinite number of higher-dimensional operators of arbitrary high dimension, this does not make it any less predictive at low en-

ergies [27, 28]. To understand this, let us look at a higher-dimensional operator $\mathcal{O}_n$, with $[\mathcal{O}_n] = E^{n-4}$ for $n > 4$, entering in the action as

$$S \supset \frac{g_n}{\Lambda^{n-4}} \int d^4x\, \mathcal{O}_n, \tag{1.2}$$

where $g_n$ is a dimensionless coupling. The correction induced by this term to processes occurring at energy $E$ scales as $(E/\Lambda)^{n-4}$, so for $E \ll \Lambda$ there is a clear hierarchy among the infinite set of higher-dimensional operators. The upshot is that using our EFT to ask physical questions at sufficiently low energies, and taking into account the limited sensitivity of our detectors, only a small number of higher-dimensional operators have to be considered in the computation of physical observables.

Applying the philosophy of EFT to the action (1.1) leads to identify the theory as an effective description valid at energies well below the scale set by $\lambda_6$, namely $\Lambda \sim 1/\sqrt{\lambda_6}$. Nature offers more interesting implementations of this scheme, some of which we will encounter later on in the context of the SM. A particularly relevant case is that of general relativity (GR), that we discuss now in some detail. We start with the Einstein–Hilbert action

$$S = \frac{1}{16\pi G_N} \int d^4x\, \sqrt{-g} R, \tag{1.3}$$

and consider fluctuations around the Minkowski metric (nonflat background metrics can also be used)

$$g_{\mu\nu} = \eta_{\mu\nu} + 2\kappa h_{\mu\nu}, \tag{1.4}$$

where

$$\kappa \equiv \sqrt{8\pi G_N}. \tag{1.5}$$

Inserting (1.4) into (1.3) and expanding in powers of $h_{\mu\nu}$ we get an action defining a theory of interacting gravitons propagating on flat spacetime [29–31]. Its interaction part contains an infinite number of terms with the structure

$$S_{\text{int}} = \sum_{n=3}^{\infty} \kappa^{n-2} \int d^4x\, \mathcal{O}_{n+2}[h, \partial], \tag{1.6}$$

where the operator $\mathcal{O}_{n+2}[h, \partial]$, which has energy dimension $n + 2$, contains $n$ graviton fields and two derivatives, while from Eq. (1.5) we see that the coupling constant has dimension $[\kappa] = E^{-1}$. In the spirit of EFT, this indicates that Einstein's gravity is not fundamental, but an effective description valid at energies below its natural energy scale set by the dimensionful gravitational constant, the so-called Planck scale

$$\Lambda_{\text{Pl}} \equiv \sqrt{\frac{\hbar c^5}{8\pi G_N}} = 2.4 \times 10^{18}\ \text{GeV}, \tag{1.7}$$

where we have restored powers of $\hbar$ and $c$. To get an idea of the size of this scale, let us just say it is about $10^{14}$ times the center-of-mass energy at which LHC currently operates.

The statement is occasionally encountered in the literature and the media that GR is impossible to quantize. This needs to be qualified. The effective action (1.6) can be consistently quantized provided we restrict our physical questions to the range of energies where it can be used, namely $E \ll \Lambda_{\rm Pl}$. In this regime, the quantum fluctuations of the background metric shown in (1.4) are of order $E/\Lambda_{\rm Pl}$ and, therefore, small. Furthermore, powers of this same quantity suppress the induced corrections and, at the level of accuracy set by our experiments, only a small number of operators in (1.6) need to be retained to compute physical observables. In other words, below the Planck energy scale quantum gravity is just a theory of weakly coupled gravitons propagating on a regular background spacetime.

This state of affairs breaks down when the energy gets close to $\Lambda_{\rm Pl}$. At this point the quantum fluctuations of the geometry become large and the hierarchy of terms in (1.6) breaks down. Physically, what happens is that our gravitons become strongly coupled and therefore cease to be the appropriate degrees of freedom to describe a quantum theory of gravity. Thus, the correct statement is not that there is no consistent theory of quantum gravity, but that we lack one *which remains valid at arbitrarily high energies*. The difference is crucial, since it is precisely the latter kind of theory needed to analyze, for example, what happens close to spacetime singularities, where quantum effects are so large as to override the semiclassical description provided by GR. Viewed as an EFT, Einstein's (quantum) gravity is expected to be subsumed near $\Lambda_{\rm Pl}$ into another theory, its ultraviolet (UV) completion, which presumably remains valid to arbitrarily high energies. Among the particle physics community string theory continues to be the favored candidate for such a framework (see, for instance, Ref. [32, 33] for a modern account).

The previous digression on EFTs leads us to the related issue of renormalizability, on which we will further elaborate in Section 10. All QFTs used in describing elementary particles, particularly the SM, lead to infinities when computing quantum corrections (terms of order $\hbar$ or higher) to classical results. The origin of these divergences lies in the behavior of the theory at very high energies. Quantum fluctuations of very short wavelength actually dominate the result, driving them to infinity. This problem was tackled already in the 1940s by the procedure of renormalization. To make a long story short, one begins by regularizing the theory by setting a maximum energy $\Lambda$, a cutoff, so fluctuations with wavelength smaller than $\Lambda^{-1}$ are ignored. This makes all results finite, albeit dependent on the otherwise arbitrary cutoff. The key observation now is that the parameters in the action (field normalizations, masses, and coupling constants) can depend on $\Lambda$, so physical observables are cutoff independent. For this to work, a further ingredient is needed: an operational definition of masses and couplings, which serves to fix the dependence of the action parameters on the cutoff (for all the details see, for example, Chapter 8 of Ref. [14] or any other of the QFT textbooks listed in the references).

In carrying out this program, two things may happen. One is that divergences can be removed with a finite number of operators in the action (most frequently, just those already present in the classical theory). This is the case of a renormalizable theory. The second situation arises when it is necessary to add an infinite number of new operators in order to absorb all the divergences in their corresponding couplings. The theory is then said to be nonrenormalizable. The SM belongs to the first type, while GR is an example of the second. As a rule of thumb, actions containing operators of dimension equal or smaller than four define renormalizable theories, while the presence of higher-dimensional operators renders the theory nonrenormalizable, at least when working in perturbation theory.

For decades, renormalizability was considered necessary for any decent theory of elementary particles. The very formulation of the SM and, most particularly, its implementation of the Brout–Englert–Higgs (BEH) mechanism [34–36] through the Higgs boson was guided by making the theory renormalizable. As a token of how important this requirement was perceived to be at the time, let us mention that the electroweak sector of the SM developed by Sheldon L. Glashow, Steven Weinberg, and Abdus Salam [37–39] only started to be taken seriously by the particle physics community after Gerard 't Hooft and Martinus Veltman mathematically demonstrated its renormalizability [40, 41].

From a modern perspective, however, the condition that a theory must be renormalizable is regarded as too restrictive, equivalent to requiring that it remains valid at all energies. As a matter of fact, there is no reason to exclude nonrenormalizable theories from our toolkit. They can be interpreted as EFTs whose natural energy scale is set by the cutoff $\Lambda$, giving accurate results for processes involving energies $E \ll \Lambda$. Furthermore, from this viewpoint, the cutoff ceases to be a mere mathematical artefact to eventually be hidden in the action parameters. Instead, it acquires a physical significance as the energy threshold of the unknown physics encoded in the higher dimensional operators of our EFTs. Otherwise expressed, nonrenormalizability has lost its bad reputation and now is taken as a hint that some unknown physics is lurking at higher energies.

To make the previous discussion more transparent, let us look at the important case of quantum chromodynamics (QCD), the theory describing the interaction of quarks and gluons. QCD is not just a renormalizable theory that can be extrapolated to arbitrary energies, but asymptotically free as well. This means that its coupling constant approaches zero as we go to higher energies, thus making perturbation theory more and more reliable. The issue, however, is that when studying its low energy dynamics, the QCD coupling grows as we decrease the energy and the theory becomes strongly coupled. This has to be handled in a way somehow reminiscent of what we explained when discussing quantum GR near the Planck scale: below a certain energy scale $\Lambda_{\rm QCD}$ we need to abandon the perturbative QCD (pQCD) description in terms of quarks and gluons, now strongly coupled, and find the "right", weakly coupled, degrees of freedom to build an operative QFT. But, simultaneously, we have a huge advantage w.r.t. the gravity case. There, the trouble arose in the unexplored region of extremely high energies, where identifying the appropriate degrees of freedom, their interactions, or just the right framework remains anybody's guess (strings? spin foam? causal sets?). By contrast, life is much easier in QCD. The problematic regime happens at low energies, so to identify the weakly coupled degrees of freedom, we only need to "look", i.e., do experiments. From them, we learn that the physics has to be described in terms of mesons and baryons, whose interactions are largely fixed by symmetries (an issue to which we will come back later). What is relevant for the present discussion is that the appropriate framework, chiral perturbation theory ($\chi$PT), is a nonrenormalizable QFT whose action contains a plethora of higher-dimensional operators. Its cutoff, however, is not some arbitrary energy $\Lambda$ whose role is just to make the theory finite, but the physical scale $\Lambda_{\rm QCD}$ at which quarks and gluons get confined into hadrons. The theory of hadron interactions should then be understood as an EFT valid at energies $E \ll \Lambda_{\rm QCD}$.

The existence of the Planck scale at which quantum gravity is expected to become the dominant interaction has led to the realization that all quantum field theories have to be regarded as EFTs with a limited range of validity. This includes even renormalizable theories that, like the SM, are well-defined in a wide range of energies. However, explaining some experimental facts, such as nonzero neutrino

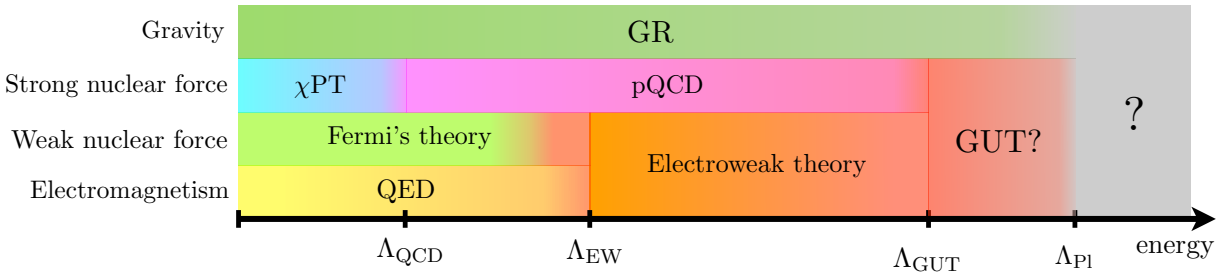



**Fig. 1:** Simplified cartoon showing the network of EFTs behind our understanding of subatomic physics.

masses, might require adding higher-dimensional operators to the theory, setting the energy scale for new physics to be explored in future high-energy facilities. At this energy, the SM will be superseded, maybe by some grand unified theory (GUT), which in turn is expected to break down at $\Lambda_{\rm Pl}$. It is in this sense that EFTs provide the foundational framework to understand nature at the smallest length scales (see Fig. 1).

## 2 From symmetry to physics

Symmetry is a central theme of contemporary physics, although its tracks go back a long way in history. More or less in disguise, symmetry-based arguments can be found in natural philosophy since classical times. In his refutation of vacuum in the fourth book of *Physics* (215a), Aristotle used the homogeneity of empty space to conclude the principle of inertia, that he however regarded as an inconsistency since it contradicted his first principle of motion: whatever moves has to be moved by something else. Galileo Galilei's assumption that reversing the velocity with which a free-rolling ball arrives at the basis of an inclined plane would make it climb exactly to the height from which it was released can be also regarded as an early *de facto* application of time reversal symmetry.

Although the origins of the mathematical study of symmetry are traced back to the first half of the 19th century with the groundbreaking works on group theory of Evariste Galois and Niels Henrik Abel, its golden age was ushered in by Felix Klein's 1872 Erlangen Program [42, 43]. Its core idea is that different geometries can be fully derived from the knowledge of the group of transformations preserving its objects (points, angles, figures, etc.). This establishes at the same time a hierarchy among geometries, determined by the relative generality of their underlying symmetry groups. In this way, Euclidean, affine, and hyperbolic geometries can be retrieved from projective geometry by restricting its group of transformations.

As an example, the whole plane Euclidean geometry emerges from the invariance under the combined action of rotations and rigid translations

$$x'^i = R^i{}_j x^j + a^i, \tag{2.1}$$

where $R^i{}_j \in \mathrm{SO}(2)$ and $a^i$ is an arbitrary two-dimensional vector. These two transformations build together the Euclidean group $E(2) \equiv \mathrm{ISO}(2)$, leaving invariant the Euclidean distance between two

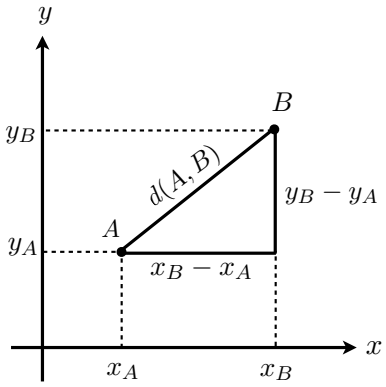



**Fig. 2:** Euclidean distance between two points on the plane.

points $A$ and $B$ with Cartesian coordinates $A = (x_A, y_A)$ and $B = (x_B, y_B)$,

$$d(A,B) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}, \tag{2.2}$$

which is just an application of the Pythagorean theorem (see Fig. 2). In a similar fashion, the geometry on the complex projective line $\mathbb{CP}^1$ (a.k.a. the Riemann sphere) follows from the invariance of geometrical objects under the projective linear group $\mathrm{PGL}(2,\mathbb{C})$, acting through Möbius transformations on $\mathbb{C}\cup\{\infty\}$

$$z' = \frac{az+b}{cz+d}, \tag{2.3}$$

where $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$. Among the invariants in this case are the four-point cross ratios associated with four points with complex coordinates $z_1$, $z_2$, $z_3$, and $z_4$

$$\mathrm{CR}(z_1, z_2, z_3, z_4) \equiv \frac{(z_1 - z_3)(z_2 - z_4)}{(z_2 - z_3)(z_1 - z_4)}, \tag{2.4}$$

as well as the chordal distance between two points $A$ and $B$ on the Riemann sphere

$$d(A,B)_{\mathrm{chordal}} = \frac{2|z_A - z_B|}{\sqrt{(1+|z_A|^2)(1+|z_B|^2)}}. \tag{2.5}$$

Möbius transformations preserve angles and maps circles to circles, so from a Kleinian point of view they are *bona fide* geometrical objects on $\mathbb{CP}^1$.

Klein's association of geometry and symmetry (i.e., group theory) revolutionized mathematics and became a game changer in physics. Beyond all early tacit uses, the systematic implementation of symmetry in physics had to wait until the end of the 19th century. In 1894 Pierre Curie used group theoretical methods to study the role of spatial symmetries in physical phenomena [44], thus introducing mathematical tools so far only applied in crystallography. This inaugurated a trend taken up later by the emerging fields of relativity and atomic physics, that led to key results like Emmy Noether's two celebrated theorems linking symmetries with conserved charges [45] (see Section 5.2).

### 2.1 Relativity from geometry

A beautiful example of geometry emerging from symmetry is provided by the geometrization of special relativity carried out in 1908 by Hermann Minkowski[1]. Einstein's formulation of special relativity in terms of events occurring in some instant $t$ at some position $\mathbf{r}$ (as measured by some inertial observer) leads naturally to introducing the four-dimensional space of all potential events, each represented by a point with spacetime coordinates $(t, \mathbf{r})$. Although switching from one inertial observer to another changes the individual coordinates of the events, the invariance of the speed of light implies the existence of an invariant. Given two arbitrary events taking place at points $\mathbf{r}$ and $\mathbf{r} + \Delta\mathbf{r}$, and separated by a time lapse $\Delta t$, their "spacetime separation"

$$\Delta s^2 \equiv \Delta t^2 - (\Delta \mathbf{r})^2 \tag{2.6}$$

remains the same for all inertial observers. The existence of this invariant with respect to the reference frame transformations introduced by Lorentz, Poincaré, and Einstein (and named after the first one) makes it natural to endow the space of events, or spacetime for short, with the metric

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2. \tag{2.7}$$

This is how spacetime geometry originates from the postulate of invariance of the speed of light.

We can take advantage of the language of tensors and write the line element (2.7) in the form

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \tag{2.8}$$

where $(x^0, x^1, x^2, x^3) \equiv (t, x, y, z)$ and $\eta_{\mu\nu} \equiv \text{diag}(1, -1, -1, -1)$ is the Minkowski metric. The most general linear transformation leaving invariant (2.8) [or (2.7)] is written as

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu, \tag{2.9}$$

where $\Lambda^\mu{}_\nu$ satisfies

$$\eta_{\mu\nu} = \eta_{\alpha\beta} \Lambda^\alpha{}_\mu \Lambda^\beta{}_\nu, \tag{2.10}$$

and $a^\mu$ is an arbitrary constant vector. The linear coordinate change (2.9) generates the Poincaré group, $\text{ISO}(1,3)$, that includes all transformations $\Lambda^\mu_\nu$ in the Lorentz group $\text{SO}(1,3)$ in addition to rigid translations. Notice that $\Lambda^\mu_\nu$ is a $4 \times 4$ matrix with 16 real components, so that the ten conditions (2.10) reduce to six independent ones. They correspond to the three parameters of a three-dimensional rotation (e.g., the Euler angles) plus the three velocity components of a generic boost. Adding the four real numbers determining a spacetime translation, we conclude that the Poincaré transformation (2.9) depends on ten independent real parameters.

Besides the invariance of the speed of light, Einstein's special relativity is also based on a second postulate, that all laws of physics take the same form for any inertial observer. This can also be recast in

[1] Einstein actually dubbed Minkowski's idea *überflussige Gelehrsamkeit* (superfluous erudition) [46], although geometrization later turned out to be the basis of his general theory of relativity.

geometric language by demanding that all equations of physics be expressed as tensor identities with the structure

$$T^{\mu_1\dots\mu_k}_{\nu_1,\dots,\nu_n}(x) = 0. \tag{2.11}$$

Under the generic Poincaré transformation (2.9), the previous equation changes as

$$T'^{\mu_1\dots\mu_k}_{\nu_1\dots\nu_n}(x') = \Lambda^{\mu_1}{}_{\alpha_1}\dots\Lambda^{\mu_k}{}_{\alpha_k}T^{\alpha_1\dots\alpha_k}_{\beta_1\dots\beta_n}(x)\Lambda^{\beta_1}{}_{\nu_1}\dots\Lambda^{\beta_n}{}_{\nu_n} = 0, \tag{2.12}$$

thus preserving the form $T'^{\mu_1\dots\mu_k}_{\nu_1,\dots,\nu_n}(x') = 0$ it had for the original observer.

**Box 1. Retrieving Lorentz transformations**

It is a trivial exercise to recover the standard expression of Lorentz transformations from the invariance of the line element (2.7). For simplicity we consider a two-dimensional spacetime, equivalent to restricting to boosts along the $x$-axis so the coordinates $y' = y$ and $z' = z$ remain unchanged. Implementing the coordinate change

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = \begin{pmatrix} \Lambda^0{}_0 & \Lambda^0{}_1 \\ \Lambda^1{}_0 & \Lambda^1{}_1 \end{pmatrix}\begin{pmatrix} t \\ x \end{pmatrix}. \tag{2.13}$$

with the condition $dt'^2 - dx'^2 = dt^2 - dx^2$ implies

$$\begin{aligned} (\Lambda^1{}_0)^1 - (\Lambda^1{}_0)^2 &= 1, \\ (\Lambda^2{}_1)^1 - (\Lambda^0{}_1)^2 &= 1, \\ \Lambda^0{}_0\Lambda^0{}_1 - \Lambda^1{}_0\Lambda^1{}_1 &= 0. \end{aligned} \tag{2.14}$$

Using the properties of the hyperbolic functions, we easily see that the first two identities are solved by $\Lambda^0{}_0 = \cosh\alpha$, $\Lambda^1{}_0 = \pm\sinh\alpha$ and $\Lambda^0{}_1 = \pm\sinh\beta$, $\Lambda^1{}_1 = \cosh\beta$, for arbitrary $\alpha$ and $\beta$, with the third one requiring $\beta = \alpha$. The sought transformation is therefore parametrized as

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = \begin{pmatrix} \cosh\alpha & -\sinh\alpha \\ -\sinh\alpha & \cosh\alpha \end{pmatrix}\begin{pmatrix} t \\ x \end{pmatrix}, \tag{2.15}$$

where the parameter $\alpha$ is called the boost rapidity. A comment on the signs is in order. First, we have taken $\Lambda^0{}_0 > 0$ so the arrow of time points in the same direction for both observers (later in page 41 we will assign a Greek name to this and call these transformations orthochronous). On the other hand, as we will see right away, the parameter $\alpha$ is related to the boost velocity. Choosing a negative sign for the off-diagonal components of the matrix in (2.15) means that $\alpha > 0$ corresponds to a boost in the direction of the positive $x$-axis.

To find the standard expression of the Lorentz transformation, we notice that the hyperbolic

functions can be alternatively parametrized as

$$\cosh\alpha = \frac{1}{\sqrt{1-V^2}}, \qquad \sinh\alpha = \frac{V}{\sqrt{1-V^2}}, \tag{2.16}$$

where the relation between the boost velocity and its rapidity is given by $V = \tanh\alpha$. Plugging these expressions into (2.15), we arrive at the well-known formulae

$$t' = \frac{t - \frac{Vx}{c^2}}{\sqrt{1-\frac{V^2}{c^2}}}, \qquad x' = \frac{x - Vt}{\sqrt{1-\frac{V^2}{c^2}}}, \tag{2.17}$$

where exceptionally we have restored powers of $c$.

Whereas the Euclidean distance (2.2) tells us about how far apart in space two points lie, the spacetime geometry (2.7) contains information about the causal relations between events. Let us consider an arbitrary event that, without loss of generality, we place at the origin of our coordinate system $x_0^\mu = (0, \mathbf{0})$. The question arises as to whether some other event $x^\mu = (t, \mathbf{r})$ may either influence what happens at $x_0^\mu$ or be influenced by it. Since the speed of light is a universal velocity limit, the question is settled by checking whether it is possible for a signal propagating with velocity $v \leq 1$ to travel from $(t, \mathbf{r})$ to $(0, \mathbf{0})$, if $t < 0$, or vice-versa for positive $t$. The condition for this to happen is

$$\frac{|\mathbf{r}|}{|t|} \leq 1 \qquad \Longrightarrow \qquad t^2 - \mathbf{r}^2 \geq 0. \tag{2.18}$$

The set of events satisfying this condition defines the interior and the surface of the light-cone associated with the event at $(0, \mathbf{0})$, that we have depicted in Fig. 3 for a $(2+1)$-dimensional spacetime. Points in the causal past of the origin lie inside or on the past light-cone ($t < 0$), whereas those on or inside the future light-cone ($t > 0$) are causally reachable from $(0, \mathbf{0})$. By contrast, events outside the light-cone cannot influence or be influenced by the event at the origin, since this would require superluminal propagation. What we have said about the origin applies to any other event: every point of the spacetime is endowed with its light-cone, defining its area of casual influence.

Thus, if two events lie outside each other's light-cones, they cannot influence one another. Mathematically this is characterized by their spacetime separation satisfying $\Delta s^2 < 0$, so they are said to be *spatially* separated. Interestingly, there always exists a reference frame in which both events happen at the same $t$, i.e. they are simultaneous. This is not possible when one event is inside the other's light-cone, in which case $\Delta s^2 > 0$ and their separation is called *timelike*. Looking at (2.6) and remembering the invariant character of $\Delta s^2$ we see that there can be no frame for which $\Delta t = 0$. Nonetheless, it is always possible to find an inertial observer for which both events happen at the same point of space, i.e. $\Delta\mathbf{r} = \mathbf{0}$. In this case $\Delta s^2$ is just the (squared) time elapsed between both events, as measured by the observer who is visiting both. Notice for two events lying on each others light-cone there is no such possibility, since they can only be joined by signals propagating at the speed of light and no observer can travel at this velocity.

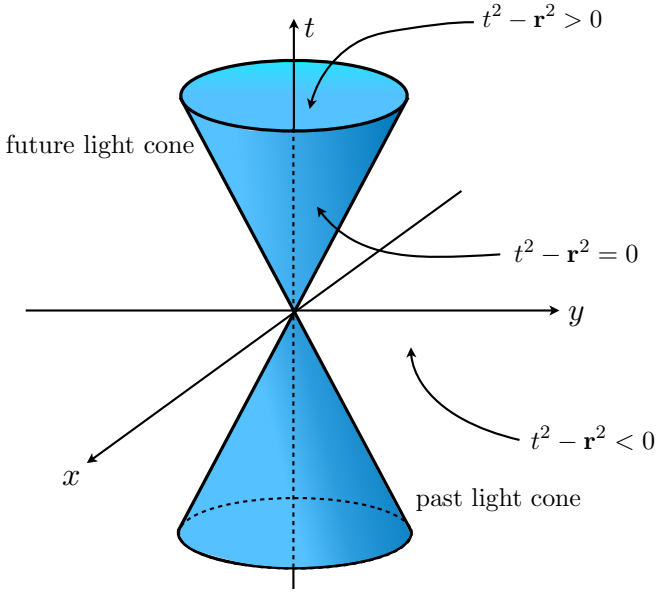
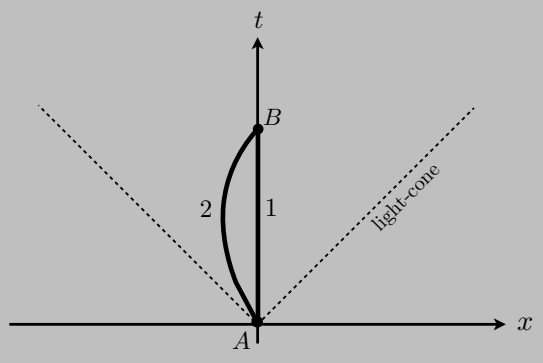



**Fig. 3:** Representation of the light cone at the origin in a $(2+1)$-dimensional spacetime.

**Box 2. There is no twin paradox**

One of the most celebrated "paradoxes" associated with special relativity is that involving two identical twins, one of which starts a round trip from Earth at very high speed while the second remains quietly behind. Relativistic time dilation implies that the clock carried by the traveling twin slows down with respect to the time set by a second clock on Earth, so at the end of the trip the returning twin looks younger than the remaining sibling. So far, so good. However, applying the same argument to the frame of reference moving with the spaceship, the conclusion seems to be the opposite: that the clock of the twin staying on Earth, that is the one moving in the reference frame of the rocket, ticks slower and after the reunion it is the Earth twin the one looking younger.

To clarify this apparent "paradox" we have to keep in mind that special relativity is about inertial observers. Thus, we are going to work with the reference frame of the twin standing on Earth, who follows the spacetime path (the worldline) indicated in the following graph as 1




The travelling twin, on the other hand, follows the worldline labelled as 2, that starts and finishes on Earth, moving back and forth along the $x$ direction. For simplicity, we restrict the movement of the

rocket to this coordinate, with the Earth located at $x = 0$.

Physical observers move along wordlines $x^\mu(\lambda)$ whose tangent at any point defines a timeline vector $\eta_{\mu\nu}\dot{x}^\mu(\lambda)\dot{x}^\nu(\lambda) > 0$. The time elapsed between two events $A$ and $B$ as measured by the clock carried by the observer (called its proper time) equals the spacetime length along the worldline $\gamma_{AB}$

$$\Delta s_{AB} = \int_{\gamma_{AB}} ds = \int_{\lambda_A}^{\lambda_B} d\lambda \sqrt{\eta_{\mu\nu}\dot{x}^\mu(\lambda)\dot{x}^\nu(\lambda)}. \tag{2.19}$$

A particularly convenient parametrization of the curve is provided by the coordinate time, $x^0 \equiv t$, so writing $x^\mu(t) = \big(t, \mathbf{R}(t)\big)$ the previous equation becomes

$$\Delta s_{AB} = \int_{t_A}^{t_B} dt' \sqrt{1 - \mathbf{v}(t')^2}, \tag{2.20}$$

with $\mathbf{v}(t) = \dot{\mathbf{R}}(t)$ the observer velocity satisfying $|\mathbf{v}(t)| < 1$.

Let us return to our twins. Both of them travel from $A$ to $B$, as shown in the graph above, but along different worldlines with different speeds. The one on Earth has $\mathbf{v} = \mathbf{0}$, so the time elapsed between the departure and arrival of the second twin is

$$\Delta s^{(1)}_{AB} = t_B - t_A. \tag{2.21}$$

For the twin on the spaceship, by contrast, we do not even need to know anything about the details of the varying speed. It is enough to notice that $0 < \sqrt{1 - \mathbf{v}(t)^2} < 1$, implying

$$\Delta s^{(2)}_{AB} < \Delta s^{(1)}_{AB}. \tag{2.22}$$

Consequently, after reunion, the traveling twin will be the younger.

A basic difference between the twins is that the one at rest is precisely the inertial observer for which the timelike separated events $A$ and $B$ happen at the same point of space. In fact, the result (2.22) reflects a property of this particular frame: its worldline represents the path of the longest proper time interpolating between two given events.

As announced, the reason why there is no paradox is because only one of the twins is an inertial observer and their descriptions cannot be simply interchanged without further ado. Seeing everything from the point of view of the spaceship leads us to give up the Minkowski metric (2.7). Indeed, by changing the coordinates

$$\begin{aligned} t' &= t, \\ \mathbf{r}' &= \mathbf{r} + \mathbf{R}(t), \end{aligned} \tag{2.23}$$

the worldlines of both twins are respectively parametrized by $x^\mu_1(t') = \big(t', -\mathbf{R}(t')\big)$ and $x^\mu_2(t') =$

$(t', \mathbf{0})$, while the spacetime metric now reads

$$ds^2 = \left[1 - \mathbf{v}(t')^2\right] dt'^2 + 2\mathbf{v}(t') \cdot d\mathbf{r}'\, dt' - d\mathbf{r}'^2, \tag{2.24}$$

which is no longer the Minkowski metric. To compute the proper time of both twins we use Eq. (2.19), replacing $\eta_{\mu\nu}$ by the line element (2.24). We then find

$$\begin{aligned} \Delta s^{(1)}_{AB} &= \int_{t'_A}^{t'_B} dt' \sqrt{1 - \mathbf{v}(t')^2 + 2\mathbf{v}(t')^2 - \mathbf{v}(t')^2} = t_B - t_A, \\ \Delta s^{(2)}_{AB} &= \int_{t'_A}^{t'_B} dt' \sqrt{1 - \mathbf{v}(t')^2} < \Delta s^{(1)}_{AB}, \end{aligned} \tag{2.25}$$

which reproduce the results obtained above. The conclusion is that, if properly analyzed, the descriptions from the points of view of both twins are absolutely consistent and no paradox arises.

As time and space coordinates combine to label a point (event) in the four-dimensional Minkowski spacetime, so do energy and momentum build up an energy–momentum four-vector $p^\mu = (E, \mathbf{p})$. For a particle of mass $m$ moving along an affinely paramerized worldline $x^\mu(s)$, the four-momentum is defined by

$$p^\mu(s) \equiv m\dot{x}^\mu(s) = \left( \frac{m}{\sqrt{1 - \mathbf{v}^2}}, \frac{m\mathbf{v}}{\sqrt{1 - \mathbf{v}^2}} \right), \tag{2.26}$$

with $\mathbf{v}$ the particle's velocity. A first thing to be noticed here is that the particle's energy is nonzero even when its velocity vanishes. Restoring powers of $c$

$$E \longrightarrow \frac{E}{c}, \qquad m \longrightarrow mc \qquad \mathbf{v} \longrightarrow \frac{\mathbf{v}}{c}, \tag{2.27}$$

we get the famous equation $E_{\text{rest}} = mc^2$. On the other hand, the particle's energy diverges as $|\mathbf{v}| \to c$. This shows that the speed of light is a physical limiting velocity for any massive particle, since reaching $|\mathbf{v}| = c$ would require pumping an infinite amount of energy into the system. The transformation of energy and momentum among inertial observers is fixed by $p^\mu$ being a four-vector, whose change under a Lorentz transformation $\Lambda^\mu{}_\nu$ is given by $p'^\mu = \Lambda^\mu{}_\nu p^\nu$. Considering a boost along the $x$ direction with velocity $V$ and using the expressions obtained in Box 1 in pages 10-11, we have

$$E' = \frac{E - Vp_x}{\sqrt{1 - V^2}}, \qquad p'_x = \frac{p_x - VE}{\sqrt{1 - V^2}}, \tag{2.28}$$

together with $p'_y = p_y$ and $p'_z = p_z$.

Equation (2.26) also implies the mass-shell condition[2]

$$E^2 - \mathbf{p}^2 = m^2. \tag{2.29}$$

[2] In covariant terms, the mass-shell condition reads $p_\mu p^\mu = m^2$ and follows from (2.26), remembering that the particle's worldline is affinely parametrized, $\eta_{\mu\nu}\dot{x}^\mu(s)\dot{x}^\nu(s) = 1$.

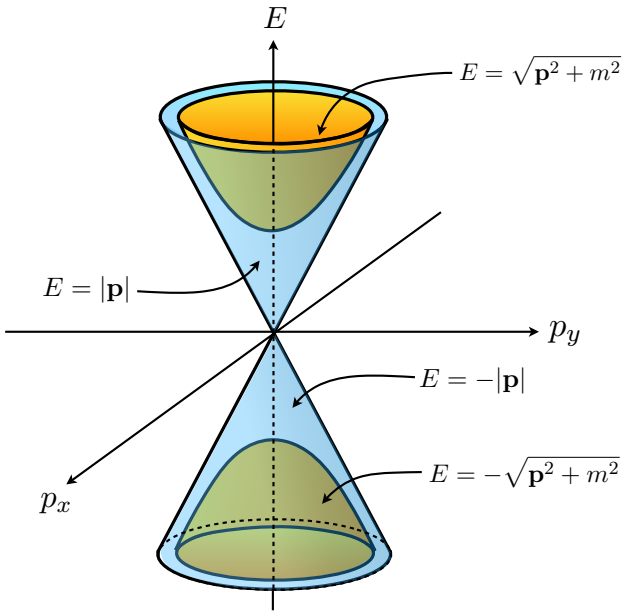



**Fig. 4:** Energy–momentum hyperboloid for a particle of mass $m \neq 0$ (orange). The energy-momentum vector of a massless particle lies on the blue cone.

In the four-dimensional energy–momentum space spanned by $E$ and $\mathbf{p}$, the particle's four-momentum $p^\mu$ lies on the two-sheeted hyperboloid $E = \pm\sqrt{\mathbf{p}^2 + m^2}$, with the two signs corresponding to the upper and lower sheet. Interestingly, the mass-shell condition has a smooth limit as $m \to 0$, where the hyperboloid degenerates into the cone $E^2 = \mathbf{p}^2$, to which all massive hyperboloids asymptote for large spatial momentum, $|\mathbf{p}| \gg m$ (see Fig. 4). Unlike Newtonian mechanics, special relativity admits the existence of zero-mass particles whose four-momenta have the form

$$p^\mu = (|\mathbf{p}|, \mathbf{p}), \tag{2.30}$$

where we have chosen the positive energy solution. In terms of its energy and momentum, the velocity of a massive particle is given by [cf. (2.26) and (2.29)]

$$\mathbf{v} = \frac{\mathbf{p}}{\sqrt{\mathbf{p}^2 + m^2}}, \tag{2.31}$$

which as $m \to 0$ gives $|\mathbf{v}| = 1$. Thus, massless particles necessarily propagate at the speed of light.

## 2.2 Relativity and quantum mechanics

So far, our analysis has left out quantum effects. Special relativity can be combined with quantum mechanics to formulate relativistic wave equations plagued with trouble. An immediate problem arises from the energy hyperboloid depicted in Fig. 4. The existence of the lower sheet implies that the system of a relativistic quantum particle coupled to an electromagnetic field has no ground state, since the particle has infinitely many available states with arbitrary negative energy to which it could decay by radiating energy. This fundamental instability of the system is impossible to solve in the context of the Klein–Gordon wave equation, while in the Dirac equation it can be avoided by "filling" all states in the

lower sheet of the hyperboloid (the Dirac sea). The Pauli exclusion principle now prevents electrons from occupying negative energy states, and the system is stable.

The Dirac sea notwithstanding, the interpretation of the Dirac equation as a single-particle relativistic wave equation is problematic, leading to puzzling results such as the Klein paradox [14, 47]. In fact, all the difficulties we run into when trying to marry quantum mechanics with special relativity stem from insisting in a single-particle description, as can be seen from a simple heuristic argument. As we know, Heisenberg's uncertainty principle correlates quantum fluctuations in the position and momentum of a particle

$$\Delta x \Delta p_x \geq \frac{\hbar}{2}. \tag{2.32}$$

Looking at physics at small distances requires taming spatial fluctuations below the scale of interest, which in turn leads to large fluctuations in the particle's momentum. When the latter reaches the scale $\Delta p_x \sim mc$, the corresponding energy fluctuations $\Delta E \sim mc^2$ are large enough to allow the creation of particles out of the vacuum and the single-particle description breaks down. Equivalently, localizing a particle below its Compton wavelength,

$$\Delta x \leq \frac{\hbar}{2mc}, \tag{2.33}$$

leads to a quantum state characterized by an indefinite number of them. Unlike what happens in non-relativistic many body physics, in the quantum-relativistic domain particle number is not conserved and creation-annihilation of particles is a central ingredient of the theory. Thus, the single-particle description inherent to the relativistic wave equation is fundamentally wrong, as indicated by the paradoxes and inconsistencies it leads to.

**Box 3. Antiparticles and causality**

One of the consequences of the Klein paradox alluded to above is the impossibility of a consistent formulation of relativistic quantum mechanics without the inclusion of antiparticles. We can reach the same conclusion by showing that antiparticles are the unavoidable ingredient to preserve causality in a relativistic quantum theory. To do so, let us consider a relativistic particle of mass $m$ that at $t = 0$ is detected at the origin. Its quantum-mechanical propagator is given by

$$G(\tau, \mathbf{r}) \equiv \langle \mathbf{r} | e^{-i\tau\sqrt{\mathbf{p}^2+m^2}} | \mathbf{0} \rangle = e^{-i\tau\sqrt{-\boldsymbol{\nabla}^2+m^2}} \delta^{(3)}(\mathbf{r}). \tag{2.34}$$

Physically, this quantity gives the probability amplitude of the particle being detected at a later time $t = \tau$ at some location $\mathbf{r}$. To explicitly evaluate the propagator, we Fourier transform the Dirac delta function and compute the resulting integral in terms of a modified Bessel function of the second kind

$$G(\tau, \mathbf{r}) = \int \frac{d^3k}{(2\pi)^3} e^{-i\tau\sqrt{\mathbf{k}^2+m^2}+i\mathbf{k}\cdot\mathbf{r}}$$

$$= \frac{1}{2\pi^2|\mathbf{r}|}\int_0^\infty kdk\,\sin(k|\mathbf{r}|)e^{-i\tau\sqrt{k^2+m^2}} \tag{2.35}$$
$$= -\frac{i}{2\pi^2}\frac{m^2 t}{\tau^2-\mathbf{r}^2}K_2\left(im\sqrt{\tau^2-\mathbf{r}^2}\right),$$

where, to write the last identity, we regularized the momentum integral by analytical continuation $\tau \to \tau - i\epsilon$. Naively, one would expect this propagator to vanish outside the light cone, $\tau^2 - \mathbf{r}^2 < 0$, since otherwise the particle would have a nonvanishing probability of being detected at points spacelike separated from the origin, its location at $t = 0$. Were this to happen, it would imply a violation of causality.

Despite expectations, the modified Bessel function in (2.35) is nonzero for both real and imaginary values of the argument and the propagator spills out of the light-cone despite being derived from a relativistic Hamiltonian. The key point to understand what is going on is that when $\mathbf{r}$ lies outside the light-cone at the origin there are frames in which the detection of the particle at the position $\mathbf{r}$ *precedes* its detection at the origin. In computing the propagator we should take this into account and consider the superposition of both processes outside and inside the light-cone

$$G(\tau,\mathbf{r}) = \begin{cases} \langle \mathbf{r}|e^{-i\tau\sqrt{\mathbf{p}+m^2}}|\mathbf{0}\rangle \quad \text{when} \quad \tau^2-\mathbf{r}^2>0 \\ \langle \mathbf{r}|e^{-i\tau\sqrt{\mathbf{p}+m^2}}|\mathbf{0}\rangle + \langle \mathbf{0}|e^{i\tau\sqrt{\mathbf{p}+m^2}}|\mathbf{r}\rangle \quad \text{when} \quad \tau^2-\mathbf{r}^2<0 \end{cases} \tag{2.36}$$

Now, from the explicit expression (2.35) we can check that $\langle \mathbf{r}|e^{-i\tau\sqrt{\mathbf{p}+m^2}}|\mathbf{0}\rangle$ is purely imaginary when $\tau^2 - \mathbf{r}^2 < 0$. Since, on the other hand,

$$\langle \mathbf{r}|e^{-i\tau\sqrt{\mathbf{p}+m^2}}|\mathbf{0}\rangle + \langle \mathbf{0}|e^{i\tau\sqrt{\mathbf{p}+m^2}}|\mathbf{r}\rangle = 2\mathrm{Re}\,\langle \mathbf{r}|e^{-i\tau\sqrt{\mathbf{p}+m^2}}|\mathbf{0}\rangle, \tag{2.37}$$

we conclude that

$$G(\mathbf{r},\tau) = -\frac{i}{2\pi^2}\frac{m^2 t}{\tau^2-\mathbf{r}^2}K_2\left(im\sqrt{\tau^2-\mathbf{r}^2}\right)\theta(\tau^2-\mathbf{r}^2), \tag{2.38}$$

and causality is consequently restored.

There exists an interesting interpretation of this cancellation mechanism due to Ernst Stueckelberg [48] and Richard Feynman [49, 50]. Our propagator can be seen as the wave function of the particle of interest, $\psi(\tau,\mathbf{r}) \equiv G(\tau,\mathbf{r})$, satisfying the boundary condition $\psi(0,\mathbf{r}) = \delta^{(3)}(\mathbf{r})$. We found that outside the light-cone there is a superposition of two processes: one in which the particle is traveling from the origin to $\mathbf{r}$ forward in time, and a second described by the wave function

$$\psi(\tau,\mathbf{r})_\Downarrow \equiv \langle \mathbf{0}|e^{i\tau\sqrt{\mathbf{p}^2+m^2}}|\mathbf{r}\rangle = \langle \mathbf{r}|e^{-i\tau\sqrt{\mathbf{p}^2+m^2}}|\mathbf{0}\rangle^* \equiv \psi(\tau,\mathbf{r})^*_\Uparrow, \tag{2.39}$$

where the particle moves backwards in time from $\mathbf{r}$ to the origin. Furthermore, writing

$$\psi(\tau,\mathbf{r})_\Downarrow = \int \frac{d^3k}{(2\pi)^3}e^{i\tau\sqrt{\mathbf{k}^2+m^2}-i\mathbf{k}\cdot\mathbf{r}} = \int \frac{d^3k}{(2\pi)^3}e^{-i\tau(-\sqrt{\mathbf{k}^2+m^2})+i(-\mathbf{k})\cdot\mathbf{r}} \tag{2.40}$$

and comparing with the first line in Eq. (2.35), we reinterpret $\psi(\tau,\mathbf{r})_{\Downarrow}$ as describing a state of mass $m$ and momentum $-\mathbf{k}$, lying in the lower sheet of the energy hyperboloid, and propagating forward in time. This represents a hole in the Dirac sea, i.e. an *antiparticle* of momentum $\mathbf{k}$. Moreover, from (2.39) we see that if our particle has charge $q$ with respect to some global U(1) symmetry, the antiparticle necessarily transforms with the opposite charge

$$\psi(\tau,\mathbf{r})_{\Uparrow} \to e^{iq\theta}\psi(\tau,\mathbf{r})_{\Uparrow} \qquad \Longrightarrow \qquad \psi(\tau,\mathbf{r})_{\Downarrow} \to e^{-iq\theta}\psi(\tau,\mathbf{r})_{\Downarrow}. \tag{2.41}$$

Antiparticles are therefore a necessary ingredient in a relativist theory of quantum processes if we want to avoid superluminal effects. They automatically imply the possibility of creation/annihilation of particle–antiparticle pairs, turning what was intended as single-particle relativistic quantum mechanics into a multiparticle theory where the number of particles is not even well defined.

A fundamental consequence of the causal structure of spacetime is that measurement of observables in regions that are spacelike separated cannot interfere with each other. In quantum theory these measurements are implemented by local operators $\mathcal{O}(x)$ smeared over the spacetime region $R$ where the measurement takes place

$$\mathcal{O}(R) \equiv \int d^4x\, \mathcal{O}(x) f_R(x), \tag{2.42}$$

where

$$f_R(x) = \begin{cases} 1 & \text{if } x \in R \\ 0 & \text{if } x \notin R \end{cases} \tag{2.43}$$

is the characteristic function associated with $R$. In mathematical terms, the noninterference of the measurements carried out in spacelike separated regions $R_1$ and $R_2$ like those shown in Fig. 5 is expressed by the vanishing of the commutator of the associated operators

$$[\mathcal{O}(R_1), \mathcal{O}(R_2)] = 0 \qquad \text{if } R_1 \text{ and } R_2 \text{ are spacelike separated,} \tag{2.44}$$

or equivalently

$$[\mathcal{O}(x), \mathcal{O}(y)] = 0 \qquad \text{if } (x-y)^2 < 0. \tag{2.45}$$

This states the *principle of microcausality*, a profound form of locality that has to be imposed on constructing any admissible QFT. To date no consistent theory has been formulated violating this principle. This is why all theories to be encountered later in these lecture will be local quantum field theories (LQFTs) in the sense of Eq. (2.44).

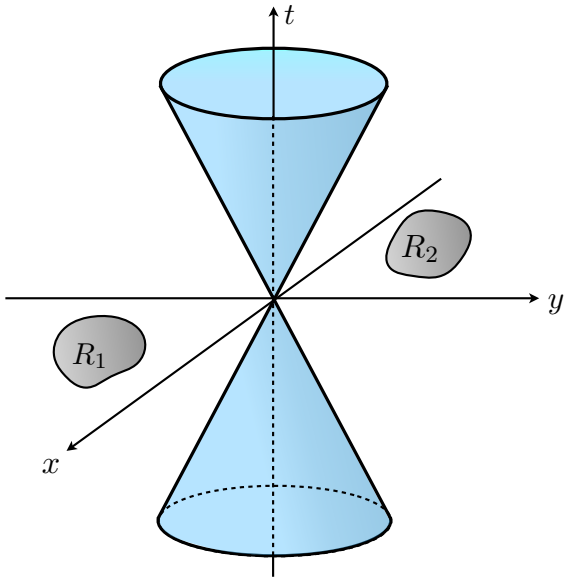



**Fig. 5:** The two spacelike-separated regions $R_1$ and $R_2$ cannot causally influence one another.

## 3 The importance of classical field theory

Maxwell's electromagnetism is arguably the mother of all classical field theories. Despite its apparent simplicity, the theory contains a number of symmetries and structures that underlie many other developments in QFT. This is the reason why it is worthwhile to spend some time extracting some lessons from classical electromagnetism that we will find useful later in our study of the SM and other theories.

### 3.1 The symmetries of Maxwell's theory

Using Heaviside units, and keeping $c = 1$ all the way, the Maxwell's equations take the form

$$\begin{aligned} \boldsymbol{\nabla} \cdot \mathbf{E} &= \rho_e, \\ \boldsymbol{\nabla} \cdot \mathbf{B} &= \rho_m, \\ \boldsymbol{\nabla} \times \mathbf{E} &= -\mathbf{j}_m - \frac{\partial \mathbf{B}}{\partial t}, \\ \boldsymbol{\nabla} \times \mathbf{B} &= \mathbf{j}_e + \frac{\partial \mathbf{E}}{\partial t}. \end{aligned} \tag{3.1}$$

Here we have introduced a color code signaling various layers of generality. Setting to zero all terms in blue and red we get the vacuum Maxwell's equations governing the evolution of electromagnetic fields in the absence of any kind of matter. If we keep the terms in blue but remove those in red, the resulting expressions describe the coupling of electric and magnetic fields to electrically charged matter, where $\rho_e$ and $\mathbf{j}_e$, respectively, represent the electric charge density and current. These are the Maxwell's equations that can be found in most textbooks on classical electrodynamics (see, for example, Ref. [51]).

Let us postpone a little bit the discussion of the terms in red and concentrate on the second and

third equations

$$\boldsymbol{\nabla}\cdot\mathbf{B}=0,$$
$$\boldsymbol{\nabla}\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t}. \tag{3.2}$$

They imply that the electric and magnetic fields can be written in terms of a scalar and a vector potential $(\phi,\mathbf{A})$ as

$$\mathbf{B}=\boldsymbol{\nabla}\times\mathbf{A},$$
$$\mathbf{E}=-\boldsymbol{\nabla}\phi-\frac{\partial\mathbf{A}}{\partial t}. \tag{3.3}$$

These potentials, however, are not uniquely defined. The electric and magnetic fields remain unchanged if we replace

$$\phi\longrightarrow\phi+\frac{\partial\epsilon}{\partial t},$$
$$\mathbf{A}\longrightarrow\mathbf{A}-\boldsymbol{\nabla}\epsilon, \tag{3.4}$$

with $\epsilon(t,\mathbf{r})$ an arbitrary well-behaved function. This *gauge invariance* is probably the most important of those structures of the electromagnetic theory that we said were of radical importance for QFT at large. Although at a classical level it might seem a mere technicality, it has profound implications for the quantum theory and is the cornerstone of the whole SM. We explore its significance in some detail in the following. For computational purposes, it is convenient sometimes to (partially) fix the gauge freedom by imposing certain conditions on $\phi$ and $\mathbf{A}$. Two popular choices in classical electromagnetism are the Coulomb gauge $\boldsymbol{\nabla}\cdot\mathbf{A}=0$ and the temporal (also called Weyl) gauge $\phi=0$. These conditions still leave a residual invariance, generated in the first case by harmonic functions $\nabla^2\epsilon(t,\mathbf{r})=0$ and by time independent functions $\epsilon(\mathbf{r})$ in the second. A covariant alternative is the Lorentz gauge

$$\boldsymbol{\nabla}\cdot\mathbf{A}+\frac{\partial\phi}{\partial t}=0, \tag{3.5}$$

preserved by gauge functions satisfying the wave equation, $\Box\epsilon(t,\mathbf{r})=0$.

Gauge invariance introduces a *redundancy* in the description in terms of the electromagnetic potentials that however cannot be reflected in physically measurable quantities such as the electric and magnetic fields. These are not the only gauge invariant quantities that can be constructed in terms of $\phi$ and $\mathbf{A}$. There is also the Wilson loop, defined by

$$U(\gamma)\equiv\exp\left(-ie\oint_\gamma d\mathbf{r}\cdot\mathbf{A}\right), \tag{3.6}$$

where $\gamma$ is a closed path in space and $e$ the electric charge. Implementing a gauge transformation on the

vector potential and using the Stokes theorem, we see that it is indeed gauge invariant

$$\exp\left(-ie\oint_\gamma d\mathbf{r}\cdot\mathbf{A}\right)\longrightarrow\exp\left(-ie\oint_\gamma d\mathbf{r}\cdot\mathbf{A}+ie\oint_\gamma d\mathbf{r}\cdot\boldsymbol{\nabla}\epsilon\right)=\exp\left(-ie\oint_\gamma d\mathbf{r}\cdot\mathbf{A}\right), \tag{3.7}$$

after taking into account that $\gamma$ is closed. Whereas $\mathbf{E}$ and $\mathbf{B}$ are *local* observables depending on the spacetime point where they are measured, the Wilson loop is *nonlocal* since it "explores" the whole region enclosed by $\gamma$.

It is enlightening to study the consequences of gauge transformations for the dynamics of a quantum particle coupled to an electromagnetic field. In quantum mechanics the prescription of minimal coupling of a particle with electric charge $e$ to the electromagnetic field

$$\mathbf{p}\longrightarrow\mathbf{p}-e\mathbf{A},\qquad H\longrightarrow H+e\phi, \tag{3.8}$$

introduces an explicit dependence of the Schrödinger equation on the electromagnetic potentials

$$i\frac{\partial\psi}{\partial t}=\left[-\frac{1}{2m}\left(\boldsymbol{\nabla}-ie\mathbf{A}\right)^2+e\phi\right]\psi. \tag{3.9}$$

To preserve the gauge invariance of this equation, the transformations (3.7) have to be supplemented by a phase shift of the wave function

$$\psi(t,\mathbf{r})\longrightarrow e^{-ie\epsilon(t,\mathbf{r})}\psi(t,\mathbf{r}), \tag{3.10}$$

which does not affect the probability density $|\psi(t,\mathbf{r})|^2$. This shows that the gauge transformations in electromagnetism belong to the Abelian group U(1) of complex rotations, parametrized by elements

$$U=e^{-ie\epsilon(t,\mathbf{r})}, \tag{3.11}$$

in terms of which Eq. (3.4) reads

$$\begin{aligned}\phi&\longrightarrow\phi+\frac{i}{e}U^{-1}\frac{\partial}{\partial t}U,\\ \mathbf{A}&\longrightarrow\mathbf{A}-\frac{i}{e}U^{-1}\boldsymbol{\nabla}U.\end{aligned} \tag{3.12}$$

**Box 4. Wilson loops and quantum interference**

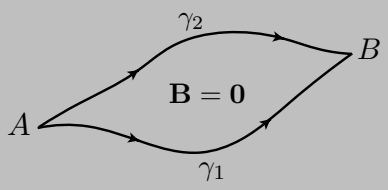
At the classical level we can live with just local observables, like the electric and magnetic fields, but not anymore when we introduce quantum effects. In this case the phase transformation of the wave function may give rise to observable interference phenomena. As we will see now, these are measured by a Wilson loop $U(\gamma)$.

We work for simplicity in the temporal gauge $\phi=0$. The action of a classical charged particle

propagating in the background of an electromagnetic potential $\mathbf{A}(t,\mathbf{r})$ is given by

$$S = \frac{1}{2}\int dt\, m\dot{\mathbf{r}}^2 - e\int_\gamma d\mathbf{r}\cdot\mathbf{A}, \tag{3.13}$$

where $\gamma$ is the particle trajectory and $e$ is the electron charge. An interesting property of the second term is that its value does not change if we smoothly deform the path $\gamma$ across any region where the magnetic field vanishes. Let us consider two paths $\gamma_1$ and $\gamma_2$ joining two points $A$ and $B$ as shown here




Computing the difference between the contributions of both paths, we find a Wilson loop

$$\int_{\gamma_1} d\mathbf{r}\cdot\mathbf{A} - \int_{\gamma_2} d\mathbf{r}\cdot\mathbf{A} = \oint_{\gamma_2^{-1}\gamma_1} d\mathbf{r}\cdot\mathbf{A} = 0, \tag{3.14}$$

where $\gamma_2^{-1}\gamma_1$ represents the closed path from $A$ to $B$ following $\gamma_1$ and back to $A$ along $\gamma_2$. To see why this term is zero, let us denote by $S$ any surface bounded by $\gamma_2^{-1}\gamma_1$. Applying the Stokes theorem, we have

$$\oint_{\gamma_2^{-1}\gamma_1} d\mathbf{r}\cdot\mathbf{A} = \int_S d\mathbf{S}\cdot(\boldsymbol{\nabla}\times\mathbf{A}) = 0, \tag{3.15}$$

since we assumed that $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A} = 0$ in the integration domain.

This topological property of the interaction term in (3.13) has an important consequence in quantum mechanics, as pointed out by Yakir Aharonov and David Bohm [52]. Let us look at a double slit experiment performed with electrons in which behind the slitted screen we place a vertical solenoid confining a constant magnetic field $\mathbf{B}$ (see Fig. 6 in page 23). The amplitude for an electron emitted from $A$ at $t=0$ to be detected at a point $P$ of the detection screen at $t=\tau$ can be computed as a coherent quantum superposition of all possible classical trajectories, expressed by the Feynman path integral

$$G(\tau;\mathbf{r}_A,\mathbf{r}_P) = \mathcal{N}\int_{\substack{\mathbf{r}(0)=\mathbf{r}_A\\ \mathbf{r}(\tau)=\mathbf{r}_P}} \mathscr{D}\mathbf{r}\exp\left(\frac{i}{2}\int_0^\tau dt\, m\dot{\mathbf{r}}^2 - ie\int_\gamma d\mathbf{r}\cdot\mathbf{A}\right), \tag{3.16}$$

with $\mathcal{N}$ a global normalization. The modulus squared of $G(\tau;\mathbf{r}_A,\mathbf{r}_P)$ gives the probability of the electron being detected at the point $P$ at time $\tau$.

Recall that the magnetic field outside the solenoid is equal to zero and we can thus apply the topological property (3.14) to conclude that the second term in the exponential of (3.16) takes the

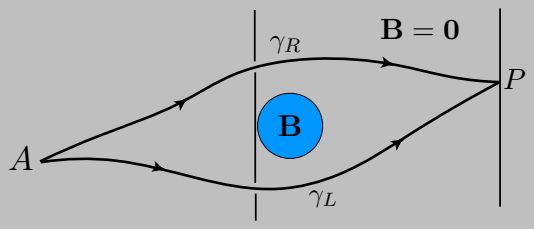
same value for *all* trajectories $\gamma_L$ passing through the left slit, and the same for *all* paths $\gamma_R$ going through the right one. The total propagator can then be written as

$$\begin{aligned} G(\tau;\mathbf{r}_A,\mathbf{r}_P) &= e^{-ie\int_{\gamma_R} d\mathbf{r}\cdot\mathbf{A}} G_R(\tau;\mathbf{r}_A,\mathbf{r}_P)_0 + e^{-ie\int_{\gamma_L} d\mathbf{r}\cdot\mathbf{A}} G_L(\tau;\mathbf{r}_A,\mathbf{r}_P)_0 \\ &= e^{-ie\int_{\gamma_R} d\mathbf{r}\cdot\mathbf{A}} \left[ G_R(\tau;\mathbf{r}_A,\mathbf{r}_P)_0 + e^{-ie\int_{\gamma_L} d\mathbf{r}\cdot\mathbf{A} + ie\int_{\gamma_R} d\mathbf{r}\cdot\mathbf{A}} G_L(\tau;\mathbf{r}_A,\mathbf{r}_P)_0 \right], \end{aligned} \tag{3.17}$$

where $G_{R,L}(\tau;\mathbf{r}_A,\mathbf{r}_P)_0$ are the propagators for the electrons going through the right (resp. left) slit in the absence of the solenoid. Now, although the global phase disappears when computing the probability amplitude, the relative phase inside the brackets of the second line of (3.17) contributes to the interference pattern to be observed on the detection screen. Using the same arguments leading to the result (3.14), we express this phase as the Wilson loop associated with the closed path $\gamma_R^{-1}\gamma_L$

$$\exp\left(-ie\int_{\gamma_L} d\mathbf{r}\cdot\mathbf{A} + ie\int_{\gamma_R} d\mathbf{r}\cdot\mathbf{A}\right) = \exp\left(-ie\oint_{\gamma_R^{-1}\gamma_L} d\mathbf{r}\cdot\mathbf{A}\right) \equiv U(\gamma_R^{-1}\gamma_L). \tag{3.18}$$

It is important to keep in mind that $\gamma_R^{-1}\gamma_L$ represents *any* closed path going through both slits and enclosing the solenoid. To evaluate this Wilson loop let us take a bird's-eye view of the Aharonov–Bohm experimental setup in Fig. 6, that we schematically represent as:




Should we apply the Stokes theorem to the calculation of $U(\gamma_R^{-1}\gamma_L)$ as we did in Eq. (3.15), the resulting integral would not be zero anymore. As we see, the surface $S$ enclosed by the loop is now pierced by the solenoid, and the magnetic field $\mathbf{B}=\boldsymbol{\nabla}\times\mathbf{A}$ is not zero everywhere. Instead

$$\oint_{\gamma_R^{-1}\gamma_L} d\mathbf{r}\cdot\mathbf{A} = \int_S d\mathbf{S}\cdot\mathbf{B} = \Phi, \tag{3.19}$$

where $\Phi$ is the magnetic flux inside the solenoid, and we have

$$U(\gamma_R^{-1}\gamma_L) = e^{-ie\Phi} \neq 1. \tag{3.20}$$

Hence, the presence of the solenoid modifies the interference pattern on the screen, even if the electrons never enter the region where the magnetic field is nonzero. The reason is that even if $\mathbf{B}=\mathbf{0}$ outside, $\mathbf{A}$ is not. Although no force is applied to them, the electrons interact with the vector potential whose global structure, codified in the nonlocal gauge-invariant quantity $U(\gamma_R^{-1}\gamma_L)$, contains information about the confined magnetic field.

Going back to the Maxwell's equations (3.1), we notice that the vacuum equations (with all blue and red terms removed) exhibit an interesting symmetry. Combining the electric and magnetic fields into

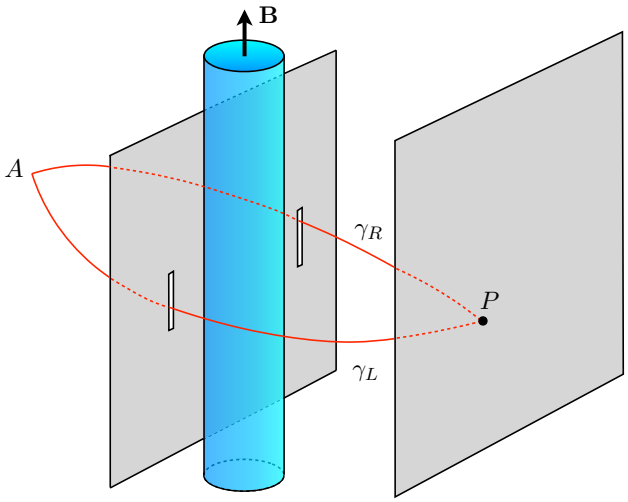



**Fig. 6:** Experimental setup to exhibit the Aharonov–Bohm effect explained in Box 4.

a single complex field $\mathbf{E}+i\mathbf{B}$, the four equations can be summarized as

$$\begin{aligned}\boldsymbol{\nabla}\cdot(\mathbf{E}+i\mathbf{B}) &= 0,\\ \boldsymbol{\nabla}\times(\mathbf{E}+i\mathbf{B}) - i\frac{\partial}{\partial t}(\mathbf{E}+i\mathbf{B}) &= 0.\end{aligned} \tag{3.21}$$

Both identities remain invariant under the transformation

$$\mathbf{E}+i\mathbf{B} \longrightarrow e^{i\theta}\big(\mathbf{E}+i\mathbf{B}\big), \tag{3.22}$$

with $\theta$ a real global angle. To be more specific, splitting the previous equation into its real and imaginary parts, we find

$$\begin{aligned}\mathbf{E} &\longrightarrow \mathbf{E}\cos\theta - \mathbf{B}\sin\theta,\\ \mathbf{B} &\longrightarrow \mathbf{E}\sin\theta + \mathbf{B}\cos\theta,\end{aligned} \tag{3.23}$$

which for $\theta=\frac{\pi}{2}$ interchanges electric and magnetic fields $(\mathbf{E},\mathbf{B})\to(-\mathbf{B},\mathbf{E})$.

This electric–magnetic duality of the vacuum equations is however broken by the source terms in the "textbook" Maxwell's equations [i.e., Eq. (3.1) without the terms in red]. The identities (3.21) are then recast as

$$\begin{aligned}\boldsymbol{\nabla}\cdot(\mathbf{E}+i\mathbf{B}) &= \rho_e,\\ \boldsymbol{\nabla}\times(\mathbf{E}+i\mathbf{B}) - i\frac{\partial}{\partial t}(\mathbf{E}+i\mathbf{B}) &= i\mathbf{j}_e.\end{aligned} \tag{3.24}$$

Since $\rho_e$ and $\mathbf{j}_e$ are both real quantities, the only transformations preserving these equations are the trivial ones which either leave invariant the electric and magnetic fields or reverse their signs (corresponding respectively to $\theta=0,\pi$), the latter one also requiring the reversal of the sign of $\rho_e$ and $\mathbf{j}_e$. Physically this

makes sense, since as far as we know there is a fundamental asymmetry in nature between electric and magnetic fields. While the first are sourced by point charges (electric monopoles) at which field lines either begin or end, magnetic fields are associated with the motion of electric charges and their field lines always close on themselves. Restoring electric-magnetic duality in the Maxwell's equations requires treating the sources of both fields symmetrically, which means introducing magnetic charge density and current. These are the terms in red in Eq. (3.1), that we rewrite now as

$$\begin{aligned} \boldsymbol{\nabla}\cdot(\mathbf{E}+i\mathbf{B}) &= \rho_e + i\rho_m, \\ \boldsymbol{\nabla}\times(\mathbf{E}+i\mathbf{B}) - i\frac{\partial}{\partial t}(\mathbf{E}+i\mathbf{B}) &= i\big(\mathbf{j}_e + i\mathbf{j}_m\big). \end{aligned} \tag{3.25}$$

These equations remain invariant under electric–magnetic duality (3.22) when supplemented by a corresponding rotation of the sources

$$\begin{aligned} \rho_e + i\rho_m &\longrightarrow e^{i\theta}\big(\rho_e + i\rho_m\big), \\ \mathbf{j}_e + i\mathbf{j}_m &\longrightarrow e^{i\theta}\big(\mathbf{j}_e + i\mathbf{j}_m\big). \end{aligned} \tag{3.26}$$

For $\theta = \frac{\pi}{2}$ the interchange of electric and magnetic fields is accompanied by a swap of the electric and magnetic sources, $(\rho_e, \mathbf{j}_e) \to (-\rho_m, -\mathbf{j}_m)$ and $(\rho_m, \mathbf{j}_m) \to (\rho_e, \mathbf{j}_e)$.

The consequences of having particles with magnetic charge were first explored by Dirac in Ref. [53]. Let us assume the existence of a point magnetic source that for simplicity we locate at the origin, $\rho_m = g\delta^{(3)}(\mathbf{r})$. The second equation in (3.1) leads to

$$\boldsymbol{\nabla}\cdot\mathbf{B} = g\delta^{(3)}(\mathbf{r}) \qquad \Longrightarrow \qquad \mathbf{B}(\mathbf{r}) = \frac{1}{4\pi}\frac{g}{r^2}\mathbf{u}_r, \tag{3.27}$$

which would be a magnetic analog of the Coulomb field. An important point to consider is that, despite the source's presence, the magnetic field's divergence still vanishes everywhere except at the monopole's position. As a consequence, away from this point we can still write $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A}$, which is solved by

$$\mathbf{A}(\mathbf{r}) = \frac{1}{4\pi}\frac{g}{r}\tan\left(\frac{\theta}{2}\right)\mathbf{u}_\varphi, \tag{3.28}$$

where we are using spherical coordinates $(r, \varphi, \theta)$. This vector potential is singular not only at the monopole location at $\mathbf{r} = 0$, but all along the line $\theta = \pi$ as well. The existence of this singular Dirac string should not be a surprise. Were $\mathbf{A}(\mathbf{r})$ be regular everywhere outside the origin, we could apply the Stokes theorem to the integral giving the magnetic flux across a closed surface $\mathcal{S}$ enclosing the monopole, to find

$$\int_{\mathcal{S}} d\mathbf{S}\cdot\mathbf{B} = \int_{\mathcal{S}} d\mathbf{S}\cdot(\boldsymbol{\nabla}\times\mathbf{A}) = \oint_{\partial\mathcal{S}} d\boldsymbol{\ell}\cdot\mathbf{A} = 0, \tag{3.29}$$

since $\partial\mathcal{S} = \emptyset$. This would contradict the calculation of the same integral applying Gauss' theorem

$$\int_{\mathcal{S}} d\mathbf{S}\cdot\mathbf{B} = \int_{\mathcal{B}_3} \boldsymbol{\nabla}\cdot\mathbf{B} = g \neq 0, \tag{3.30}$$

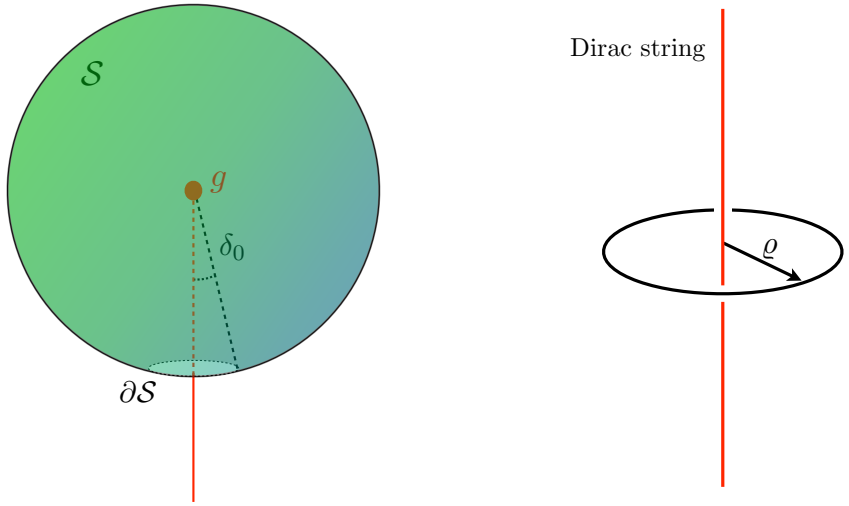



**Fig. 7:** Left: Section of a sphere around a Dirac magnetic monopole with charge $g$, resulting from cutting out a region around the south pole. Its boundary $\partial\mathcal{S}$ surrounds the singular Dirac string located along $\theta = \pi$ (in red). Right: Closed path surrounding the Dirac string.

where $\mathcal{B}_3$ denotes the three-dimensional region bounded by $\mathcal{S}$ and containing the monopole. Notice that this second calculation is free of trouble, since the magnetic field (3.27) is regular everywhere on $\mathcal{S}$. The catch, of course, is that the vector potential is singular at $\theta = \pi$ and the surface $\mathcal{S}$ in (3.29) cannot be closed. As shown on the left of Fig. 7, its boundary is a circle surrounding the singularity and the integral gives a nonzero result

$$\oint_{\partial\mathcal{S}} d\boldsymbol{\ell}\cdot\mathbf{A} = \frac{1}{2}g\sin\delta_0\tan\left(\frac{\delta_0}{2}\right) \xrightarrow{\delta_0\to 0} g, \tag{3.31}$$

where the last limit corresponds to shrinking the boundary to a point, reproducing the result of Eq. (3.30).

Even if mathematically unavoidable, the existence of a singularity is always a source of concern in physics. A way to restore our peace of mind in this case might be to make the Dirac string an artefact that somehow is rendered unobservable. One may think that a way to accomplish this is to apply a gauge transformation, since the vector potential is not uniquely defined. This, however, does not eliminate the Dirac string, just changes its location.

Let us look a bit closer at the vector potential (3.28) near the Dirac string. Denoting by $\varrho$ the linear distance to the string (see the right of Fig. 7), in the limit $\varrho \to 0$ we can write

$$\mathbf{A} \approx \frac{1}{2\pi}\frac{g}{\varrho}\mathbf{u}_\varphi. \tag{3.32}$$

This expression should be familiar from elementary electrodynamics, since it represents the vector potential outside an infinite solenoid. The Dirac string can be pictured then as an infinitely thin solenoid pumping magnetic flux into the monopole which, according to the limiting value of the integral in Eq. (3.31), is actually equal to the outgoing flux through a closed surface surrounding the monopole.

In Box 4 we learned a way to "detect solenoids" by their imprints on the wave function of charged quantum particles detectable by interference experiments. The Wilson loop of a particle with electric

charge $e$ going around the Dirac string is computed from the vector potential (3.32) and gives [see also Eq. (3.31)]

$$U(\gamma) = \exp\left(-ie \oint_\gamma d\boldsymbol{\ell} \cdot \mathbf{A}\right) = e^{-ieg}. \tag{3.33}$$

The absence of detectable interference requires this phase to be equal to one for any electrically charged particle, which amounts to the condition

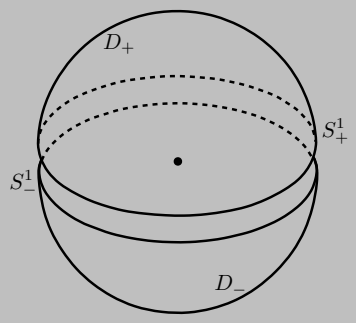
$$eg = 2\pi n \qquad \Longrightarrow \qquad e = \frac{2\pi}{g} n. \tag{3.34}$$

with $n$ an integer. This is a very interesting result, stating that the existence of a single magnetic monopole anywhere in the universe implies by consistency that electric charges have to be *quantized*. The quantization condition (3.34) remains invariant under electric–magnetic duality with $\theta = \frac{\pi}{2}$.

Unconfirmed sightings in cosmic rays notwithstanding [54,55], no evidence exists of magnetically charged particles at the energies explored. They are, however, an almost ubiquitous prediction of many theories beyond the SM, where they usually emerge as solitonic objects resulting from the spontaneous breaking in unified field theories leaving behind unbroken U(1)'s. Although they acquire masses of the order of the symmetry breaking scale, magnetic monopoles should have been created in huge amounts at the early stages of the universe's history. One of the original aims of cosmological inflation models was to dilute their presence in the early universe, thus accounting for their apparent absence.

**Box 5. Magnetic monopoles from topology**

The origin of all our troubles with the Dirac monopole was after all *topological*: although the vector potential of the magnetic monopole is locally well defined anywhere away from the origin, it cannot be extended globally to the sphere surrounding the monopole. There is however a way to avoid the singular Dirac string, which was pointed out by Tai Tsun Wu and Chen Ning Yang [56]. When computing the flux integral (3.30), instead of covering the sphere with a single patch cutting out the region around the place where the Dirac string crosses the surface (in our case, the south pole), we can be more sophisticated and use two patches, respectively centered at the north and south poles and overlapping at the equator. This is what we represent in the picture below, with $D_\pm$ the upper and lower hemispheres glued together along their respective boundaries $S^1_\pm$




On both $D_+$ and $D_-$ we can write vector potentials whose curls reproduce the expression of the

monopole field (3.27)

$$\mathbf{A}(\mathbf{r})_+ = \frac{1}{4\pi}\frac{g}{r}\tan\left(\frac{\theta}{2}\right)\mathbf{u}_\varphi \qquad 0 \le \theta \le \frac{\pi}{2},$$
$$\mathbf{A}(\mathbf{r})_- = -\frac{1}{4\pi}\frac{g}{r}\cot\left(\frac{\theta}{2}\right)\mathbf{u}_\varphi \qquad \frac{\pi}{2} \le \theta \le \pi. \tag{3.35}$$

The important point here is that both expressions are perfectly regular in their respective domains, so our vector potential is regular everywhere on the sphere $S^2 = D_+ \cup D_-$. An apparent obstacle arises in their overlap at the equator $\theta = \frac{\pi}{2}$, where the two expressions do not agree

$$\mathbf{A}(\mathbf{r})_+\Big|_{S^1_+} - \mathbf{A}(\mathbf{r})_-\Big|_{S^1_-} = \frac{1}{2\pi}\frac{g}{r}\mathbf{u}_\varphi. \tag{3.36}$$

This is however not a problem since, as we know, the vector potential is not uniquely defined. It is physically acceptable that the identification of the vector potentials at the equator is made modulo a gauge transformation, which is indeed the case here

$$\epsilon = -\frac{g}{2\pi}\varphi \qquad \Longrightarrow \qquad \mathbf{A}(\mathbf{r})_+\Big|_{S^1_+} = \mathbf{A}(\mathbf{r})_-\Big|_{S^1_-} - \boldsymbol{\nabla}\epsilon. \tag{3.37}$$

The magnetic flux due to the magnetic monopole at its center can be evaluated using these expressions as

$$\begin{aligned}\int_{S^2} d\mathbf{S}\cdot\mathbf{B} &= \int_{D_+} d\mathbf{S}\cdot(\boldsymbol{\nabla}\times\mathbf{A}_+) + \int_{D_-} d\mathbf{S}\cdot(\boldsymbol{\nabla}\times\mathbf{A}_-)\\ &= \oint_{S^1_+} d\boldsymbol{\ell}\cdot\mathbf{A}_+ + \oint_{S^1_-} d\boldsymbol{\ell}\cdot\mathbf{A}_-\\ &= \epsilon(2\pi) - \epsilon(0) = g,\end{aligned} \tag{3.38}$$

correctly reproducing (3.30). Notice that the two boundaries $S^1_\pm = \partial D_\pm$ have opposite orientations, so using Eq. (3.37) the second line combines into a single integral of $\epsilon'(\varphi)$ from 0 to $2\pi$.

The gauge function $\epsilon(\varphi)$ relating the vector potentials along the equator is not single-valued on $S^1$. This might pose a problem in the presence of quantum charged particles, since their wave functions also change under gauge transformations [see Eq. (3.10)]. In order to avoid multivaluedness of the wave function, we must require

$$e^{-ie\epsilon(0)} = e^{-ie\epsilon(2\pi)} \qquad \Longrightarrow \qquad e^{ieg} = 1, \tag{3.39}$$

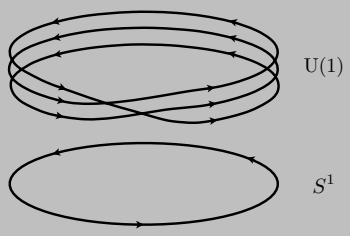
and the Dirac quantization condition (3.34) is retrieved. Alternatively, we can also notice that under a gauge transformation the action of a particle moving along the equator changes by $\Delta S = -eg$, as can be easily checked from Eq. (3.13). This has no effect in the Feynman path integral provided $eg = 2\pi n$, with $n \in \mathbb{Z}$, and the same result is obtained.

The Wu–Yang construction highlights the topological structure underlying the magnetic

monopole. Implementing the quantization condition $eg = 2\pi n$, the U(1) transformation (3.37) relating the vector potential of both hemispheres takes the form [cf. (3.11)]

$$U = e^{in\varphi}. \tag{3.40}$$

Since U(1) is the multiplicative group of complex phases, it can be identified with the unit circle. As we move once along the equator and the azimuthal angle $\varphi$ changes from 0 to $2\pi$, the gauge transformation (3.40) wraps $n$ times around U(1), as we illustrate here for the particular case $n = 3$




More technically speaking, when mapping the circle $S^1$ onto U(1) we encounter infinitely many sectors that cannot be smoothly deformed into one another and are distinguished by how many times the circle wraps around U(1). The corresponding integer is an element of the first homotopy group $\pi_1[\mathrm{U}(1)] = \mathbb{Z}$ classifying the continuous maps $U : S^1 \to \mathrm{U}(1)$ (see, for example, Refs. [57–60] for physicist-oriented overviews of basic concepts in differential geometry).

This should not come as a surprise. After all, at face value, our insistence in expressing the magnetic field as the curl of the vector potential is incompatible with having a nonvanishing value for $\boldsymbol{\nabla}\cdot\mathbf{B}$ as in Eq. (3.27). To reconcile these two facts we have to assume that although $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A}$ is valid on a contractible coordinate patch, there is no vector field $\mathbf{A}$ globally defined on the sphere with this property. This is why in our case the topologically trivial configuration $n = 0$ corresponds to zero magnetic charge and a vanishing magnetic field.

Looking at the symmetries of classical electrodynamics, we notice one conspicuously absent from the Maxwell's equations (3.1): Galilean invariance. It is amusing that Maxwell composed a fully relativistic invariant field theory some forty years before Einstein's formulation of special relativity. It took the latter's genius to realize that the tension between classical mechanics and electrodynamics was to be solved giving full credit to the Maxwell's equations and their spacetime symmetries. The price to pay was to modify Newtonian mechanics to make it applicable to systems involving velocities close to the speed of light.

### 3.2 Quantum electromagnetism

The easiest way to show the relativistic invariance of the Maxwell's equations is to rewrite them as tensor equations with respect to Poincaré transformations. To do so, we combine the scalar and vector electromagnetic potentials into a single four-vector

$$A^\mu \equiv (\phi, \mathbf{A}), \tag{3.41}$$

while electric and magnetic fields are codified in the field strength two-tensor

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{3.42}$$

The latter can be explicitly computed to be

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{pmatrix}, \tag{3.43}$$

where $\mathbf{E} = (E_x, E_y, E_z)$ and $\mathbf{B} = (B_x, B_y, B_z)$. The gauge transformations (3.4) are now expressed in the more compact form

$$A_\mu \longrightarrow A_\mu + \partial_\mu \epsilon, \tag{3.44}$$

which obviously leave $F_{\mu\nu}$ invariant. It is also convenient to define the dual field strength

$$\widetilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}F^{\alpha\beta}, \tag{3.45}$$

whose components are obtained from (3.43) by replacing $\mathbf{E} \to \mathbf{B}$ and $\mathbf{B} \to -\mathbf{E}$. Charge densities and currents are also merged into four-vectors

$$\begin{aligned} j_e^\mu &\equiv (\rho_e, \mathbf{j}_e), \\ j_m^\mu &\equiv (\rho_m, \mathbf{j}_m), \end{aligned} \tag{3.46}$$

in terms of which the four Maxwell's equations (3.1) are recast as

$$\begin{aligned} \partial_\mu F^{\mu\nu} &= j_e^\nu, \\ \partial_\mu \widetilde{F}^{\mu\nu} &= j_m^\nu. \end{aligned} \tag{3.47}$$

Some comments about the magnetic current are in order here. It should be noticed that the definition (3.42) automatically implies the Bianchi identity

$$\partial_\mu \widetilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\nu\sigma\alpha\beta}\partial_\sigma F_{\alpha\beta} = \epsilon^{\nu\sigma\alpha\beta}\partial_\sigma\partial_\alpha A_\beta = 0, \tag{3.48}$$

contradicting the second equation in (3.47). In fact, we have already encountered this problem in its noncovariant version when discussing magnetic monopoles: writing $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$ is incompatible with having $\boldsymbol{\nabla} \cdot \mathbf{B} \neq 0$. The solution given there is also applicable here. What happens is that (3.42) is valid locally but *not globally*. Magnetic monopoles can be described using the vector potential $A_\mu$, but the gauge field configuration needs to be topologically nontrivial.

The tensors $F_{\mu\nu}$ and $\widetilde{F}_{\mu\nu}$ can be used to construct quantities that are relativistic invariant. By

contracting them, we find the two invariants

$$F_{\mu\nu}F^{\mu\nu} = \widetilde{F}_{\mu\nu}\widetilde{F}^{\mu\nu} = -2\big(\mathbf{E}^2 - \mathbf{B}^2\big),$$
$$F_{\mu\nu}\widetilde{F}^{\mu\nu} = 2\mathbf{E}\cdot\mathbf{B}. \tag{3.49}$$

This implies that the complex combinations

$$\big(\mathbf{E} \pm i\mathbf{B}\big)^2 = \mathbf{E}^2 - \mathbf{B}^2 \pm 2i\mathbf{E}\cdot\mathbf{B}, \tag{3.50}$$

also remain invariant under the Lorentz group[3]. The present discussion is very relevant for building an action principle for classical electrodynamics. In particular, noticing that $F_{\mu\nu}\widetilde{F}^{\mu\nu} = 2\partial_\mu(A_\nu F^{\mu\nu})$ is a total derivative, the obvious choice is

$$\begin{aligned} S &= \int d^4x \left(-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + j^\mu A_\mu\right) \\ &= \int dtd^3x \left[\frac{1}{2}\big(\mathbf{E}^2 - \mathbf{B}^2\big) + \rho\phi - \mathbf{j}\cdot\mathbf{A}\right], \end{aligned} \tag{3.51}$$

which is also gauge invariant provided charge is conserved, $\partial_\mu j^\mu = 0$. Since from now on we will ignore the presence of magnetic charges, we drop the color code used so far, as well as the subscript in the electric density and current.

Although obtaining the Maxwell field equations from the action in (3.51) is straightforward, the canonical formalism is tricky. The reason is that $\dot{\phi}$ does not appear in the action and as a consequence the momentum conjugate to $A_0$ is identically zero. Thus, we have a constrained system that has to be dealt with using Dirac's formalism (see, for example, Ref. [14] for the details). At a practical level, we regard $\mathbf{A}$ and $\mathbf{E}$ as a pair of canonically conjugated variables

$$\big\{A_i(t,\mathbf{r}), E_j(t,\mathbf{r}')\big\}_{\rm PB} = \delta_{ij}\delta^{(3)}(\mathbf{r}-\mathbf{r}'). \tag{3.52}$$

Using $\dot{\mathbf{A}} = -\mathbf{E} - \boldsymbol{\nabla}\phi$, we construct the Hamiltonian

$$\begin{aligned} H &= \int dtd^3x \left[-\dot{\mathbf{A}}\cdot\mathbf{E} - \frac{1}{2}\big(\mathbf{E}^2 - \mathbf{B}^2\big) - \rho\phi + \mathbf{j}\cdot\mathbf{A}\right] \\ &= \int dtd^3x \left[\frac{1}{2}\big(\mathbf{E}^2 + \mathbf{B}^2\big) + \phi\big(\boldsymbol{\nabla}\cdot\mathbf{E} - \rho\big) + \mathbf{j}\cdot\mathbf{A}\right], \end{aligned} \tag{3.53}$$

where the term $-\mathbf{E}\cdot\boldsymbol{\nabla}\phi$ has been integrated by parts and the substitution $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A}$ is understood. Gauss' law $\boldsymbol{\nabla}\cdot\mathbf{E} = \rho$ emerges as a constraint preserved by time evolution

$$\big\{\boldsymbol{\nabla}\cdot\mathbf{E} - \rho, H\big\}_{\rm PB} = -\boldsymbol{\nabla}\cdot\mathbf{j} - \dot{\rho} \approx 0, \tag{3.54}$$

where we follow Dirac's notation and denote by $\approx$ identities that are satisfied after the equations of

[3] They change however under electric–magnetic duality, which mixes the two quantities introduced in (3.49).

motions are implemented. It also generates the gauge transformations of the vector potential

$$\delta\mathbf{A}(t,\mathbf{r}) = \Big\{\mathbf{A}(t,\mathbf{r}), \int d^3r'\epsilon(t,\mathbf{r}')\big[\boldsymbol{\nabla}\cdot\mathbf{E}(t,\mathbf{r}') - \rho(t,\mathbf{r}')\big]\Big\}_{\rm PB} = -\boldsymbol{\nabla}\epsilon(t,\mathbf{r}). \tag{3.55}$$

Solving the vacuum field equations written in terms of the gauge potential

$$\Box A_\mu - \partial_\mu\partial_\nu A^\nu = 0, \tag{3.56}$$

requires fixing the gauge freedom (3.44). To preserve relativistic covariance it is convenient to use the Lorenz gauge $\partial_\mu A^\mu = 0$ introduced in (3.5), so the gauge potential satisfies the wave equation $\Box A_\mu = 0$. Trying a plane wave ansatz

$$A_\mu(x) \sim \varepsilon_\mu(k,\lambda)e^{-ik_\mu x^\mu}, \tag{3.57}$$

the wave equation implies that the momentum vector $k^\mu$ is null

$$k_\mu k^\mu = 0 \qquad \Longrightarrow \qquad k^0 = \pm|\mathbf{k}|. \tag{3.58}$$

The parameter $\lambda$ in $\varepsilon_\mu(k,\lambda)$ labels the number of independent polarization vectors, which the Lorentz gauge condition force to be transverse

$$k^\mu\varepsilon_\mu(\mathbf{k},\lambda) = 0. \tag{3.59}$$

Using this condition we eliminate the temporal polarization in terms of the other three

$$\varepsilon_0(\mathbf{k},\lambda) = \frac{1}{|\mathbf{k}|}\mathbf{k}\cdot\boldsymbol{\varepsilon}(\mathbf{k},\lambda). \tag{3.60}$$

In addition, there is a residual gauge freedom preserving the Lorentz condition implemented on the plane wave solutions by shifts of the polarization vector proportional to the wave momentum

$$\varepsilon_\mu(\mathbf{k},\lambda) \longrightarrow \varepsilon_\mu(\mathbf{k},\lambda) + \alpha(\mathbf{k})k_\mu. \tag{3.61}$$

Using this freedom to set $\varepsilon_0(\mathbf{k},\lambda)$ to zero, we are left with just two independent transverse polarizations satisfying $\mathbf{k}\cdot\boldsymbol{\varepsilon}(\mathbf{k},\lambda) = 0$. The plane wave solution then reads

$$\mathbf{A}(t,\mathbf{r}) \sim \boldsymbol{\varepsilon}(\mathbf{k},\lambda)e^{-i|\mathbf{k}|t+i\mathbf{k}\cdot\mathbf{r}}, \tag{3.62}$$

with $A_0 = 0$ and $\lambda = \pm 1$ labelling the two transverse polarizations, that in the following we will respectively identify with right–left circular polarizations[4], $\boldsymbol{\varepsilon}(\mathbf{k},\lambda)^* = \boldsymbol{\varepsilon}(\mathbf{k},-\lambda)$. They moreover satisfy

$$\boldsymbol{\varepsilon}(\mathbf{k},\lambda)\cdot\big[\mathbf{k}\times\boldsymbol{\varepsilon}(\mathbf{k},\lambda')\big] = i\lambda|\mathbf{k}|\delta_{\lambda,-\lambda'}. \tag{3.63}$$

[4]For a massive vector field the Lorentz condition $\partial_\mu A^\mu = 0$ is still satisfied as an integrability condition of the equations of motion $\partial_\mu F^{\mu\nu} + m^2A^\nu = 0$ and Eq. (3.60) therefore holds. The key difference lies in that the residual freedom (3.61) is absent and we have an additional longitudinal polarization (i.e., aligned with $\mathbf{k}$) in addition to the two transverse ones.

This identity will be useful later on.

Since the field equations are linear, a general solution can be written as a superposition of the plane wave solutions (3.62) and their complex conjugates. Upon quantization the coefficients in this expansion become operators and we can write a general expression for the gauge field operator

$$\widehat{\mathbf{A}}(t,\mathbf{r}) = \sum_{\lambda=\pm 1}\int \frac{d^3k}{(2\pi)^3}\frac{1}{2|\mathbf{k}|}\left[\boldsymbol{\varepsilon}(\mathbf{k},\lambda)\widehat{a}(\mathbf{k},\lambda)e^{-i|\mathbf{k}|t+i\mathbf{k}\cdot\mathbf{r}} + \boldsymbol{\varepsilon}(\mathbf{k},\lambda)^*\widehat{a}(\mathbf{k},\lambda)^\dagger e^{i|\mathbf{k}|t-i\mathbf{k}\cdot\mathbf{r}}\right], \tag{3.64}$$

where, with our gauge fixing, $\widehat{A}_0(t,\mathbf{r}) = 0$. The integration measure appearing in this expression results from integrating over all four-dimensional momenta lying on the upper light-cone in Fig. 4

$$\int \frac{d^4k}{(2\pi)^4}\delta(k_\mu k^\mu)\theta(k^0)[\ldots] = \int \frac{d^3k}{(2\pi)^3}\frac{1}{2|\mathbf{k}|}[\ldots], \tag{3.65}$$

and is by construction Lorentz invariant. The quantum states of the theory are vectors in the space of states the operator (3.64) acts on. To determine it and therefore the excitations of the quantum field, we establish first the algebra of operators and then find a representation. This is done by applying the canonical quantization prescription replacing classical Poisson brackets with quantum commutators

$$i\{\cdot,\cdot\}_{\mathrm{PB}} \longrightarrow [\cdot,\cdot]. \tag{3.66}$$

Using the definition $\widehat{\mathbf{E}} = \partial_0\widehat{\mathbf{A}}$, the electric field operator is computed to be

$$\widehat{\mathbf{E}}(t,\mathbf{r}) = -\frac{i}{2}\sum_{\lambda=\pm 1}\int \frac{d^3k}{(2\pi)^3}\left[\boldsymbol{\varepsilon}(\mathbf{k},\lambda)\widehat{a}(\mathbf{k},\lambda)e^{-i|\mathbf{k}|t+i\mathbf{k}\cdot\mathbf{r}} - \boldsymbol{\varepsilon}(\mathbf{k},\lambda)^*\widehat{a}(\mathbf{k},\lambda)^\dagger e^{i|\mathbf{k}|t-i\mathbf{k}\cdot\mathbf{r}}\right]. \tag{3.67}$$

Classically, the electric field is canonically conjugate to the vector potential [see Eq. (3.52)], so the prescription (3.66) gives its equal-time commutator with the gauge field

$$[A_i(t,\mathbf{r}), E_i(t,\mathbf{r}')] = i\delta_{ij}\delta^{(3)}(\mathbf{r}-\mathbf{r}') \tag{3.68}$$

that translates into the following commutation relations for the operators $\widehat{a}(\mathbf{k},\lambda)$ and their Hermitian conjugates

$$\begin{aligned}[\widehat{a}(\mathbf{k},\lambda),\widehat{a}(\mathbf{k}',\lambda')^\dagger] &= (2\pi)^3 2|\mathbf{k}|\delta_{\lambda\lambda'}\delta^{(3)}(\mathbf{k}-\mathbf{k}'),\\ [\widehat{a}(\mathbf{k},\lambda),\widehat{a}(\mathbf{k}',\lambda')] &= [\widehat{a}(\mathbf{k},\lambda)^\dagger,\widehat{a}(\mathbf{k}',\lambda')^\dagger] = 0.\end{aligned} \tag{3.69}$$

This algebra is reminiscent of the one of creation–annihilation operators in the quantum harmonic oscillator. Introducing a properly normalized vacuum state $|0\rangle$ to be annihilated by all $\widehat{a}(\mathbf{k};\lambda)$, we define the vector

$$|\mathbf{k},\lambda\rangle = \widehat{a}(\mathbf{k},\lambda)^\dagger|0\rangle, \tag{3.70}$$

representing a one-photon state with momentum $\mathbf{k}$ and helicity $\lambda$. These states are covariantly normalized

according to

$$\langle \mathbf{k}, \lambda | \mathbf{k}', \lambda' \rangle = (2\pi)^3 2|\mathbf{k}| \delta_{\lambda\lambda'} \delta^{(3)}(\mathbf{k} - \mathbf{k}'), \tag{3.71}$$

as can be seen from Eq. (3.69). Multiple photon states are obtained by successive application of creation operators

$$|\mathbf{k}_1, \lambda_1; \mathbf{k}_2, \lambda_2; \ldots; \mathbf{k}_n, \lambda_n\rangle = \widehat{a}(\mathbf{k}_1, \lambda_1)^\dagger \widehat{a}(\mathbf{k}_2, \lambda_2)^\dagger \ldots \widehat{a}(\mathbf{k}_n, \lambda_n)^\dagger |0\rangle. \tag{3.72}$$

From the commutation relation of creation operators given in (3.69) we see that the multi-photon state is even under the interchange of whatever two photons, as it should be for bosons.

Although we have been talking about photons, we must check that the states (3.70) have the quantum numbers corresponding to these particles. So, first we compute their energy by writing the quantum Hamiltonian. Going back to Eq. (3.53), we set the sources to zero ($\rho = 0$ and $\mathbf{j} = \mathbf{0}$) and replace the electric and magnetic field for their corresponding operators. A first thing to notice is that the electric field (3.67) satisfies the Gauss law $\boldsymbol{\nabla} \cdot \widehat{\mathbf{E}} = 0$ as a consequence of the transversality condition of the polarizations vectors. Computing in addition $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$ and after some algebra, we find

$$\widehat{H} = \sum_{\lambda=\pm 1} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2|\mathbf{k}|} |\mathbf{k}| \widehat{a}(\mathbf{k}, \lambda)^\dagger \widehat{a}(\mathbf{k}, \lambda) + \frac{1}{2} \sum_{\lambda=\pm 1} \int d^3k \, |\mathbf{k}| \delta^{(3)}(\mathbf{0}). \tag{3.73}$$

The second term on the right-hand side represents the energy of the vacuum state

$$\widehat{H}|0\rangle = \left( \frac{1}{2} \sum_{\lambda=\pm 1} d^3k \, |\mathbf{k}| \delta^{(3)}(\mathbf{0}) \right) |0\rangle \tag{3.74}$$

and is doubly divergent. One infinity originates in the delta function and comes about because we are working at infinite volume, a type of divergence that in QFT is designated as *infrared* (IR). It can be regularized by setting our system in a box of volume $V$, which replaces $(2\pi)^3 \delta^{(3)}(\mathbf{0})$. Proceeding in this way, we write the energy density of the vacuum as

$$\rho_{\text{vac}} \equiv \frac{E_{\text{vac}}}{V} = \frac{1}{2} \sum_{\lambda=\pm 1} \int \frac{d^3k}{(2\pi)^3} |\mathbf{k}|. \tag{3.75}$$

This expression has the obvious interpretation of being the result of adding the zero-point energies of infinitely many harmonic oscillators, each with frequency $\omega = |\mathbf{k}|$. It is still divergent, and since the infinity originates in the integration over arbitrarily high momenta, it is called *ultraviolet* (UV). A way to get rid of it is assuming that $|\mathbf{k}| < \Lambda_{\text{UV}}$, so that after carrying out the integral, the vacuum energy density is given by

$$\rho_{\text{vac}} = \frac{1}{16\pi^2} \Lambda_{\text{UV}}^4. \tag{3.76}$$

In the spirit of effective field theory this UV cutoff is physically interpreted as the energy scale at which our description of the electromagnetic field breaks down and has to be replaced by some more general

theory.

The vacuum energy density (3.76) is at the origin of the cosmological constant problem. Due to its strong dependence on the UV cutoff, when we add the contributions of all known quantum fields to $\rho_{\rm vac}$ the result is many orders of magnitude larger than the one measured through cosmological observations. The way to handle this mismatch is by assuming the existence of a nonzero cosmological constant $\Lambda_c$ contribution to the total vacuum energy of the universe as

$$\rho_{\rm vac} = \frac{\Lambda_c}{8\pi G_N} + \sum_i \rho_{{\rm vac},i}, \tag{3.77}$$

where the sum is over all quantum fields in nature. Identifying the UV cutoff with the Planck energy, $\Lambda_{\rm UV} \simeq \Lambda_{\rm Pl}$, the cosmological constant has to be fine tuned over 120 orders of magnitude in order to cancel the excess contribution of the quantum fields to the vacuum energy density of the universe (see, for example, Refs. [61–63] for comprehensive reviews).

Let us get rid of the vacuum energy for the time being by subtracting it from the Hamiltonian (3.73). Acting with this subtracted Hamiltonian on the multiparticle states (3.72), we find they are energy eigenstates

$$\widehat{H}|\mathbf{k}_1,\lambda_1;\mathbf{k}_2,\lambda_2;\ldots;\mathbf{k}_n,\lambda_n\rangle = \big(|\mathbf{k}_1|+|\mathbf{k}_2|+\ldots+|\mathbf{k}_n|\big)|\mathbf{k}_1,\lambda_1;\mathbf{k}_2,\lambda_2;\ldots;\mathbf{k}_n,\lambda_n\rangle, \tag{3.78}$$

with the eigenvalue giving the energy of $n$ free photons with momenta $\mathbf{k}_1,\mathbf{k}_2,\ldots,\mathbf{k}_n$. The field momentum, on the other hand, is given by the Poynting operator

$$\begin{aligned}\widehat{\mathbf{P}} &= \int d^3r\, \mathbf{E}(t,\mathbf{r})\times\mathbf{B}(t,\mathbf{r}) \\ &= \sum_{\lambda=\pm1}\int \frac{d^3k}{(2\pi)^3}\frac{1}{2|\mathbf{k}|}\mathbf{k}\,\widehat{a}(\mathbf{k},\lambda)^\dagger\widehat{a}(\mathbf{k},\lambda),\end{aligned} \tag{3.79}$$

where, unlike for the Hamiltonian, here there is no vacuum contribution due to the rotational invariance of $|0\rangle$. Its action on the states (3.72) gives

$$\widehat{\mathbf{P}}|\mathbf{k}_1,\lambda_1;\mathbf{k}_2,\lambda_2;\ldots;\mathbf{k}_n,\lambda_n\rangle = \big(\mathbf{k}_1+\mathbf{k}_2+\ldots+\mathbf{k}_n\big)|\mathbf{k}_1,\lambda_1;\mathbf{k}_2,\lambda_2;\ldots;\mathbf{k}_n,\lambda_n\rangle, \tag{3.80}$$

showing that the vector $\mathbf{k}$ labelling the one-particle states (3.70) is rightly interpreted as the photon momentum. Finally, we compute the spin momentum operator

$$\begin{aligned}\widehat{\mathbf{S}} &= \int d^3x\,\widehat{\mathbf{A}}\times\widehat{\mathbf{E}} \\ &= i\sum_{\lambda,\lambda'=\pm1}\int\frac{d^3k}{(2\pi)^2}\frac{1}{2|\mathbf{k}|}\boldsymbol{\varepsilon}(\mathbf{k},\lambda)\times\boldsymbol{\varepsilon}(\mathbf{k},\lambda')^*\,\widehat{a}(\mathbf{k},\lambda)^\dagger\widehat{a}(\mathbf{k},\lambda).\end{aligned} \tag{3.81}$$

Acting on a one-particle state (3.70), we find

$$\widehat{\mathbf{S}}|\mathbf{k},\lambda\rangle = i \sum_{\lambda,\lambda'=\pm 1} \boldsymbol{\varepsilon}(\mathbf{k},\lambda)\times\boldsymbol{\varepsilon}(\mathbf{k},\lambda')^*|\mathbf{k},\lambda\rangle. \tag{3.82}$$

We now project this expression on the direction of the photon's momentum, to find the helicity operator acting on the single photon state

$$\widehat{h}|\mathbf{k},\lambda\rangle \equiv \frac{\mathbf{k}}{|\mathbf{k}|}\cdot\widehat{\mathbf{S}}|\mathbf{k},\lambda\rangle = \frac{i}{|\mathbf{k}|}\sum_{\lambda,\lambda'=\pm 1}\mathbf{k}\cdot\left[\boldsymbol{\varepsilon}(\mathbf{k},\lambda)\times\boldsymbol{\varepsilon}(\mathbf{k},\lambda')^*\right]|\mathbf{k},\lambda\rangle. \tag{3.83}$$

Using the relation (3.63) to evaluate the mixed product inside the sum, we arrive at

$$\widehat{h}|\mathbf{k},\lambda\rangle = \lambda|\mathbf{k},\lambda\rangle, \tag{3.84}$$

which shows that $\lambda$ is indeed the helicity of the photon. We have convinced ourselves that our interpretation of the quantum numbers describing the Hamiltonian eigenstates was correct, and they describe states with an arbitrary number of free photons of definite momenta and helicities. Photons therefore emerge as the elementary excitations of the quantum electromagnetic field.

### 3.3 Some comments on quantum fields

The previous calculation also teaches an important lesson: the space of states of a free quantum field (in this case the electromagnetic field) is in fact a Fock space, i.e., the direct sum of Hilbert spaces spanned by the $n$-particle states (3.72),

$$\mathscr{F} = \bigoplus_{n=0}^{\infty}\mathscr{H}_n, \tag{3.85}$$

where we take $\mathscr{H}_0 = L\{|0\rangle\}$, the one-dimensional linear space generated by the vacuum state $|0\rangle$. We have shown that the canonical commutation relations (3.68) admit a representation in the Fock space. Although we have done this for the free sourceless Maxwell's theory, it is also the case for any other free field theory, as we will see in other examples below. Including interactions does not change this, provided they are sufficiently weak and to be treated in perturbation theory. Thus, the first step in describing a physical system is to identify the weakly coupled degrees of freedom, whose multiparticle states span the Fock space representing the asymptotic states in scattering experiments of the type carried out everyday in high energy facilities around the world. This is well illustrated by the case of QCD discussed in the Introduction (see page 6), where while the asymptotic states are described by hadrons, the fundamental interactions taking place are described in terms of weakly coupled quarks and gluons[5].

[5] A technical caveat: Haag's theorem [64], however, states that for a general interacting QFT there exists no Fock space representation of the canonical commutation relation. This is usually interpreted as implying that full interacting QFT is not a theory of particles [65–67].

**Box 6. Complex fields and antiparticles**

The analysis presented for electrodynamics carries over to the quantization of other free fields. A simple but particularly interesting example is provided by a complex scalar field, with action

$$S = \int d^4x \left( \partial_\mu \varphi^* \partial^\mu \varphi - m^2 \varphi^* \varphi \right). \tag{3.86}$$

Life is now simpler since there is no gauge freedom and the Hamiltonian formalism is straightforward. We compute the conjugate momentum and the canonical Poisson brackets

$$\pi(t,\mathbf{r}) = \frac{\delta S}{\delta \partial_0 \varphi(t,\mathbf{r})} = \partial_0 \varphi(t,\mathbf{r})^* \quad \Longrightarrow \quad \left\{ \varphi(t,\mathbf{r}), \pi(t,\mathbf{r}') \right\}_{\rm PB} = \delta^{(3)}(\mathbf{r}-\mathbf{r}'), \tag{3.87}$$

with the corresponding expression for the complex conjugate fields, $\varphi(t,\mathbf{r})^*$ and $\pi(t,\mathbf{r})^*$. The Hamiltonian is then given by

$$H = \int d^3r \left[ \pi^* \pi + (\boldsymbol{\nabla} \varphi^*) \cdot (\boldsymbol{\nabla} \varphi) + m^2 \varphi^* \varphi \right]. \tag{3.88}$$

The equation of motion derived from the action (3.86) is the Klein–Gordon equation

$$\left( \Box + m^2 \right) \varphi = 0, \tag{3.89}$$

which admits plane wave solutions of the form

$$\varphi(x) \sim e^{i p_\mu x^\mu}, \tag{3.90}$$

with $p_\mu$ satisfying the mass-shell condition

$$p_\mu p^\mu = m^2 \qquad \Longrightarrow \qquad p^0 \equiv \pm E_{\mathbf{p}} = \pm \sqrt{\mathbf{p}^2 + m^2}. \tag{3.91}$$

As with the electromagnetic field, the corresponding quantum fields are an operator-valued superposition of plane waves

$$\begin{aligned} \widehat{\varphi}(t,\mathbf{r}) &= \int \frac{d^3p}{(2\pi)^3} \frac{1}{2E_{\mathbf{p}}} \left[ \widehat{\alpha}(\mathbf{p}) e^{-iE_{\mathbf{p}}t + i\mathbf{p}\cdot\mathbf{r}} + \widehat{\beta}(\mathbf{p})^\dagger e^{iE_{\mathbf{p}}t - i\mathbf{p}\cdot\mathbf{r}} \right], \\ \widehat{\varphi}(t,\mathbf{r})^\dagger &= \int \frac{d^3p}{(2\pi)^3} \frac{1}{2E_{\mathbf{p}}} \left[ \widehat{\beta}(\mathbf{p}) e^{-iE_{\mathbf{p}}t + i\mathbf{p}\cdot\mathbf{r}} + \widehat{\alpha}(\mathbf{p})^\dagger e^{iE_{\mathbf{p}}t - i\mathbf{p}\cdot\mathbf{r}} \right], \end{aligned} \tag{3.92}$$

while the operator associated to the canonically conjugate momentum is given by

$$\begin{aligned} \widehat{\pi}(t,\mathbf{r}) &= -\frac{i}{2} \int \frac{d^3p}{(2\pi)^3} \left[ \widehat{\beta}(\mathbf{p}) e^{-iE_{\mathbf{p}}t + i\mathbf{p}\cdot\mathbf{r}} - \widehat{\alpha}(\mathbf{p})^\dagger e^{iE_{\mathbf{p}}t - i\mathbf{p}\cdot\mathbf{r}} \right], \\ \widehat{\pi}(t,\mathbf{r})^\dagger &= \frac{i}{2} \int \frac{d^3p}{(2\pi)^3} \left[ \widehat{\alpha}(\mathbf{p}) e^{-iE_{\mathbf{p}}t + i\mathbf{p}\cdot\mathbf{r}} - \widehat{\beta}(\mathbf{p})^\dagger e^{iE_{\mathbf{p}}t - i\mathbf{p}\cdot\mathbf{r}} \right]. \end{aligned} \tag{3.93}$$

The key observation here is that, since $\widehat{\varphi}$ is not Hermitian, the two operators $\widehat{\alpha}(\mathbf{p})$ and $\widehat{\beta}(\mathbf{p})$ cannot be identified, as it was the case with the electromagnetic field. Imposing the equal-time canonical commutation relations induced by the canonical Poisson brackets [see Eq. (3.87)] leads to the following algebra of operators

$$\begin{aligned}[\widehat{\alpha}(\mathbf{p}),\widehat{\alpha}(\mathbf{p}')^\dagger] &= (2\pi)^3 2E_{\mathbf{p}}\delta^{(3)}(\mathbf{p}-\mathbf{p}'),\\ [\widehat{\alpha}(\mathbf{p}),\widehat{\alpha}(\mathbf{p}')] &= [\widehat{\alpha}(\mathbf{p})^\dagger,\widehat{\alpha}(\mathbf{p}')^\dagger] = 0,\end{aligned} \tag{3.94}$$

and corresponding expressions for $\widehat{\beta}(\mathbf{p})$ and $\widehat{\beta}(\mathbf{p})^\dagger$, with both types of operators commuting with each other. As with the photons, the Fock space of states is built by acting with $\widehat{\alpha}(\mathbf{p})^\dagger$'s and $\widehat{\beta}(\mathbf{p})^\dagger$'s on the vacuum state $|0\rangle$, which is itself annihilated by $\widehat{\alpha}(\mathbf{p})$'s and $\widehat{\beta}(\mathbf{p})$'s

$$|\mathbf{p}_1,\ldots,\mathbf{p}_n;\mathbf{q}_1,\ldots,\mathbf{q}_m\rangle = \widehat{\alpha}(\mathbf{p}_1)^\dagger\ldots\widehat{\alpha}(\mathbf{p}_n)^\dagger\widehat{\beta}(\mathbf{q}_1)^\dagger\ldots\widehat{\beta}(\mathbf{q}_1)^\dagger|0\rangle, \tag{3.95}$$

where we have distinguished the momenta associated with the two kinds of creation operators. Notice that, since the operators on the right-hand side of this expression commute with each other, the order in which we list the momenta $\mathbf{p}_1,\ldots,\mathbf{p}_n$ and $\mathbf{q}_1,\ldots,\mathbf{q}_m$ is irrelevant, signalling that both types of excitations are bosons.

The states constructed in (3.95) in fact diagonalize the Hamiltonian

$$\widehat{H} = \int \frac{d^3p}{(2\pi)^3}\frac{1}{2E_{\mathbf{p}}}E_{\mathbf{p}}\Big[\widehat{\alpha}(\mathbf{p})^\dagger\widehat{\alpha}(\mathbf{p}) + \widehat{\beta}(\mathbf{p})^\dagger\widehat{\beta}(\mathbf{p})\Big], \tag{3.96}$$

where we have subtracted a UV and IR divergent vacuum contribution similar to the one encountered in Eq. (3.73). Indeed, it is not difficult to show that

$$\begin{aligned}&\widehat{H}|\mathbf{p}_1,\ldots,\mathbf{p}_n;\mathbf{q}_1,\ldots,\mathbf{q}_m\rangle\\ &\qquad = \big(E_{\mathbf{p}_1}+\ldots+E_{\mathbf{p}_n}+E_{\mathbf{q}_1}+\ldots+E_{\mathbf{q}_m}\big)|\mathbf{p}_1,\ldots,\mathbf{p}_n;\mathbf{q}_1,\ldots,\mathbf{q}_m\rangle,\end{aligned} \tag{3.97}$$

from where we conclude that the elementary excitations of the quantum real scalar field are free scalar particles with well-defined energy and momentum. These particles come in two different types depending on whether they are created by $\widehat{\alpha}(\mathbf{p})^\dagger$ or $\widehat{\beta}(\mathbf{p})^\dagger$, since they share the same dispersion relation, they have equal masses.

The obvious question is what distinguishes physically one from the other. To answer, we have to study the symmetries of the classical theory. A look at the action (3.86) shows that it is invariant under global phase rotations of the complex field

$$\varphi(x) \longrightarrow e^{i\vartheta}\varphi(x), \qquad \varphi(x)^* \longrightarrow e^{-i\vartheta}\varphi(x), \tag{3.98}$$

with $\vartheta$ a constant real parameter. Noether's theorem (see page 57) states that associated to this

symmetry there must be a conserved current, whose expression turns out to be

$$j^\mu = i\varphi^* \overleftrightarrow{\partial}^\mu \varphi \equiv i\varphi^* \partial^\mu \varphi - i(\partial^\mu \varphi^*)\varphi \quad \Longrightarrow \quad \partial_\mu j^\mu = 0. \tag{3.99}$$

In particular, the conserved charge is given by

$$Q = \int d^3r \left(\varphi^* \pi^* - \pi\varphi\right), \tag{3.100}$$

and once classical fields are replaced by their operator counterparts (and complex by Hermitian conjugation), we have the following form for the charge operator:

$$\widehat{Q} = \int \frac{d^3p}{(2\pi)^3} \frac{1}{2E_{\mathbf{p}}} \Big[\widehat{\alpha}(\mathbf{p})^\dagger \widehat{\alpha}(\mathbf{p}) - \widehat{\beta}(\mathbf{p})^\dagger \widehat{\beta}(\mathbf{p})\Big]. \tag{3.101}$$

By acting with it on one-particle states, we get

$$\begin{aligned} \widehat{Q}|\mathbf{p};0\rangle &= |\mathbf{p};0\rangle, \\ \widehat{Q}|0;\mathbf{q}\rangle &= -|0;\mathbf{q}\rangle, \end{aligned} \tag{3.102}$$

showing that the conserved charge distinguishes the excitations generated by $\widehat{\alpha}(\mathbf{p})^\dagger$ from those generated by $\widehat{\beta}(\mathbf{p})^\dagger$. Moreover, the complex scalar field can be coupled to the electromagnetic field by identifying the current (3.99) with the one appearing in the Maxwell action (3.51), its conservation guaranteeing gauge invariance of the combined action. Thus, the two kinds of particles with the same mass and spin have opposite electric charges and are identified as particles and antiparticles. The complex (i.e., non-Hermitian) character of the scalar field is crucial to have both particles and antiparticles. In the case of the gauge field $\widehat{\mathbf{A}}$, hermiticity identifies the operators associated with positive and negative energy plane wave solutions as conjugate to each other, making the photon its own antiparticle.

It is time we address another symmetry present in Maxwell's electrodynamics that is of pivotal importance for QFT as a whole: scale invariance. Looking at the free electromagnetic action

$$S_{\text{EM}} = -\frac{1}{4} \int d^4x \, F_{\mu\nu} F^{\mu\nu}, \tag{3.103}$$

we notice the absence of any dimensionful parameters, unlike in the case of the complex scalar field action (3.86), where we have a parameter $m$ that turns out to be the mass of its elementary quantum excitations. It seems that the free Maxwell's theory should be invariant under changes of scale.

To formulate the idea of scale invaraince in more general and precise mathematical terms, let us assume a scale transformation of the coordinates

$$x^\mu \longrightarrow \lambda x^\mu, \tag{3.104}$$

with $\lambda$ a nonzero real parameter, combined with the following scaling of the fields in the theory

$$\Phi(x) \longrightarrow \lambda^{-\Delta_\Phi}\Phi(\lambda^{-1}x), \tag{3.105}$$

where $\Delta_\Phi$ is called the field's scaling dimension. Applying these transformations to the particular case of the action (3.103), we find

$$S_{\rm EM} \longrightarrow \lambda^{2-2\Delta_A} S_{\rm EM}, \tag{3.106}$$

so that by setting $\Delta_A = 1$ the action remains invariant under scale transformations.

We will explore now whether the scale invariance of the free Maxwell's theory is preserved by the coupling of the electromagnetic field to charged matter. As an example, let us consider the complex scalar field we studied in Box 6, but now coupled to an electromagnetic field

$$\begin{aligned} S &= \int d^4x \left\{ \partial_\mu\varphi^*\partial^\mu\varphi - m\varphi^*\varphi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + ie\big[\varphi^*\partial_\mu\varphi - (\partial_\mu\varphi^*)\varphi\big]A^\mu + e^2\varphi^*\varphi A_\mu A^\mu \right\} \\ &= \int d^4x \left[ \big(\partial_\mu + ieA_\mu\big)\varphi^*\big(\partial^\mu - ieA^\mu\big)\varphi - m^2\varphi^*\varphi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \right]. \end{aligned} \tag{3.107}$$

Here, besides the coupling $j_\mu A^\mu$ suggested by the Maxwell's equations, we also have the term $e^2\varphi^*\varphi A_\mu A^\mu$, that has to be added to preserve the invariance of the whole action under the gauge transformations[6]

$$\varphi \to e^{ie\epsilon(x)}\varphi^*, \qquad \varphi^* \to e^{-ie\epsilon(x)}\varphi^*, \qquad A_\mu \to A_\mu + \partial_\mu\epsilon(x). \tag{3.108}$$

Setting the scaling dimension of the scalar field to one, $\Delta_\varphi = 1$, we easily check that the scale invariance of the action (3.107) is only broken by the mass term of the scalar field

$$m\int d^4x\,\varphi^*\varphi \longrightarrow \lambda^2 m\int d^4x\,\varphi^*\varphi. \tag{3.109}$$

This confirms our intuition that classical scale invariance is incompatible with the presence of dimensionful parameters in the action. It also shows that taking $m = 0$ the photon can be coupled to scalar charged matter preserving the classical scale invariance of the free Maxwell theory. Several essential field theories share this property besides the example just analyzed, most notably QCD once all quark masses are set to zero.

The discussion above has emphasized the term *classical* whenever referring to scale invariance. The reason is that this is a very fragile symmetry once quantum effects are included. For example, let us go back to the action (3.107) but now take $m = 0$. The classical scale invariance is broken by quantum effects in the sense that, once the quantum corrections induced by interactions are taken into account, physics depends on the energy scale at which experiments are carried out. One way in which this happens is by the electric charge of the elementary excitations of the field depending on the energy at which it is

[6]Notice that the combination $(\partial_\mu - ieA_\mu)\varphi$ appearing in the second line of Eq. (3.107) transforms as the complex scalar field itself. It defines the gauge covariant derivative of $\varphi$, its name reflecting its covariant transformation under gauge transformations, $D_\mu\varphi \to e^{ie\epsilon(x)}D_\mu\varphi$.

measured[7]. We will further elaborate on this phenomenon in Section 10.

## 4 Some group theory and some more wave equations

Scalars and vectors are relatively intuitive objects, which is why we did not need to get into sophisticated mathematics to handle them. In nature, however, elementary scalar fields are rare (as of today, we know just one, the Higgs field) and vector fields only describe interactions, not matter. To describe fundamental physics we need fields whose excitations are particles with spin-$\frac{1}{2}$, such as the electron, the muon, and the quarks. We have to plunge into group theory before we can formulate these objects rigorously.

### 4.1 Special relativity and group theory

Let us begin by giving a more technical picture of the Lorentz group. We have defined it as the set of linear transformations of the spacetime coordinates $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ satisfying (2.10) and therefore preserving the Minkowski metric. The first thing to be noticed is that this condition implies the inequality

$$(\Lambda^0{}_0)^2 - \sum_{i=1}^{3} (\Lambda^i{}_0)^2 = 1 \qquad \Longrightarrow \qquad |\Lambda^0{}_0| \geq 1. \tag{4.1}$$

The sign of $\Lambda^0{}_0$ indicates whether or not the transformed time coordinate "flows" in the same direction as the original one, this being why transformations with $\Lambda^0{}_0 \geq 1$ are called *orthochronous*. At the same time, Eq. (2.10) also implies

$$(\det \Lambda)^2 = 1 \qquad \Longrightarrow \qquad \det \Lambda = \pm 1. \tag{4.2}$$

Since it is not possible to change the signs of $\Lambda^0{}_0$ or $\det \Lambda$ by continuously deforming Lorentz transformations, the full Lorentz group is seen to be composed of four different connected components:

$$\begin{aligned}
&\mathfrak{L}_+^\uparrow : \text{ proper, orthochronous transformations with } \Lambda^0{}_0 \geq 1 \text{ and } \det \Lambda = 1, \\
&\mathfrak{L}_+^\downarrow : \text{ proper, non-orthochronous transformations with } \Lambda^0{}_0 \leq -1 \text{ and } \det \Lambda = 1, \\
&\mathfrak{L}_-^\uparrow : \text{ improper, orthochronous transformations with } \Lambda^0{}_0 \geq 1 \text{ and } \det \Lambda = -1, \\
&\mathfrak{L}_-^\downarrow : \text{ improper, non-orthochronous transformations with } \Lambda^0{}_0 \leq -1 \text{ and } \det \Lambda = -1.
\end{aligned} \tag{4.3}$$

The set of proper orthochronous transformations $\mathfrak{L}_+^\uparrow$ contains the identity, while the remaining ones respectively include the time reversal operation (T : $x^0 \to -x^0$), parity (P : $x^i \to -x^i$), and the composition of both. As indicated in Fig. 8, these discrete transformations also map the identity's connected component to the other three,

$$\mathrm{T} : \mathfrak{L}_+^\uparrow \longrightarrow \mathfrak{L}_-^\downarrow, \qquad \mathrm{P} : \mathfrak{L}_+^\uparrow \longrightarrow \mathfrak{L}_-^\uparrow, \qquad \mathrm{PT} : \mathfrak{L}_+^\uparrow \longrightarrow \mathfrak{L}_+^\downarrow. \tag{4.4}$$

[7] Incidentally, most scale invariant QFTs are also invariant under the full conformal group, i.e., the group of coordinate transformations preserving the light cone.

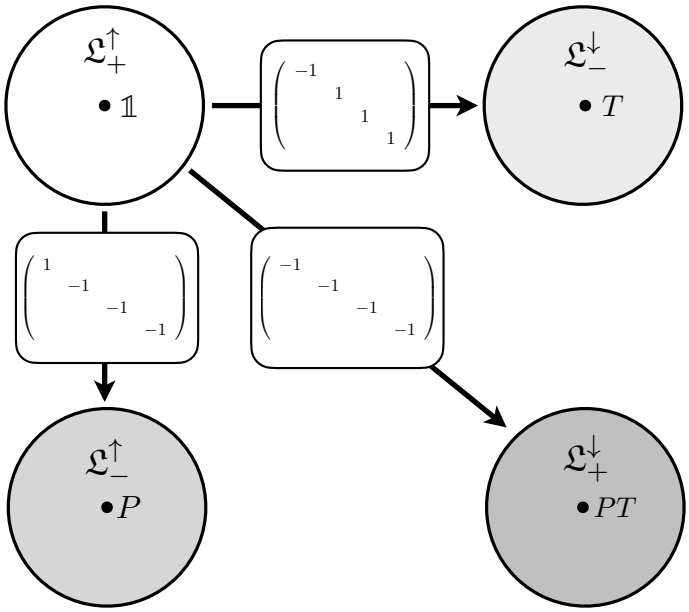



**Fig. 8:** The four connected components of the Lorentz group. The matrices indicate the transformations $P$, $T$, and $PT$ mapping the connected component of the identity $\mathfrak{L}_+^\uparrow$ to the other three.

Thus, to study the irreducible representations (irreps) of the Lorentz group it is enough to restrict our attention to $\mathfrak{L}_+^\uparrow \equiv$ SO(1,3).

As discussed in page 9, the proper group Lorentz SO(1,3) is composed by two kinds of transformations: rotations with angle $0 \leq \phi < 2\pi$ around an axis defined by the unit vector $\mathbf{u}$ and boosts with rapidity $\lambda$ along the direction set by the unit vector $\mathbf{e}$. Since we are on the connected component of the identity, the transformations can be written by exponentiation of the Lie algebra generators

$$\begin{aligned} R(\phi, \mathbf{u}) &= e^{-i\phi \mathbf{u}\cdot\mathbf{J}}, \\ B(\lambda, \mathbf{e}) &= e^{-i\lambda \mathbf{e}\cdot\mathbf{M}}, \end{aligned} \tag{4.5}$$

where $\mathbf{J} = (J_1, J_2, J_3)$ and $\mathbf{M} = (M_1, M_2, M_3)$ are the generators of rotations and boost, respectively. They satisfy the algebra[8]

$$\begin{aligned} [J_i, J_j] &= i\epsilon_{ijk} J_k, \\ [J_i, M_j] &= i\epsilon_{ijk} M_k, \\ [M_i, M_j] &= -i\epsilon_{ijk} J_k. \end{aligned} \tag{4.6}$$

Although the calculation leading to them is relatively easy, the previous commutation relations can also be heuristically understood. The first commutator reproduces the usual algebra of infinitesimal rotations familiar from elementary quantum mechanics. The second one is the simple statement that the generators of the boost along the three spatial directions transform as vectors under three-dimensional rotations. The

[8] The six generators $(J_i, M_i)$ of the proper Lorentz group can be fit into a rank-2 antisymmetric tensor with components $\mathscr{J}_{0i} = M_i$ and $\mathscr{J}_{ij} = \epsilon_{ijk} J_k$, satisfying the algebra $[\mathscr{J}_{\mu\nu}, \mathscr{J}_{\alpha\beta}] = i\eta_{\mu\alpha}\mathscr{J}_{\nu\beta} - i\eta_{\mu\beta}\mathscr{J}_{\nu\alpha} + i\eta_{\nu\beta}\mathscr{J}_{\mu\alpha} - i\eta_{\nu\alpha}\mathscr{J}_{\mu\beta}$.

third identity is the less obvious. It amounts to saying that if we carry out two boosts along the directions set by unit vectors $\mathbf{e}_1$ and $\mathbf{e}_2$, the ambiguity in the order of the boost is equivalent to a three-dimensional rotation with respect to the axis defined by $\mathbf{e}_1 \times \mathbf{e}_2$.

We could now try to find irreducible representations of the algebra (4.6). Life gets simpler if we relate this algebra to the one of a group we are more familiar with. This can be done in this case by introducing the new set of generators

$$J_i^{\pm} = \frac{1}{2}\big(J_i \pm iM_i\big), \tag{4.7}$$

in terms of which, the algebra (4.6) reads

$$\begin{aligned} [J_i^+, J_j^+] &= i\epsilon_{ijk}J_k^+, \\ [J_i^-, J_j^-] &= i\epsilon_{ijk}J_k^-, \\ [J_i^+, J_j^-] &= 0. \end{aligned} \tag{4.8}$$

One thing we gain with this is that we have decoupled an algebra of six generators into two algebras of three generators each commuting with one another. But the real bonus here is that the individual algebras are those of SU(2), whose representation theory can be found in any quantum mechanics group. Thus, SO(1,3) $=$ SU(2)$_+ \times$ SU(2)$_-$ and its irreps are obtained by providing a pair of irreps of SU(2), labeled by their total spins $(\mathbf{s}_+, \mathbf{s}_-)$, with $\mathbf{s}_\pm = \mathbf{0}, \frac{\mathbf{1}}{\mathbf{2}}, \mathbf{1}, \frac{\mathbf{3}}{\mathbf{2}}, \ldots$ Since $J_i$ is a pseudovector, it does not change under parity transformations, whereas the boost generators $M_i$ do reverse sign

$$\mathrm{P} : J_i \longrightarrow J_i, \qquad \mathrm{P} : M_i \longrightarrow -M_i. \tag{4.9}$$

As a consequence, parity interchanges the two SU(2) factors

$$\mathrm{P} : (\mathbf{s}_+, \mathbf{s}_-) \longrightarrow (\mathbf{s}_-, \mathbf{s}_+). \tag{4.10}$$

Finally, the generators of the group SO(3) $\approx$ SU(2) of spatial rotations are given by

$$J_i = J_i^+ + J_i^-, \tag{4.11}$$

so the irrep $(\mathbf{s}_+, \mathbf{s}_-)$ decomposes into those of SU(2) with $j = \mathbf{s}_+ + \mathbf{s}_-, \mathbf{s}_+ + \mathbf{s}_- - 1, \ldots, |\mathbf{s}_+ - \mathbf{s}_-|$.

Let us illustrate this general analysis with some relevant examples. We begin with the trivial irrep $(\mathbf{s}_+, \mathbf{s}_-) = (\mathbf{0}, \mathbf{0})$, whose generators are $J_i^\pm = 0$. Fields transforming in this representation are scalar, which under a Lorentz transformation $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ change according to

$$\varphi'(x') = \varphi(x). \tag{4.12}$$

Another parity invariant representation is $(\mathbf{s}_+, \mathbf{s}_-) = (\frac{\mathbf{1}}{\mathbf{2}}, \frac{\mathbf{1}}{\mathbf{2}})$, with generators $J_i^+ = J_i^- = \frac{1}{2}\sigma^i$. Decomposing this irrep with respect to those of spatial rotations, we see that they include a scalar ($j = 0$) and a three-vector ($j = 1$). These correspond respectively to the zero and spatial components of a spin-one

vector field $V^\mu(x)$ transforming as

$$V^\mu(x') = \Lambda^\mu{}_\nu V^\nu(x). \tag{4.13}$$

Finally, we look at $(\mathbf{s}_+, \mathbf{s}_-) = (\mathbf{1}, \mathbf{1})$. This is decomposed in terms of three irreps of SU(2) $\approx$ SO(3) with $j = 2, 1, 0$. Together, they build a rank-two symmetric-traceless tensor field $h^{\mu\nu}(x) = h^{\nu\mu}(x)$, $\eta_{\mu\nu}h^{\mu\nu}(x) = 0$, transforming as

$$h'^{\mu\nu}(x') = \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta h^{\alpha\beta}(x), \tag{4.14}$$

the three irreps of SU(2) corresponding respectively to $h^{ij} - \frac{1}{3}\delta^{ij}h^{00}$, $h^{0i} = h^{i0}$, and $h^{00}$. This is a spin-two field like the one used to describe a graviton.

We look next at parity-violating representations, starting with $(\mathbf{s}_+, \mathbf{s}_-) = (\frac{\mathbf{1}}{\mathbf{2}}, \mathbf{0})$. Its generators are

$$J_k^+ = \frac{1}{2}\sigma^k, \qquad J_k^- = 0. \tag{4.15}$$

Hence, objects transforming in this representation have two complex components changing under rotations and boost according to

$$\chi_+ \longrightarrow e^{-\frac{i}{2}(\phi\mathbf{u} - i\boldsymbol{\lambda})\cdot\boldsymbol{\sigma}}\chi_+, \tag{4.16}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ is the boost's rapidity. In particular, we see that $\chi_+$ transforms as a SO(3) spinor. A field transforming in this representation is a *positive helicity* Weyl spinor. Very soon we will learn the reason for its name.

### 4.2 Chiral (and also nonchiral) fermions

After all these group-theoretical considerations, it is time to start thinking about physics. To construct an action principle for Weyl spinors, we need to build Lorentz invariant quantities from these fields. To begin with, we notice that the Hermitian conjugate spinor $u_+^\dagger$ also transforms in the $(\frac{\mathbf{1}}{\mathbf{2}}, \mathbf{0})$ representation of the Lorentz group, since the representations of SU(2) are real. A general bilinear $\chi_+^\dagger A \chi_+$, on the other hand, transforms under the group SO(3) $\approx$ SU(2) of three-dimensional rotations in the product representation $\frac{\mathbf{1}}{\mathbf{2}} \otimes \frac{\mathbf{1}}{\mathbf{2}} = \mathbf{1} \otimes \mathbf{0}$. Computing the appropriate Clebsh–Gordan coefficients, we find

$$\begin{aligned} \chi_+^\dagger \chi_+ \quad &\Longrightarrow \quad j = 0, \\ \chi_+^\dagger \sigma^i \chi_x \quad &\Longrightarrow \quad j = 1. \end{aligned} \tag{4.17}$$

They represent the time and spatial components of a four-vector

$$\chi_+^\dagger \sigma_+^\mu \chi_+, \tag{4.18}$$

where $\sigma_+^\mu \equiv (\mathbb{1}, \sigma^i)$. With this, we construct an action for the Weyl field as

$$S_+ = \int d^4x\, i\chi_+^\dagger \sigma_+^\mu \partial_\mu \chi_+. \tag{4.19}$$

Notice that although $\chi_+^\dagger \chi_+$ is invariant under rotations it does transform under boosts. Therefore it is not a Lorentz scalar and cannot be added to the action as a mass term.

As for the $(\mathbf{s}_+, \mathbf{s}_-) = (\mathbf{0}, \frac{\mathbf{1}}{\mathbf{2}})$ irrep of SO(1,3), a *negative helicity* Weyl spinor, the analysis is similar to the one just presented and the corresponding expressions are obtained from the ones derived above by applying a parity transformation. In particular, we find its transformations under rotations and boosts to be

$$\chi_- \longrightarrow e^{-\frac{i}{2}(\phi \mathbf{u} + i\boldsymbol{\lambda})\cdot \boldsymbol{\sigma}} \chi_-, \tag{4.20}$$

showing that they also transform as SO(3) spinors. Their free dynamics is derived from the action

$$S_- = \int d^4x\, i\chi_-^\dagger \sigma_-^\mu \partial_\mu \chi_-, \tag{4.21}$$

where $\sigma_-^\mu \equiv (\mathbb{1}, -\sigma^i)$.

Let us analyze in some more detail the physics of Weyl spinor fields. The equations of motion derived from the actions (4.19) and (4.21) are

$$i\sigma_\pm^\mu \partial_\mu \chi_\pm = 0 \qquad \Longrightarrow \qquad \big(\partial_0 \mp \boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\big)\chi_\pm = 0. \tag{4.22}$$

As in other cases, we search for positive energy ($k^0 > 0$) plane wave solutions of the form

$$\chi_\pm(x) \sim u_\pm(\mathbf{k}) e^{-ik\cdot x}, \tag{4.23}$$

where $u_\pm(\mathbf{k})$ are $(\frac{\mathbf{1}}{\mathbf{2}}, \mathbf{0})$ and $(\mathbf{0}, \frac{\mathbf{1}}{\mathbf{2}})$ spinors normalized according to

$$u_\pm(\mathbf{k})^\dagger \sigma_\pm^\mu u_\pm(\mathbf{k}) = 2k^\mu \mathbb{1}. \tag{4.24}$$

Using this Ansatz, the wave equations (4.22) then take the form

$$\big(k_0 \mp \mathbf{k}\cdot\boldsymbol{\sigma}\big) u_\pm(\mathbf{k}) = 0. \tag{4.25}$$

Multiplying by $k_0 \pm \mathbf{k}\cdot\boldsymbol{\sigma}$ on the left and using $k_i k_j \sigma^i \sigma^j = \mathbf{k}^2 \mathbb{1}$, we obtain the dispersion relation of a massless particle, $k_0 = |\mathbf{k}|$. Equation (4.25) implies the condition

$$\left(\mathbb{1} \mp \frac{\mathbf{k}}{|\mathbf{k}|}\cdot\boldsymbol{\sigma}\right) u_\pm(\mathbf{k}) = 0 \qquad \Longrightarrow \qquad \left(\frac{\mathbf{k}}{|\mathbf{k}|}\cdot\mathbf{s}\right) u_\pm(\mathbf{k}) = \pm\frac{1}{2} u_\pm(\mathbf{k}), \tag{4.26}$$

where $\mathbf{s} \equiv \frac{1}{2}\boldsymbol{\sigma}$ is the spin operator. Helicity is defined as the projection of the particle's spin on its direction of motion and the previous identity shows that $u_\pm(k)$ are spinors with positive and negative helicity, respectively. Since the generic Weyl spinors $\chi_\pm$ can be written as a superposition of the plane

wave solutions (4.23), this explains the terminology introduced above.

To write a general positive (resp. negatively) helicity Weyl spinor, we also need to consider negative energy plane waves $v_\pm(\mathbf{k})e^{-ik\cdot x}$, where $k^0 < 0$. Imposing this to solve Eq. (4.22), we find that $v_\pm(\mathbf{k})$ satisfies

$$\big(k^0 \pm \mathbf{k}\cdot\boldsymbol{\sigma}\big)v_\pm(\mathbf{k}) = 0, \tag{4.27}$$

where we set the normalization

$$v_\pm(\mathbf{k})^\dagger \sigma^\mu_\pm v_\pm(\mathbf{k}) = 2k^\mu \mathbb{1}. \tag{4.28}$$

In addition, it can also be shown that the positive and negative energy solutions satisfy the orthogonality relations

$$u(-\mathbf{k})^\dagger v(\mathbf{k}) = v(-\mathbf{k})^\dagger u(\mathbf{k}) = 0. \tag{4.29}$$

These identities will be important later in determining the spectrum of excitations of the free quantum Weyl spinor field.

Classical Weyl spinors are complex fields and their actions (4.19) and (4.21) are invariant under global phase rotations $\chi_\pm \longrightarrow e^{i\vartheta}\chi_\pm$. The associated Noether currents (see page 57) are the bilinear Lorentz vector constructed in Eq. (4.18), and the corresponding expression for negative helicity,

$$j^\mu_\pm = \chi^\dagger_\pm \sigma^\mu_\pm \chi_\pm. \tag{4.30}$$

Plugging this current into Eq. (3.51) we couple the Weyl spinors to the electromagnetic field

$$\begin{aligned} S_\pm &= \int d^4x \left( i\chi^\dagger_\pm \sigma^\mu_\pm \partial_\mu \chi_\pm + e\chi_\pm \sigma^\mu_\pm \chi_\pm A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\right) \\ &= \int d^4x \left[ i\chi^\dagger_\pm \sigma^\mu_\pm \big(\partial_\mu - ieA_\mu\big)\chi_\pm - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\right], \end{aligned} \tag{4.31}$$

where in the second line we find again the gauge covariant derivative first introduced in Eq. (3.107). This action is invariant under gauge transformations, acting on the Weyl spinor by *local* phase rotations $\chi_\pm \longrightarrow e^{ie\epsilon(x)}\chi_\pm$. Moreover, given the absence of any dimensionful parameter in the action, we can expect the classical theory to be scale invariant. This is indeed the case, with the Weyl spinors having scaling dimension $\Delta_\chi = \frac{3}{2}$.

To quantize the Weyl field, we begin with the computation of the canonical Poisson algebra. The momentum canonically conjugate to the spinor is given by

$$\pi_\pm \equiv \frac{\delta S_\pm}{\delta \partial_0 \chi_\pm} = i\chi^\dagger_\pm, \tag{4.32}$$

leading to

$$\{\chi_{\pm,a}(t,\mathbf{r}),\chi_{\pm,b}(t,\mathbf{r}')^\dagger\}_{\rm PB}=-i\delta_{ab}\delta^{(3)}(\mathbf{r}-\mathbf{r}'), \tag{4.33}$$

where $a,b$ denote the spinor indices and all other Poisson brackets are equal to zero. The Hamiltonian then reads

$$H_\pm=\pm i\int d^3x\,\chi_\pm^\dagger(\boldsymbol{\sigma}\cdot\boldsymbol{\nabla})\chi_\pm. \tag{4.34}$$

So much for the classical theory. Quantum Weyl spinor fields are written as operator-valued superpositions of positive- and negative-energy plane wave solutions

$$\widehat{\chi}_\pm(t,\mathbf{r})=\int\frac{d^3k}{(2\pi)^3}\frac{1}{2|\mathbf{k}|}\Big[\widehat{b}(\mathbf{k},\pm)u_\pm(\mathbf{k})e^{-i|\mathbf{k}|t+i\mathbf{k}\cdot\mathbf{r}}+\widehat{d}(\mathbf{k},\pm)^\dagger v_\pm(\mathbf{k})^* e^{i|\mathbf{k}|t-i\mathbf{k}\cdot\mathbf{r}}\Big]. \tag{4.35}$$

It is important to remember that the previous operator is not Hermitian. Similarly to what we learned from the analysis of the complex scalar field, this implies that the operators $\widehat{b}(\mathbf{k},\pm)$ and $\widehat{d}(\mathbf{k},\pm)$ are independent and unrelated to each other by Hermitian conjugation. However, we need to be careful when constructing the algebra of field operators. For example, the spin-statistics theorem states that particles with half-integer spin are fermions, and their quantum states should be antisymmetric under the interchange of two of them. To achieve this, the prescription (3.66) has to be modified and Poisson brackets are replaced by *anticommutators* instead of commutators

$$i\{\cdot,\cdot\}_{\rm PB}\longrightarrow\{\cdot,\cdot\}. \tag{4.36}$$

Accordingly, we impose

$$\{\widehat{\chi}_{\pm,a}(t,\mathbf{r}),\widehat{\chi}_{\pm,b}(t,\mathbf{r}')^\dagger\}_{\rm PB}=\delta_{ab}\delta^{(3)}(\mathbf{r}-\mathbf{r}'), \tag{4.37}$$

which, using the normalization $u_\pm(\mathbf{k})^\dagger u_\pm(\mathbf{k})=2|\mathbf{k}|$ [cf. (4.24)], leads to the operator algebra

$$\begin{aligned}\{\widehat{b}(\mathbf{k},\pm),\widehat{b}(\mathbf{k}',\pm)^\dagger\}&=(2\pi)^3 2|\mathbf{k}|\delta_{ab}\delta^{(3)}(\mathbf{r}-\mathbf{r}'),\\ \{\widehat{d}(\mathbf{k},\pm),\widehat{d}(\mathbf{k}',\pm)^\dagger\}&=(2\pi)^3 2|\mathbf{k}|\delta_{ab}\delta^{(3)}(\mathbf{r}-\mathbf{r}'),\end{aligned} \tag{4.38}$$

with all remaining anticommutators equal to zero. As in the case of the complex scalar field analyzed in Box 6, here we also get two types of particles generated by the two kinds of creation operators acting on the vacuum

$$|\mathbf{k}_1,\ldots,\mathbf{k}_n;\mathbf{p}_1,\ldots,\mathbf{p}_m\rangle_\pm=\widehat{b}(\mathbf{k}_1,\pm)^\dagger\ldots\widehat{b}(\mathbf{k}_n,\pm)^\dagger\widehat{d}(\mathbf{p}_1,\pm)^\dagger\ldots\widehat{d}(\mathbf{p}_m,\pm)^\dagger|0\rangle. \tag{4.39}$$

As expected, the state is antisymmetric under the interchange of two particles of the same type, due to the anticommutation of the creation operators. Similarly to the complex scalar field, the two types of

particles are distinguished by the charge operator defined by the conserved current (4.30),

$$\widehat{Q} = \int d^3\mathbf{r}\, \widehat{\chi}_\pm(t,\mathbf{r})^\dagger \widehat{\chi}_\pm(t,\mathbf{r}) \qquad \Longrightarrow \qquad \begin{cases} \widehat{Q}|\mathbf{k};0\rangle_\pm = |\mathbf{k};0\rangle_\pm \\ \widehat{Q}|0;\mathbf{k}\rangle_\pm = -|0;\mathbf{k}\rangle_\pm \end{cases}, \tag{4.40}$$

so the states $|0;\mathbf{k}\rangle_\pm$ are naturally identified as the antiparticles of $|\mathbf{k};0\rangle_\mp$.

The calculation of the Hamiltonian operator follows the lines outlined in previous cases. Replacing classical fields by operators in the Hamiltonian (4.34), and using the properties of the positive and negative energy solutions $u(\mathbf{k})$ and $v(\mathbf{k})$, we find after some algebra

$$\widehat{H}_\pm = \int \frac{d^3k}{(2\pi)^3} \frac{1}{2|\mathbf{k}|} \Big[ |\mathbf{k}| \widehat{b}(\mathbf{k},\pm)^\dagger \widehat{b}(\mathbf{k},\pm) + |\mathbf{k}| \widehat{d}(\mathbf{k},\pm)^\dagger \widehat{d}(\mathbf{k},\pm) \Big] - \int d^3k\, |\mathbf{k}| \delta^{(3)}(\mathbf{0}). \tag{4.41}$$

We see from the first term on the right-hand side that the multiparticle states (4.39) diagonalize the Hamiltonian, with particles and antiparticles having zero mass, $E_\mathbf{k} = |\mathbf{k}|$. In this Hamiltonian we find once more the UV and IR divergent zero-point contribution, that once regularized gives a vacuum energy density

$$\rho_{\text{vac}} = -\frac{1}{8\pi^2} \Lambda_{\text{UV}}^4. \tag{4.42}$$

Although it will eventually be subtracted, it is worthwhile to stop a moment and compare this with the expression (3.76). A first thing meeting the eye is the relative factor of two in the Weyl spinor case. This reflects that while a real scalar field has a single propagating degree of freedom, here we have two, associated with the complex field's real and imaginary parts. The second and physically very relevant feature is the different sign, boiling down to having anticommutators rather than commutators. It implies that bosons and fermions contribute to the vacuum energy with opposite signs. This is the reason why supersymmetric theories, which have as many bosonic as fermionic degrees of freedom and therefore zero vacuum energy, have been invoked to solve the problem of the cosmological constant mentioned in page 35, or at least to ameliorate it[9].

**Box 7. Dirac spinors**

Although the theory of a single Weyl spinor violates parity, it is possible to construct a parity-invariant theory by taking together two Weyl spinors with opposite chiralities. They can be combined into a single object, a Dirac spinor

$$\psi \equiv \begin{pmatrix} \chi_+ \\ \chi_- \end{pmatrix}, \tag{4.43}$$

which obviously transforms in the parity-invariant reducible representation $(\mathbf{\frac{1}{2}},\mathbf{0}) \oplus (\mathbf{0},\mathbf{\frac{1}{2}})$. The corresponding free action is obtained by adding the ones already written in eqs. (4.19) and (4.19)

[9] Since supersymmetry must be broken at low energies (after all, we do not "see" the same number of bosons as fermions), there is still a nonvanishing contribution to the vacuum energy proportional to the fourth power of the scale of supersymmetry breaking, $\Lambda_{\text{SUSY}}$, rather than the much higher $\Lambda_{\text{Pl}}$.

for Weyl spinors of different chiralities, namely

$$S = \int d^4x \left( i\chi_+^\dagger \sigma_+^\mu \partial_\mu \chi_+ + i\chi_-^\dagger \sigma_-^\mu \chi_- \right) = i \int d^4x \, \psi^\dagger \begin{pmatrix} \sigma_+^\mu & 0 \\ 0 & \sigma_-^\mu \end{pmatrix} \partial_\mu \psi. \tag{4.44}$$

An important point to be taken into account now is that $u_\pm$ and $u_\pm^*$ do have opposite helicities. This is the reason why $u_\pm^\dagger \sigma_\pm^\mu u_\pm \equiv u_{\pm,a}^* (\sigma_\pm^\mu)_{ab} u_{\pm,b}$ defines a Lorentz vector, since $(\mathbf{\frac{1}{2}}, \mathbf{0}) \otimes (\mathbf{0}, \mathbf{\frac{1}{2}}) = (\mathbf{\frac{1}{2}}, \mathbf{\frac{1}{2}})$ and $(\sigma_\pm^\mu)_{ab}$ are the Clebsh–Gordan coefficients decomposing the product representation into its irreps. As a consequence, whereas $\psi^*$ does not transform in the same representation as $\psi$, the spinor

$$\overline{\psi}^T \equiv \begin{pmatrix} u_-^* \\ u_+^* \end{pmatrix} = \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix} \psi^* \tag{4.45}$$

does. This suggests recasting the action (4.44) as

$$S = i \int d^4x \, \overline{\psi} \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix} \begin{pmatrix} \sigma_+^\mu & 0 \\ 0 & \sigma_-^\mu \end{pmatrix} \partial_\mu \psi = i \int d^4x \, \overline{\psi} \begin{pmatrix} 0 & \sigma_-^\mu \\ \sigma_+^\mu & 0 \end{pmatrix} \partial_\mu \psi, \tag{4.46}$$

It seems natural to introduce a new set of $4 \times 4$ matrices, the *Dirac matrices*, defined by

$$\gamma^\mu \equiv \begin{pmatrix} 0 & \sigma_-^\mu \\ \sigma_+^\mu & 0 \end{pmatrix}, \tag{4.47}$$

and satisfying the Clifford algebra

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} \mathbb{1}, \tag{4.48}$$

as can be easily checked using the anticommutation relations of the Pauli matrices. The generators of the representation of $(\mathbf{\frac{1}{2}}, \mathbf{0}) \oplus (\mathbf{0}, \mathbf{\frac{1}{2}})$ are then given in terms of the Dirac matrices by (see the footnote in page 42)

$$\mathscr{J}^{\mu\nu} = -\frac{i}{4}[\gamma^\mu, \gamma^\nu] \equiv \sigma^{\mu\nu}. \tag{4.49}$$

Denoting by $\mathcal{U}(\Lambda)$ the matrix implementing the Lorentz transformation $\Lambda^\mu{}_\nu$ on Dirac spinors and using the property $\gamma^{\mu\dagger} = \gamma^0 \gamma^\mu \gamma^0$, it is easy to show that $\mathcal{U}(\Lambda)^\dagger = \gamma^0 \mathcal{U}(\Lambda)^{-1} \gamma^0$. This implies that, while $\psi \to \mathcal{U}(\Lambda)\psi$, the conjugate spinor transforms contravariantly, $\overline{\psi} \to \overline{\psi}\mathcal{U}(\Lambda)^{-1}$, and the Dirac matrices themselves satisfy $\mathcal{U}(\Lambda)^{-1} \gamma^\mu \mathcal{U}(\Lambda) = \Lambda^\mu{}_\nu \gamma^\nu$. Let this serve as *a posteriori* justification of the introduction of the conjugate field $\overline{\psi}$.

The previous discussion shows that $\overline{\psi}\psi$ is a Lorentz scalar that can be added to the Dirac action (4.46), that we now write in a much more compact form

$$S = \int d^4x \left( i\overline{\psi}\gamma^\mu \partial_\mu \psi - m\overline{\psi}\psi \right). \tag{4.50}$$

The associated field equations admit positive energy plane wave solutions of the form $\psi(x) \sim u(\mathbf{k},s)e^{-ik\cdot x}$, with $s=\pm\frac{1}{2}$ labelling the two possible values of the spin third component

$$\big(i\gamma^\mu\partial_\mu - m\big)\psi(x)=0 \qquad \Longrightarrow \qquad \big(\not{k}-m\big)u(\mathbf{k},s)=0. \tag{4.51}$$

Here we have introduced the Feynman slash notation $\not{a}\equiv\gamma^\mu a_\mu$ that we will use throughout these lectures. Acting on the equation to the right of (4.51) with $\not{k}+m$ and implementing the identity $\not{k}\not{k}=k^2\mathbb{1}$, we find the massive dispersion relation $k^0\equiv E_{\mathbf{k}}=\sqrt{\mathbf{k}^2+m^2}$.

To get a better idea about the role played by the mass term in the Dirac equation, it is instructive to write the equation $\big(\not{k}-m\big)u(\mathbf{k},s)=0$ in terms of the two helicity components of the Dirac spinor

$$\begin{aligned}\big(E_{\mathbf{k}}\mathbb{1}-\mathbf{k}\cdot\boldsymbol{\sigma}\big)u_+(\mathbf{k},s)&=mu_-(\mathbf{k},s),\\ \big(E_{\mathbf{k}}\mathbb{1}+\mathbf{k}\cdot\boldsymbol{\sigma}\big)u_-(\mathbf{k},s)&=mu_+(\mathbf{k},s).\end{aligned} \tag{4.52}$$

These expressions show that the mass terms mix the two helicities. Introducing the chirality matrix

$$\gamma_5\equiv -i\gamma^0\gamma^1\gamma^2\gamma^3=\begin{pmatrix}\mathbb{1} & 0\\ 0 & -\mathbb{1}\end{pmatrix}, \tag{4.53}$$

the previous identity is recast as

$$\begin{pmatrix}\frac{\mathbf{k}}{|\mathbf{k}|}\cdot\mathbf{s} & 0\\ 0 & \frac{\mathbf{k}}{|\mathbf{k}|}\cdot\mathbf{s}\end{pmatrix}u(\mathbf{k},s)=\frac{1}{2}\left(\frac{E_{\mathbf{k}}}{|\mathbf{k}|}\mathbb{1}-\frac{m}{|\mathbf{k}|}\gamma^0\right)\gamma_5 u(\mathbf{k},s), \tag{4.54}$$

with $\mathbf{s}=\frac{1}{2}\boldsymbol{\sigma}$ the spin, so the matrix on the left-hand side of this expression is the helicity operator $h$ acting on a four-component Dirac spinor.

The chirality matrix satisfies $\gamma_5^2=\mathbb{1}$ and anticommutes with all Dirac matrices, $\{\gamma_5,\gamma^\mu\}=0$. As a consequence, its commutator with the Lorentz generators vanishes, $[\gamma_5,\sigma^{\mu\nu}]=0$, and by Schur's lemma this means that the spinors $P_+\psi$ and $P_-\psi$ transform in different irreps of the Lorentz group, with $P_\pm=\frac{1}{2}(\mathbb{1}\pm\gamma_5)$ the projector onto the two chiralities. The spinor's chirality is therefore a Lorentz invariant.

A look at Eq. (4.54) shows that for a *massive* Dirac, spinor helicity (the projection of the spin onto the direction of motion) and chirality (the eigenvalue of the chirality matrix) are very different things. The former is not even a Lorentz invariant, since for a massive fermion with positive/negative helicity we can switch to a moving frame overcoming the particle and make the helicity negative/positive. Taking, however, the massless limit $m\to 0$ we have $E_{\mathbf{k}}\to|\mathbf{k}|$ and chirality and helicity turn out to be equivalent

$$h=\frac{1}{2}\gamma_5 \qquad (m=0). \tag{4.55}$$

This is why, when dealing with massless spin-$\frac{1}{2}$ fermions, both terms can be used indistinctly, although in the case of massive particles one should be very careful in using the one appropriate to the physical situation under analysis.

To quantize the theory, we write an expansion of the Dirac field operator into its positive and negative energy solutions

$$\widehat{\psi}(t,\mathbf{r}) = \sum_{s=\pm\frac{1}{2}} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2E_{\mathbf{k}}} \left[ \widehat{b}(\mathbf{k},s) u(\mathbf{k},s) e^{-i|\mathbf{k}|t+i\mathbf{k}\cdot\mathbf{r}} + \widehat{d}(\mathbf{k},s)^\dagger v(\mathbf{k},s)^* e^{i|\mathbf{k}|t-i\mathbf{k}\cdot\mathbf{r}} \right], \quad (4.56)$$

where the negative energy solutions $v(\mathbf{k},s)$ are defined by the equation $(\not{k}+m)v(\mathbf{k},s) = 0$. The canonical anticommutation relations of the Dirac field with its Hermitian conjugate imply that $\widehat{b}(\mathbf{k},s)$ and $\widehat{b}(\mathbf{k},s)^\dagger$ are a system of fermionic creation–annihilation operators for particles, while $\widehat{d}(\mathbf{k},s)$ and $\widehat{d}(\mathbf{k},s)^\dagger$ respectively annihilate and create antiparticles out of the vacuum. The multiparticle states obtained by acting with creation operators on the Fock vacuum are eigenstates of the Dirac Hamiltonian, with the elementary excitations $\widehat{b}(\mathbf{k},s)^\dagger|0\rangle$ and $\widehat{d}(\mathbf{k},s)^\dagger|0\rangle$ representing spin $\frac{1}{2}$ particles (resp. antiparticles) of momentum $\mathbf{k}$, energy $E_{\mathbf{k}} = \sqrt{\mathbf{k}^2+m^2}$, and spin third component $s$. The details of this analysis are similar to the ones presented above for Weyl fermions and can be found in any of the QFT textbooks listed in the references.

Finally, let us mention that Dirac spinors can be coupled to the electromagnetic field as we did in Eq. (4.31) for the Weyl spinors. The Dirac action (4.50) is invariant under a global phase rotation of the spinor, $\psi \to e^{i\alpha}\psi$, leading to the existence of a conserved current due to the first Noether theorem (see page 57)

$$j^\mu = \overline{\psi}\gamma^\mu\psi. \quad (4.57)$$

We can use this conserved current to couple fermions to the electromagnetic field and write the QED action

$$\begin{aligned} S &= \int d^4x \left[ -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \overline{\psi}(i\not{\partial} - m)\psi + eA_\mu\overline{\psi}\gamma^\mu\psi \right] \\ &= \int d^4x \left[ -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \overline{\psi}(i\not{D} - m)\psi \right], \end{aligned} \quad (4.58)$$

where once again we encounter the covariant derivative $D_\mu = \partial_\mu - ieA_\mu$ and the slash notation introduced in Eq. (4.51) is used. This action describes the interaction of spinors with the electromagnetic field, that upon quantization is called quantum electrodynamics (QED). It is an interacting theory of charged particles (e.g., electrons) and photons that, unlike the free theories we have been dealing with so far, cannot be exactly solved. One particularly effective way to extract physical information is perturbation theory. This assumes that the coupling is sufficiently weak, so that physics can be reliably described in terms of the interaction among the excitations of the free theory.

Before closing our discussion of the irreps of the Lorentz group, let us mention some more relevant examples. The representations $(\mathbf{s}_+,\mathbf{s}_-) = (\mathbf{1},\mathbf{0})$ and $(\mathbf{s}_+,\mathbf{s}_-) = (\mathbf{0},\mathbf{1})$ correspond to rank-2

| Representation | Field | Parity |
|---|---|---|
| $(\mathbf{0},\mathbf{0})$ | Scalar | ✓ |
| $(\frac{\mathbf{1}}{\mathbf{2}},\mathbf{0})$ | Positive helicity Weyl spinor | × |
| $(\mathbf{0},\frac{\mathbf{1}}{\mathbf{2}})$ | Negative helicity Weyl spinor | × |
| $(\frac{\mathbf{1}}{\mathbf{2}},\frac{\mathbf{1}}{\mathbf{2}})$ | Vector | ✓ |
| $(\frac{\mathbf{1}}{\mathbf{2}},\mathbf{0})\oplus(\mathbf{0},\frac{\mathbf{1}}{\mathbf{2}})$ | Dirac spinor | ✓ |
| $(\mathbf{1},\mathbf{0})$ | Self-dual rank-2 antisymmetric tensor | × |
| $(\mathbf{0},\mathbf{1})$ | Anti-self-dual rank-2 antisymmetric tensor | × |
| $(\mathbf{1},\mathbf{0})\oplus(\mathbf{0},\mathbf{1})$ | Antisymmetric rank-2 tensor | ✓ |
| $(\mathbf{1},\mathbf{1})$ | Symmetric-traceless rank-2 tensor | ✓ |

**Table 1:** Summary of some relevant representations of the Lorentz group and their parity properties.

antisymmetric tensor fields $B_{\mu\nu} = B_{[\mu\nu]}$ respectively satisfying self-dual $(+)$ and anti-self-dual $(-)$ conditions

$$B_{\mu\nu} = \pm\frac{1}{2}\epsilon_{\mu\nu\alpha\beta}B^{\alpha\beta}. \tag{4.59}$$

An example of the $(\mathbf{1},\mathbf{0})$ and $(\mathbf{0},\mathbf{1})$ irreps are the complex combinations $\mathbf{E}\pm i\mathbf{B}$ that we encountered in our discussion of electric–magnetic duality in page 24. The two irreps can be added to form the parity-invariant reducible representation $(\mathbf{1},\mathbf{0})\oplus(\mathbf{0},\mathbf{1})$, corresponding to a generic rank-2 antisymmetric tensor field such as the electromagnetic field strength[10].

Finally, multiplying together two vector representations we have

$$\left(\frac{\mathbf{1}}{\mathbf{2}},\frac{\mathbf{1}}{\mathbf{2}}\right)\otimes\left(\frac{\mathbf{1}}{\mathbf{2}},\frac{\mathbf{1}}{\mathbf{2}}\right) = (\mathbf{1},\mathbf{1})\oplus\big[(\mathbf{1},\mathbf{0})\oplus(\mathbf{0},\mathbf{1})\big]\oplus(\mathbf{0},\mathbf{0}). \tag{4.60}$$

This is just group theory lingo to express the decomposition of the product $V_\mu W_\nu$ of two four-vectors into its symmetric-traceless, antisymmetric, and trace pieces

$$V_\mu W_\nu = \left(V_{(\mu}W_{\nu)} - \frac{1}{4}\eta_{\mu\nu}V_\alpha W^\alpha\right) + V_{[\mu}W_{\nu]} + \frac{1}{4}\eta_{\mu\nu}V_\alpha W^\alpha. \tag{4.61}$$

This leads to identify the $(\mathbf{1},\mathbf{1})$ irrep as corresponding to a symmetric-traceless rank-2 tensor field. For the reader's benefit, we have summarized in Table 1 the different representations of the Lorentz group discussed in this section, indicating as well whether or not they preserve parity.

[10] Rank-2 antisymmetric tensor fields are ubiquitous in string theories, including those satisfying the (anti-)self-dual condition (4.59).

### 4.3 Some more group theory

Having got some practice with the language of group theory, we close this section by enlarging our vocabulary with many important group-theoretic concepts that will become handy later on (see Refs. [68, 69] for some physics oriented textbooks on group theory, or Appendix B of Ref. [14] for a quick survey of basic facts). Next, we focus on the relevant groups for the SM, namely SU(3), SU(2), and U(1) associated with the strong and electroweak interactions. We have encountered the Abelian group U(1) when discussing electromagnetism and learned there that it has a single generator, let us call it $Q$, so its elements are written as $U(\vartheta) = e^{i\vartheta Q}$. This is the only irrep of this group, all others being reducible to a diagonal form.

Concerning SU(2), its properties are well know from the theory of angular momentum in quantum mechanics and we have already used many of them in our analysis of the representations of the Lorentz group. Its three generators satisfy the algebra

$$[T^a_{\mathbf{R}}, T^b_{\mathbf{R}}] = i\epsilon^{abc} T^c_{\mathbf{R}}, \tag{4.62}$$

where the subscript $\mathbf{R}$ denotes the representation. Up to this point, we have labelled the irreps of SU(2) by their spin $\mathbf{s} = \mathbf{0}, \frac{\mathbf{1}}{\mathbf{2}}, \mathbf{1}, \ldots$, although they are also frequently referred to by their dimension $2\mathbf{s}+1$, as it is customary for all unitary groups SU($N$). As an example, the fundamental representation $\mathbf{s} = \frac{\mathbf{1}}{\mathbf{2}}$ is denoted by $\mathbf{2}$ and the adjoint $\mathbf{s} = \mathbf{1}$ by $\mathbf{3}$. In the former case the generators are written in terms of the three Pauli matrices as $T^a_{\mathbf{2}} = \frac{1}{2}\sigma_a$, a fact we used when studying Weyl spinors.

As for the group SU(3), less familiar from elementary physics, it has eight generators satisfying the Lie algebra

$$[T^a_{\mathbf{R}}, T^b_{\mathbf{R}}] = i f^{abc} T^c_{\mathbf{R}} \qquad\qquad (a, b, c = 1, \ldots, 8), \tag{4.63}$$

where the structure constants are given by

$$f^{123} = 1, \quad f^{147} = -f^{156} = f^{246} = f^{257} = f^{345} = -f^{367} = \frac{1}{2}, \quad f^{458} = f^{678} = \frac{\sqrt{3}}{2}, \tag{4.64}$$

the remaining ones being either zero or fixed from the ones just given by antisymmetry. The group elements are written as exponentials of linear combinations of the algebra generators

$$U(\alpha)_{\mathbf{R}} = e^{i\alpha^a T^a_{\mathbf{R}}}, \tag{4.65}$$

where the condition $\det U(\alpha)_{\mathbf{R}} = 1$ implies $\operatorname{tr} T^a_{\mathbf{R}} = 0$ and the generators can be chosen to satisfy the orthogonality relations

$$\operatorname{tr}\left(T^a_{\mathbf{R}} T^b_{\mathbf{R}}\right) = T_2(\mathbf{R})\delta^{ab}. \tag{4.66}$$

Although similar in many aspects, there are however important differences between SU(2) and SU(3) concerning the character of their irreps. For any Lie algebra representation with generators $T^a_{\mathbf{R}}$ it is very easy to check that $-T^{a*}_{\mathbf{R}}$ satisfies the same Lie algebra, defining the complex conjugate

representation denoted by $\overline{\mathbf{R}}$. A representation is said to be *real* or *pseudoreal* whenever it is related to its complex conjugate irrep by a similarity transformation

$$T^a_{\overline{\mathbf{R}}} \equiv -T^{a*}_{\mathbf{R}} = S^{-1} T^a_{\mathbf{R}} S, \tag{4.67}$$

with $S$ either symmetric (real representation) or antisymmetric (pseudoreal representation). For SU(2) all irreps are real or pseudoreal. This is the reason why we only have one independent irrep of a given dimension labelled by its spin. The group SU(3), on the other hand, has complex irreps. This is the case of the fundamental and an antifundamental representations, $\mathbf{3}$ and $\overline{\mathbf{3}}$, whose generators are given by

$$T^a_{\mathbf{3}} = \frac{1}{2}\lambda_a \qquad \text{and} \qquad T^a_{\overline{\mathbf{3}}} = -\frac{1}{2}\lambda_a^T, \tag{4.68}$$

where $\lambda_a$ are the eight Gell-Mann matrices, given by

$$\begin{aligned}
\lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\
\lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \\
\lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & -\frac{2}{\sqrt{3}} \end{pmatrix}.
\end{aligned} \tag{4.69}$$

Two instances of the group SU(3) exist in the SM. One is the color gauge symmetry of QCD, which we will study in some detail in later sections. The second is the global SU(3)$_f$ flavor symmetry of the eightfold way, originally formulated by Murray Gell-Mann [70] and Yuval Ne'eman [71]. With the hindsight provided by the quark model, this classification scheme is based on the assumption that the strong nuclear force does not distinguish among different quark flavors[11]. Let us consider the action for three quark flavors $q_i$ ($i = 1, 2, 3$),

$$\begin{aligned}
S &= \sum_{i=u,d,s} \int d^4x\, \overline{q}_i \big(i\partial\!\!\!/ - m_i\big) q_i + S_{\text{int}} \\
&= \int d^4x\, \overline{\boldsymbol{q}} \big(i\partial\!\!\!/ \mathbb{1} - \boldsymbol{m}\big) \boldsymbol{q} + S_{\text{int}},
\end{aligned} \tag{4.70}$$

where $S_{\text{int}}$ represents interaction terms that we will not care about for the time being and in the second line we have grouped the quarks into a triplet $\boldsymbol{q}$ and rewrote the action in matrix notation, with $\boldsymbol{m} =$

[11] Quarks were proposed as hadron constituents in Refs. [72,73], some three years after the formulation of the eightfold way. The name, as with quarks, was invented by Gell-Mann drawing this time not from James Joyce but from the Noble Eightfold Path of Buddhism: Right View, Right Intention, Right Speech, Right Conduct, Right Livelihood, Right Effort, Right Mindfulness, and Right Meditation.

$\mathrm{diag}(m_u, m_d, m_s)$. Under SU(3)$_f$ the quark triplet transforms in the fundamental irrep $\mathbf{3}$ as $\boldsymbol{q} \to U\boldsymbol{q}$. This results in the following transformation of the free action

$$\int d^4x\, \overline{\boldsymbol{q}}\big(i\partial\!\!\!/\mathbb{1} - \boldsymbol{m}\big)\boldsymbol{q} \longrightarrow \int d^4x\, \overline{\boldsymbol{q}}\big(i\partial\!\!\!/\mathbb{1} - U^\dagger \boldsymbol{m} U\big)\boldsymbol{q}, \tag{4.71}$$

where $\boldsymbol{m} = \mathrm{diag}(m_u, m_d, m_s)$. Since all three quark masses are different, $\boldsymbol{m}$ is not proportional to the identity and $U^\dagger \boldsymbol{m} U \neq \boldsymbol{m}$, and the mass term breaks the global SU(3)$_f$ invariance. Moreover, the strong interaction does not distinguish quark flavors and $S_{\mathrm{int}}$ remains invariant. Thus, we conclude that SU(3)$_f$ is an approximate symmetry of QCD that becomes exact in the limit of equal, in particular zero, quark masses (also called, for obvious reasons, the chiral limit).

Mesons are bound states of a quark and an antiquark, the later transforming in the antifundamental $\overline{\mathbf{3}}$ irrep. Their classification into SU(3)$_f$ multiplets follows from decomposing into irreps the product of the fundamental and the antifundamental

$$\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}. \tag{4.72}$$

The octet contains the $\pi^0$, $\pi^\pm$, $K^0$, $\overline{K}^0$, $K^\pm$, and $\eta_8$ mesons, while the singlet is the $\eta_1$ meson. In fact, the $\eta_1$ and $\eta_8$ mesons mix together into the $\eta$ and the $\eta'$ mesons, which are the interaction eigenstates in the electroweak sector of the SM. A similar classification scheme works for the baryons. Being composed of three quarks, the baryon multiplets emerge from decomposing the product of three fundamental representations

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1}. \tag{4.73}$$

The proton and the neutron are in one of the octets, together with the $\Sigma^0$, $\Sigma^\pm$, $\Xi^0$, and $\Xi^-$ particles of nonzero strangeness. Were SU(3)$_f$ an exact symmetry, the masses of all hadrons within a single multiplet would be equal. However, the differences in the quark masses induce a mass split, which in the case of the octet containing the proton and the neutron is about $30\%$ of the average mass. By contrast, the mass split between the proton and the neutron is only $0.1\%$ of their average mass. The wider mass gap with the other octet members results from the larger mass of the strange quark, $m_s > m_u \sim m_d$.

## 5 A tale of many symmetries

Symmetry is probably the most important heuristic principle at our disposal in fundamental physics. The formulation of particle physics models starts with selecting those symmetries/invariances to be implemented in the theory, which usually restrict drastically the types of interactions allowed. In the SM gauge, for example, invariance plus the condition that the action only contains operators of dimension four or less fixes the action, up to a relatively small number of numerical parameters to be experimentally measured in high energy facilities.

### 5.1 The symmetries of physics

Our approach to symmetry up to here has been rather casual. It is time to be more precise, beginning with a discussion of the types of symmetries we encounter in QFT and how they are implemented.

i) **Kinematic (or spacetime) symmetries**. They act on the spacetime coordinates and field indices. This class of symmetries includes Lorentz, Poincaré, scale, and conformal transformations that we already encountered in previous sections.

ii) **Discrete symmetries.** They include parity P, charge conjugation C, time reversal T, and the compositions CP and CPT. If gravity and electromagnetism were the only interactions in nature, the universe would be invariant under C, P, and T separately. However, nuclear (both weak and strong) interactions break P, C, T and CP in different degrees.
CPT, however, turns out to be a symmetry of QFT forced upon us by the basic requirements of Poincaré invariance and locality. Moreover, it is a completely general result that can be demonstrated without relying on the specific form of any Hamiltonian (for a detailed proof of this result, called the CPT theorem, see Chapter 11 of [14]).

iii) **Global continuous symmetries**. These are transformations depending on a continuous constant parameter. One example is the invariance of the complex scalar field action (3.86) under spacetime constant phase rotation (3.98). The current view in QFT is that global symmetries are accidental properties of the low energy theories, whereas, in the UV, all fundamental symmetries should be local (see next).

iv) **Local (gauge) invariance**. Unlike the previous case, the theory is invariant under a set of continuous transformations that vary from point to point in spacetime. The archetypical example is the gauge invariance of the Maxwell's equations found in (3.4). Unlike standard quantum mechanical symmetries, gauge invariance does not map one physical state into another, but represents a redundancy in the labeling of the physical states. This is the price we pay to describe fields with spin one and two in a way that manifestly preserves locality and Lorentz invariance. To highlight this fundamental feature, we will refrain from talking about gauge symmetry and stick to gauge invariance (we will qualify this statement below).

v) **Spontaneously/softly broken symmetries.** In all instances discussed above, we have assumed that the symmetries/invariances are realized at the action level and in the spectrum of the quantum theory. Classically, it is possible that the symmetries of the action are not reflected in their solutions which implies that in the quantum theory, the spectrum does not remain invariant under the symmetry. When this happens, we say that the symmetry (or invariance) is *spontaneously broken*. Since the breaking takes place by the choice of vacuum, it does not affect the UV behavior of the theory. Another situation when this also happens is when adding terms to the action that explicitly break the symmetry but do not modify the UV behavior of the theory (e.g., mass terms). In this case, the symmetry is *softly broken*.

vi) **Anomalous symmetries.** Usually, symmetries are identified in the classical action and then implemented in the quantum theory. This tacitly assumes that all classical symmetries remain after quantization, and this is not always the case. Sometimes, the classical symmetry is impossible to implement quantum mechanically, and it is said to be *anomalous*. Anomalies originate in very

profound mathematical properties of QFT and they have important physical consequences.

Let us see now how symmetries are implemented in QFT. We know from quantum mechanics that symmetries are maps among rays in the theory's Hilbert space that preserve probability amplitudes. More precisely, for two arbitrary states $|\alpha\rangle$ and $|\beta\rangle$, a symmetry is implemented by some operator $U$ acting as

$$|\alpha\rangle \longrightarrow |U\alpha\rangle, \qquad |\beta\rangle \longrightarrow |U\beta\rangle, \tag{5.1}$$

and satisfying the condition that probability amplitudes are preserved

$$|\langle\alpha|\beta\rangle| = |\langle U\alpha|U\beta\rangle|. \tag{5.2}$$

There are two ways in which this last condition can be achieved. One is that

$$\langle\alpha|\beta\rangle = \langle U\alpha|U\beta\rangle, \tag{5.3}$$

implying that the operator $U$ is *unitary*. But there also exists a second alternative to fullfil Eq. (5.2),

$$\langle U\alpha|U\beta\rangle = \langle\alpha|\beta\rangle^*. \tag{5.4}$$

In this case the operator $U$ is said to be *antiunitary*. Notice that consistency requires that in this case the operator $U$ implementing the symmetry should be antilinear:

$$U\big(a|\alpha\rangle + b|\beta\rangle\big) = a^*|U\alpha\rangle + b^*|U\beta\rangle, \tag{5.5}$$

for any two states $|\alpha\rangle$ and $|\beta\rangle$, and $a, b \in \mathbb{C}$.

Our discussion has led us to Wigner's theorem [74]: symmetries are implemented quantum-mechanically either by unitary or antiunitary operators. In fact, continuous symmetries are always implemented by the first kind. This can be understood by thinking that a family of operators $U(\lambda)$, depending on a continuous parameter, can always be smoothly deformed to the identity, a linear and not an antilinear operator. On the other hand, there are two critical discrete symmetries implemented by antiunitary operators: time reversal T and CPT.

### 5.2 Noether's two theorems

In the case of continuous symmetries, we have the celebrated theorem due to Noether linking them to the existence of conserved quantities [45]. What is often called "the" Noether theorem is actually the first of two theorems, dealing with the consequences of *global* and *local* symmetries respectively. Let us begin with the first one considering a classical field theory of $n$ fields whose field equations remain invariant under infinitesimal variations $\phi_i \to \phi_i + \delta_\epsilon \phi_i$ linearly depending on $N$ continuous parameters $\epsilon_A$. There are two essential things about the transformations we are talking about. First, they form a group, as can be seen by noticing that the composition of two symmetries is itself a symmetry and, that for each transformation, there exists its inverse obtained by reversing the signs of $\epsilon_A$. The second fact is that the

infinitesimal transformations can be exponentiated to cover all transformations that can be continuously connected to the identity. The latter statement is rather subtle in the case of diffeomorphisms (i.e., coordinate transformations), but we will not worry about them here.

Since the transformations leave invariant the field equations, the theory's Lagrangian density must change at most by a total derivative, namely

$$S = \int d^4x\, \mathcal{L}(\phi_i, \partial_\mu \phi_i) \qquad \Longrightarrow \qquad \delta_\epsilon S = \int d^4x\, \partial_\mu K^\mu, \tag{5.6}$$

where $K^\mu$ is linear in the $\epsilon_A$'s. At the same time, a general variation of the action can be written as

$$\delta_\epsilon S = \int d^4x \left\{ \left[ \frac{\partial \mathcal{L}}{\partial \phi_i} - \partial_\mu \left( \frac{\partial \mathcal{L}}{\partial\, \partial_\mu \phi_i} \right) \right] \delta_\epsilon \phi_i + \partial_\mu \left( \frac{\partial \mathcal{L}}{\partial\, \partial_\mu \phi_i} \delta_\epsilon \phi_i \right) \right\}, \tag{5.7}$$

so equating expressions (5.6) and (5.7), we find

$$\int d^4x \left\{ \left[ \frac{\partial \mathcal{L}}{\partial \phi_i} - \partial_\mu \left( \frac{\partial \mathcal{L}}{\partial\, \partial_\mu \phi_i} \right) \right] \delta_\epsilon \phi_i + \partial_\mu \left( \frac{\partial \mathcal{L}}{\partial\, \partial_\mu \phi_i} \delta_\epsilon \phi_i - K^\mu \right) \right\} = 0, \tag{5.8}$$

which is valid for arbitrary $\epsilon$. From this equation we identify the conserved current

$$j^\mu(\epsilon) = \frac{\partial \mathcal{L}}{\partial\, \partial_\mu \phi_i} \delta_\epsilon \phi_i - K^\mu \qquad \Longrightarrow \qquad \partial_\mu j^\mu(\epsilon) = \left[ \partial_\mu \left( \frac{\partial \mathcal{L}}{\partial\, \partial_\mu \phi_i} \right) - \frac{\partial \mathcal{L}}{\partial \phi_i} \right] \delta_\epsilon \phi_i \approx 0, \tag{5.9}$$

where again we used the Dirac notation first introduced in page 31. Notice that since the expression of the current is linear in the parameters $\epsilon_A$ the current can be written as $j^\mu(\epsilon) = \epsilon_A j^\mu_A$, and (5.9) is satisfied for arbitrary values of $\epsilon_A$, we conclude that there are a total of $N$ conserved currents $\partial_\mu j^\mu_A$. An important point glaring in the previous analysis is that current conservation happens *on-shell*, i.e., once the equations of motion are implemented.[12]

The second Noether theorem deals with local symmetries depending on a number of point-dependent parameters $\epsilon_A(x)$. It is important to keep in mind that the first theorem remains valid in this case, in the sense that there exists a current $j_\mu$ whose divergence is proportional to the equations of motion. To simplify expressions, let us denote the latter as

$$E_i(\phi) \equiv \partial_\mu \left( \frac{\partial \mathcal{L}}{\partial\, \partial_\mu \phi_i} \right) - \frac{\partial \mathcal{L}}{\partial \phi_i}, \tag{5.10}$$

and consider that our theory is invariant under field transformations involving only $\epsilon_A(x)$ and their first derivatives

$$\delta_\epsilon \phi_i = R_{i,a}(\phi_k)\epsilon_A + R^\mu_{i,A}(\phi_k)\partial_\mu \epsilon_A. \tag{5.11}$$

This includes, for example, the gauge transformations of electromagnetism, $\delta_\epsilon A_\mu = \partial_\mu \epsilon$ (the argument

[12] A note of warning: the term on-shell is employed in physics with at least two different meanings. In the one used here we say that an identity is valid on-shell whenever it holds after the equations of motion are implemented. The second use applies to the four-momentum of a particle with mass $m$. The momentum $p^\mu$ (or the particle carrying it) is said to be on-shell if it satisfies $p^2 = m^2$. As an example, particles running in loops in Feynman diagrams are off-shell in this sense.

here can be easily generalized to include transformations depending up to the $k$-th derivative of the gauge functions). The general variation of the action $\delta_\epsilon S$ has the structure shown in Eq. (5.8),

$$\int d^4x \Big[ - E_i(\phi)\delta_\epsilon\phi_i + \partial_\mu j^\mu(\epsilon)\Big] = 0, \tag{5.12}$$

with $\delta_\epsilon\phi_i$ given in (5.11) and $j^\mu$ the Noether current implied by the first theorem and defined in Eq. (5.9). A crucial difference now is that, since $j^\mu(\epsilon)$ is linear in $\epsilon_A$, when these parameters vanish at infinity the boundary term on the right-hand side appearing when integrating by parts is zero

$$\delta_\epsilon S = -\int d^4x\, \epsilon_A(x)\Big\{R_{i,A}(\phi_k)E_i(\phi_k) - \partial_\mu\Big[R^\mu_{i,A}(\phi_k)E_i(\phi_k)\Big]\Big\}. \tag{5.13}$$

Thus, if this is a symmetry, $\delta_\epsilon S = 0$ for any $\epsilon_A(x)$, we obtain the identities

$$R_{i,A}(\phi_k)E_i(\phi_k) - \partial_\mu\Big[R^\mu_{i,A}(\phi_k)E_i(\phi_k)\Big] = 0, \tag{5.14}$$

where we should remember that $A = 1, \ldots, N$, with $N$ the number of gauge functions (i.e., the dimension of the symmetry's Lie algebra). This result is Noether's second theorem: invariance of a field theory under local transformations implies the existence of several differential identities among the field equations, meaning that some are redundant.

As to the existence of conserved currents associated with local invariance, using Eq. (5.14) it can be shown that

$$\partial_\mu\Big[\epsilon_A(x)R^\mu_{i,A}(\phi_k)E_i(\phi_k)\Big] = E_i(\phi_k)\delta_\epsilon\phi_i, \tag{5.15}$$

from where we read the conserved current

$$S^\mu(\epsilon) \equiv \epsilon_A(x)R^\mu_{i,A}(\phi_k)E_i(\phi_k) \qquad \Longrightarrow \qquad \partial_\mu S^\mu(\epsilon) = E_i(\phi_k)\delta_\epsilon\phi_i \approx 0. \tag{5.16}$$

This quantity is however trivial, in the sense that it vanishes on-shell, $S^\mu(\epsilon) \approx 0$. Notice, however, that the conserved current obtained as the result of the first Noether theorem also applies to the gauge case. Indeed, considering transformations such that $\epsilon_A(x)$ does not vanish at infinity, we find from (5.12)

$$\partial_\mu j^\mu(\epsilon) = E_i(\phi_k)\delta_\epsilon\phi_i \approx 0, \tag{5.17}$$

where $j^\mu$ is explicitly given by the expression on the left of Eq. (5.9). This shows that for theories with local invariances the only nontrivial conserved currents are the ones provided by Noether's first theorem, associated with transformations that do not vanish at infinity (see also the discussion in Box 9 below).

Together with the conserved current from the first Noether theorem, there exists a conserved charge defined by its time component,

$$Q(\epsilon) = \int_\Sigma d^3r\, j^0(\epsilon), \tag{5.18}$$

where $\Sigma$ is a three-dimensional spatial section of spacetime. Using current conservation it is easy to see

that the time derivative of the charge vanishes on-shell

$$\dot{Q}(\epsilon) \approx -\int_\Sigma d^3r\, \boldsymbol{\nabla}\cdot\mathbf{j}(\epsilon) = \int_{\partial\Sigma} d\mathbf{S}\cdot\mathbf{j}(\epsilon) = 0, \tag{5.19}$$

provided the spatial components of the current $\mathbf{j}(\epsilon)$ are zero at $\partial\Sigma$ or, equivalently, there is no flux of charge entering or leaving the spatial sections at infinity.

Applying the first Noether theorem to different symmetries, we get a number of conserved quantities:

- The energy–momentum tensor $T^\mu{}_\nu$ is the conserved current associated with the invariance of field theories under spacetime translations, $x^\mu \to x^\mu + a^\mu$. Its general expression is

$$T^\mu{}_\nu = \frac{\partial\mathcal{L}}{\partial\,\partial_\mu\phi_i}\partial_\nu\phi_i - \delta^\mu_\nu\mathcal{L}, \tag{5.20}$$

  with $\partial_\mu T^\mu{}_\nu = 0$. Notice that this canonical is not necessarily symmetric as, for example, in Maxwell's electrodynamics

$$T^\mu{}_\nu = -F^{\mu\alpha}\partial_\nu A_\alpha + \frac{1}{4}\delta^\mu_\nu F_{\alpha\beta}F^{\alpha\beta}. \tag{5.21}$$

  It can nevertheless be symmetrized by adding a term of the form $\partial_\sigma K^{\sigma\mu}{}_\nu$, with $K^{\sigma\mu}{}_\nu = -K^{\mu\sigma}{}_\nu$, that does not spoil its conservation [75, 76]. In the case of the electromagnetism, the resulting Belinfante–Rosenfeld energy-momentum tensor reads

$$K^{\mu\nu}{}_\sigma = F^{\mu\nu}A_\sigma \qquad \Longrightarrow \qquad \widetilde{T}^\mu{}_\nu = -F^{\mu\alpha}F_{\nu\alpha} + \frac{1}{4}\delta^\mu_\nu F_{\alpha\beta}F^{\alpha\beta}. \tag{5.22}$$

  This modified energy–momentum tensor not only is symmetric but, unlike (5.21), also gauge invariant. Notice that since conserved currents are quantities evaluated on-shell, we can apply the vacuum field equations $\partial_\mu F^{\mu\nu} = 0$.
- Invariance under infinitesimal Lorentz transformations $\delta x^\mu = \omega^\mu{}_\nu x^\nu$, with $\omega_{\mu\nu} = -\omega_{\nu\mu}$, implies the conservation of the total angular momentum

$$J^\mu{}_{\nu\sigma} = T^\mu{}_\nu x_\sigma - T^\mu{}_\sigma x_\nu + S^\mu{}_{\nu\sigma}, \tag{5.23}$$

  where $J^\mu{}_{\nu\sigma} = -J^\mu{}_{\sigma\nu}$ and $\partial_\mu J^\mu{}_{\nu\sigma} = 0$. The first two terms on the right-hand side represent the "orbital" contribution induced by the Lorentz variation of the spacetime coordinates, while $S^\mu{}_{\nu\sigma}$ is the "intrinsic" angular momentum (or spin) coming from the spacetime transformation properties of the field itself. For a scalar field this last part vanishes[13].
- As a further application, let us mention the invariance of complex fields under phase rotation, already anticipated in various examples in previous pages. For instance, in the case of the complex scalar field studied in Box 6, applying (5.9) to infinitesimal variations $\delta_\vartheta\phi = i\vartheta\phi$, $\delta_\vartheta\phi^* = -i\vartheta\phi^*$

[13] To connect with the notation employed in our discussion of the first Noether theorem, let us indicate that the conserved current (5.9) associated to the invariance under spacetime translations is written by $j^\mu(a^\sigma) = T^\mu{}_\nu a^\nu$, whereas $j^\mu(\omega^{\alpha\beta}) = J^\mu{}_{\nu\sigma}\omega^{\nu\sigma}$ is the current whose conservation follows from Lorentz invariance.

leads to the conserved current (3.99). The corresponding analysis for Weyl spinors gives (4.30).

### 5.3 Quantum symmetries: to break or not to break (spontaneously)

In the quantum theory symmetries are realized on the Hilbert space of physical states. In particular, the charge (5.18) is promoted to a Hermitian operator $\widehat{Q}(\epsilon)$ implementing infinitesimal transformations on the fields

$$\delta_\epsilon \widehat{\phi}_k = -i[\widehat{Q}(\epsilon), \widehat{\phi}_k], \tag{5.24}$$

whereas, due to the conservation equation (5.19), it commutes with the Hamiltonian, $[\widehat{Q}(\epsilon), \widehat{H}] = 0$. In the case of rigid transformations, the parameters $\epsilon_A$ can be taken outside the integral in (5.18) to write $\widehat{Q}(\epsilon) = \epsilon_A \widehat{Q}^A$. Finite transformations in the connected component of the identity are obtained then by exponentiating the charge operator

$$\widehat{\mathscr{U}}(\epsilon) = e^{i\epsilon_A \widehat{Q}^A} \qquad \Longrightarrow \qquad \widehat{\mathscr{U}}(\epsilon)^\dagger \widehat{\phi}_k(x) \widehat{\mathscr{U}}(\epsilon) = \mathscr{U}_{k\ell}(\epsilon) \widehat{\phi}_\ell(x), \tag{5.25}$$

where $\mathscr{U}_{k\ell}(\epsilon)$ is the representation of the symmetry group acting on the field indices and the Hermiticity of $\widehat{Q}$ guarantees the unitarity of $\widehat{\mathscr{U}}(\epsilon)$. The implication for the free theory is that the creation–annihilation operators transform covariantly under the symmetry. Consequently, to determine the action of $\widehat{\mathscr{U}}(\epsilon)$ on the Fock space of the theory, we need to know how the charge acts on the vacuum. Here, we may have two possibilities corresponding to different realization of the symmetry.

**Wigner–Weyl realization:** the vacuum state is left invariant by the symmetry

$$\widehat{\mathscr{U}}(\epsilon)|0\rangle = |0\rangle \qquad \Longrightarrow \qquad \widehat{Q}_a|0\rangle = 0. \tag{5.26}$$

If this is the case, the symmetry is manifest in the spectrum, falling into representations of the symmetry group. Since the whole Fock space is generated by successive application of the fields $\widehat{\phi}_k(x)$ on the vacuum, it is enough to know how the symmetry acts on the states $|\phi_k\rangle \equiv \widehat{\phi}_k(x)|0\rangle$,

$$\widehat{\mathscr{U}}(\epsilon)|\phi_k\rangle = \mathscr{U}_{k\ell}(\epsilon)|\phi_\ell\rangle, \tag{5.27}$$

where $\mathscr{U}_{k\ell}(\epsilon)$ is the representation of the symmetry group introduced in (5.25).

This is what happens, for example, in the hydrogen atom. Its ground state has $j = 0$ and therefore remains invariant under a generic rotation labelled by the Euler angles $\phi$, $\theta$, and $\psi$,

$$\widehat{\mathscr{R}}(\phi, \theta, \psi)|0, 0, 0\rangle = |0, 0, 0\rangle, \tag{5.28}$$

while the other states transform in irreps of the rotation group SO(3) $\simeq$ SU(2),

$$\widehat{\mathscr{R}}(\phi, \theta, \psi)|n, j, m\rangle = \sum_{m'=-j}^{j} \mathscr{D}^{(j)}_{mm'}(\phi, \theta, \psi)|n, j, m'\rangle, \tag{5.29}$$

where $\mathscr{D}^{(j)}_{mm'}(\phi,\theta,\psi)$ is the spin $j$ rotation matrix [77]. From this point of view, the angular momentum and magnetic quantum numbers introduced to account for certain properties of atomic spectra are just group theory labels indicating how the atomic state transforms under spatial rotations. Symmetries in quantum mechanical systems with finite degrees of freedom are usually realized à la Wigner–Weyl, since tunneling among different vacua results in an invariant ground state. We will return to this issue on page 64.

**Nambu–Goldstone realization:** the vacuum state is not invariant under the symmetry. This means that the conserved charge does not annihilate the vacuum

$$\widehat{Q}(\epsilon)|0\rangle \neq 0. \tag{5.30}$$

Whenever this happens, the symmetry is said to be *spontaneously broken*. Notice that the previous equation does not imply that $\widehat{Q}_a|0\rangle \neq 0$ for all $a$. There might be a subset of charges satisfying $\widehat{Q}_A|0\rangle = 0$, with $\{A\} \subset \{a\}$ that we refer to as *unbroken* generators. It is easy to see that, since $[\widehat{Q}_A, \widehat{Q}_B]|0\rangle = 0$, they must form a closed subalgebra under commutation.

Let us illustrate this mode of realization of the symmetry with the example of $N$ real scalar fields $\varphi^i$ with action

$$S = \int d^4x \left[\frac{1}{2}\partial_\mu\varphi^i\partial^\mu\varphi^i - V(\varphi^i\varphi^i)\right]. \tag{5.31}$$

This theory is invariant under global infinitesimal transformations

$$\delta_\epsilon\varphi^i = \epsilon_a (T^a_{\mathbf{f}})^i{}_j\varphi^j, \tag{5.32}$$

with $T^a_{\mathbf{f}}$ the generators in the fundamental representation of SO($N$). Using the standard procedure, we compute the associated Hamiltonian

$$H = \int d^3x \left[\frac{1}{2}\pi^i\pi^i + \frac{1}{2}(\boldsymbol{\nabla}\varphi^i)\cdot(\boldsymbol{\nabla}\varphi^i) + V(\varphi^i\varphi^i)\right], \tag{5.33}$$

with $\pi^i = \partial_0\varphi^i$ the conjugate momenta. From this expression we read the SO($N$)-invariant potential energy

$$\mathscr{V}(\varphi^i) = \int d^3x \left[\frac{1}{2}(\boldsymbol{\nabla}\varphi^i)\cdot(\boldsymbol{\nabla}\varphi^i) + V(\varphi^i\varphi^i)\right]. \tag{5.34}$$

Its minimum is attained for spatially constant configurations $\boldsymbol{\nabla}\varphi^i = 0$ lying at the bottom of the potential $V(\varphi^i\varphi^i)$. This is known as the *vacuum expectation value* (vev) of the field and is represented as $\langle\varphi^i\rangle$. Its value is determined by

$$\left.\frac{\partial V}{\partial\varphi^i}\right|_{\varphi^k=\langle\varphi^k\rangle} = 0. \tag{5.35}$$

Once the vev $\langle\varphi^i\rangle$ is known, we can expand the fields around it by writing $\varphi^i = \langle\varphi^i\rangle + \xi^i$. Substituting

in (5.31) we obtain the action for the fluctuations $\xi^i$ whose quantization gives the elementary excitations (particle) of the field in this vacuum.

Here we may encounter two possible situations. One is that the vev of the field is SO($N$) invariant, $(T_{\mathbf{f}})^i{}_j\langle\varphi^j\rangle = 0$. In this case the action of the fluctuations $\xi^i$ inherits the global symmetry of the parent theory that is then realized à la Wigner–Weyl. Here we want to explore the second alternative, the vev breaks at least part of the symmetry. Let us split the SO($N$) generators into $T^a_{\mathbf{f}} = \{K^\alpha_{\mathbf{f}}, H^A_{\mathbf{f}}\}$, such that

$$(K^\alpha_{\mathbf{f}})^i{}_j\langle\varphi^j\rangle \neq 0, \qquad (H^A_{\mathbf{f}})^i{}_j\langle\varphi^j\rangle = 0, \tag{5.36}$$

and the global symmetry SO($N$) is spontaneously broken. As argued after Eq. (5.30), the generators preserving the symmetry must form a Lie subalgebra generating the unbroken subgroup $H \subset$ SO($N$) and we have the spontaneous symmetry breaking (SSB) pattern SO($N$) $\to H$.

Generically, the action for the field fluctuations around the vev can be written as

$$S = \int d^4x \left(\frac{1}{2}\partial_\mu\xi^i\partial^\mu\xi^i - \frac{1}{2}M^2_{ij}\xi^i\xi^j + \dots\right), \tag{5.37}$$

where the ellipsis stands for interactions terms and the mass-squared matrix $M^2_{ij}$ is given by

$$M^2_{ij} \equiv \left.\frac{\partial^2 V}{\partial\varphi^i\partial\varphi^j}\right|_{\varphi^k=\langle\varphi^k\rangle}. \tag{5.38}$$

The SO($N$) invariance of the potential $\delta_\epsilon V = 0$ implies

$$\epsilon_a\frac{\partial V}{\partial\varphi^i}(T^a_{\mathbf{f}})^i{}_j\varphi^j = 0 \qquad \Longrightarrow \qquad \epsilon_a\frac{\partial^2 V}{\partial\varphi^k\partial\varphi^i}(T^a_{\mathbf{f}})^i{}_j\varphi^j + \epsilon_a\frac{\partial V}{\partial\varphi^i}(T^a_{\mathbf{f}})^i{}_k = 0, \tag{5.39}$$

where in the equation on the right we have taken a further derivative with respect to $\varphi^k$. Evaluating this expression at the vev, and taking into account eqs. (5.35) and (5.38), we find

$$M_{ik}(T^a_{\mathbf{f}})^k{}_j\langle\varphi^j\rangle = 0. \tag{5.40}$$

This equation is trivially satisfied for the unbroken generators $H^A_{\mathbf{f}}$, but has very nontrivial physical implications for $K^\alpha_{\mathbf{R}}$. It states that there are as many zero eigenvalues of the mass matrix as broken generators, i.e., the theory contains one massless particle for each generator not preserving the vacuum. This result is the Goldstone theorem [78, 79], and the corresponding massless particles emerging as the result of spontaneous symmetry breaking are known as Nambu–Goldstone (NG) modes [80, 81]. Although obtained here using a particular example and in a classical setup, the result is also valid quantum mechanically and applicable to any field theory with a global symmetry group $G$ spontaneously broken down to a subgroup $H \subset G$, where the broken part of the symmetry is the coset space $G/H$. One way to prove the Goldstone theorem in the quantum theory is by considering instead of the classical action the quantum effective action and replacing $V(\varphi^i\varphi^i)$ with the effective potential, including all interactions among the scalar fields resulting from resumming quantum effects. It can also be shown that the NG modes always

have zero spin, also known as NG bosons.

Although we are mostly concerned with applications to particle physics, the idea of SSB, in general, and the Goldstone theorem, in particular, have critical applications to nonrelativistic systems, particularly in condensed matter physics.[14] In particular, the notion of SSB is intimately related to the theory of phase transitions [82–84]. It is frequently the case that the phase change is associated with the system changing its ground state. For example, the translational symmetry present in a liquid is spontaneously broken at its freezing point when the full group of three-dimensional translation is broken down to the crystalographic group preserving the lattice in the solid phase. The corresponding NG bosons are the three species of acoustic phonons. These are massless quasiparticles in the sense that their dispersion relation at low momentum takes the form $E_{\mathbf{k}} \simeq c_s|\mathbf{k}|$, with $c_s$ the speed of sound, so it has no mass gap. Another well-known example is a ferromagnet below the Curie point. The rotationally symmetric ground state at high temperature is replaced by a lowest energy configuration where atomic magnetic moments align, generating a macroscopic magnetization that spontaneously breaks rotational symmetry. Magnetic waves, called magnons, are the associated NG gapless modes.

Besides their intrinsic physical interest, these condensed matter examples are useful in bringing home a very important aspect of NG bosons: they do not need to be elementary states. Indeed, phonons and magnons are quasiparticles and, therefore, collective excitations of the system. But also in high energy physics we encounter situations where the NG bosons are bound states of elementary constituents. The most relevant example are the pions, appearing as NG bosons associated with the spontaneous breaking of chiral symmetry in QCD (see Box 8 below).

It is frequently stated that systems with SSB present vacuum degeneracy. Although technically the theory might possess various vacua, there are important subtleties involved in the infinite volume limit preventing quantum transitions among them, that would restore the broken symmetry through tunneling. Let us consider a theory at finite volume $V$ and with a family of degenerate vacua labelled by a properly normalized real parameter $\xi$. It can be shown that the overlap between any two of these vacua is exponentially suppressed but nonzero (see Chapter 7 of Ref. [14] for a more detailed analysis)

$$|\langle\xi'|\xi\rangle| = e^{-\frac{1}{4}(\xi'-\xi)^2 V^{\frac{2}{3}}}|\langle\xi|\xi\rangle|. \tag{5.41}$$

This means that transitions among Fock states built on different vacua are allowed, resulting in a unique ground state invariant under the original symmetry. As a consequence, no SSB can happen at finite volume and symmetries are usually realized à la Wigner–Weyl.

The situation is radically different in the $V \to \infty$ limit when the overlap between any two vacua vanishes, $\langle\xi'|\xi\rangle \to 0$. This means that the Fock space of states builds on different vacua are mutually orthogonal, and no transition among them can occur. At a more heuristic level, what happens is that at infinite volume switching from one vacuum to another requires a nonlocal operation acting at each spacetime point. Notice, however, that at a practical level if the volume is "large enough" compared with the system's microscopic characteristic scale we can consider the vacua as orthogonal for all purposes. This

[14] It should be stressed that historically the very notion of SSB and of NG bosons was inspired by solid state physics, as it is clear in the seminal works by Yoichiro Nambu [80] and Jeffrey Goldstone [78]. Another example of this cross-fertilization between the fields of condensed matter and high energy physics can be found in the formulation of the Brout–Englert–Higgs mechanism to be discussed in Section 5.4.

is why we see SSB in finite samples, as illustrated by the examples of ferromagnets and superconductors.

**Box 8. Of quarks, chiral symmetry breaking, and pions**

The SM offers a very important implementation of SSB as a consequence of quark low-energy dynamics. Let us consider a generalization of the action in Eq. (4.70), now with $N_f$ different quark flavors. Writing $\boldsymbol{q}^T = (q_1, \ldots, q_{N_f})$, the action reads

$$\begin{aligned} S &= \int d^4x\, \overline{\boldsymbol{q}}\big(i\partial\!\!\!/\mathbb{1} - \boldsymbol{m}\big)\boldsymbol{q} + S_{\rm int} \\ &= \int d^4x \left(i\overline{\boldsymbol{q}}_R\partial\!\!\!/\boldsymbol{q}_R + i\overline{\boldsymbol{q}}_L\partial\!\!\!/\boldsymbol{q}_L - \overline{\boldsymbol{q}}_R\boldsymbol{m}\boldsymbol{q}_L - \overline{\boldsymbol{q}}_L\boldsymbol{m}\boldsymbol{q}_R\right) + S_{\rm int}, \end{aligned} \tag{5.42}$$

where in the second line we split the quark fields into its right- and left-handed chiralities and in $S_{\rm int}$ we include all interaction terms. This theory is invariant under global $\mathrm{U}(N_f)$ transformations acting on the fermion fields as

$$\boldsymbol{q}_{R,L} \to \mathscr{U}(\alpha)\boldsymbol{q}_{R,L} \qquad \text{where} \qquad \mathscr{U}(\alpha) = e^{i\alpha^A T^A_{\mathbf{R}}}, \tag{5.43}$$

and $(T^A_{\mathbf{R}})^i{}_j$, with $A = 1, \ldots, N_f^2$, are the $\mathrm{U}(N_f)$ generators in the representation $\mathbf{R}$ with dimension $N$. We observe that it is the presence of the mass term, mixing right- and left-handed quarks, that forces the two chiralities to transform under the same transformation of $\mathrm{U}(N_f)$. This is why in the chiral limit (i.e., zero quark masses $\boldsymbol{m} \to 0$) the global symmetry is enhanced from $\mathrm{U}(N_f)$ to $\mathrm{U}(N_f)_R \times \mathrm{U}(N_f)_L$, acting independently on the two chiralities

$$\boldsymbol{q}_R \to \mathscr{U}(\alpha_R)\boldsymbol{q}_R, \qquad \boldsymbol{q}_L \to \mathscr{U}(\alpha_L)\boldsymbol{q}_L, \tag{5.44}$$

where $\alpha^a_R$ and $\alpha^a_L$ are independent. Thus, there are two independent Noether currents

$$j^\mu_R(\alpha) = \alpha^A_R\overline{\boldsymbol{q}}_R\gamma^\mu T^A_{\mathbf{R}}\boldsymbol{q}_R, \qquad j^\mu_L(\alpha) = \alpha^A_L\overline{\boldsymbol{q}}_L\gamma^\mu T^A_{\mathbf{R}}\boldsymbol{q}_L \tag{5.45}$$

as well as $2 \times N_f^2$ conserved charges

$$Q^A_R = \int d^3x\, \boldsymbol{q}^\dagger_R T^A_{\mathbf{R}}\boldsymbol{q}_R, \qquad Q^A_L = \int d^3x\, \boldsymbol{q}^\dagger_L T^A_{\mathbf{R}}\boldsymbol{q}_L. \tag{5.46}$$

Upon quantization, these charges are replaced by the corresponding operators $\widehat{Q}^A_{R,L}$, whose commutator realizes the algebra of generators of $\mathrm{U}(N_f)_R \times \mathrm{U}(N_f)_L$.

Taking into account that $\mathrm{U}(N_f) = \mathrm{U}(1) \times \mathrm{SU}(N_f)$, the theory's global symmetry group can be written as

$$\mathrm{U}(N_f)_R \times \mathrm{U}(N_f)_L = \mathrm{U}(1)_B \times \mathrm{U}(1)_A \times \mathrm{SU}(N_f)_R \times \mathrm{SU}(N_f)_L. \tag{5.47}$$

The first two factors on the right-hand side act on the quark fields respectively as

$$\boldsymbol{q} \to e^{i\alpha}\boldsymbol{q}, \qquad \boldsymbol{q} \to e^{i\beta\gamma_5}\boldsymbol{q}, \tag{5.48}$$

the former symmetry leading to baryon number conservation (hence the subscript). The $\mathrm{U}(1)_A$ factor is an axial vector transformation acting on the two chiralities with opposite phases and is broken by anomalies (more on this in Section 7). The action of the two $\mathrm{SU}(N_f)_{R,L}$ factors, on the other hand, is defined by

$$\mathrm{SU}(N_f)_R : \begin{cases} \boldsymbol{q}_R & \to & U_R\boldsymbol{q}_R \\ \boldsymbol{q}_L & \to & \boldsymbol{q}_L \end{cases} \qquad \mathrm{SU}(N_f)_L : \begin{cases} \boldsymbol{q}_R & \to & \boldsymbol{q}_R \\ \boldsymbol{q}_L & \to & U_L\boldsymbol{q}_L \end{cases} \tag{5.49}$$

with

$$U_{R,L} \equiv e^{i\alpha^I_{L,R}t^I_{\mathbf{f}}} \tag{5.50}$$

and $t^I_{\mathbf{f}}$ ($I = 1, \ldots, N_f^2 - 1$) the generators of the fundamental irrep of $\mathrm{SU}(N_f)$.

At low energies the strong quark dynamics triggers quark condensation, giving a non-zero vev to the scalar quark bilinear $\overline{q}_i q_j$

$$\langle 0|\overline{q}_i q_j|0\rangle \equiv \langle 0|\big(\overline{q}_{i,R}q_{j,L} + \overline{q}_{i,L}q_{i,R}\big)|0\rangle = \Lambda^3_{\chi\mathrm{SB}}\delta_{ij}, \tag{5.51}$$

where $\Lambda_{\chi\mathrm{SB}}$ is the energy scale associated with the condensation. This vev, however, is only invariant under the "diagonal" subgroup of the $\mathrm{SU}(N_f)_R \times \mathrm{SU}(N_f)_L$ transformations (5.49) consisting of transformations with $U_R = U_L$. What happens is that the global $\mathrm{SU}(N_f)_R \times \mathrm{SU}(N_f)_L$ chiral symmetry is spontaneously broken down to its vector subgroup

$$\mathrm{U}(1)_B \times \mathrm{SU}(N_f)_R \times \mathrm{SU}(N_f)_L \longrightarrow \mathrm{U}(1)_B \times \mathrm{SU}(N_f)_V. \tag{5.52}$$

Goldstone's theorem implies that associated with each spontaneously broken generator there should be a massless NG boson. In our case there are $N_f^2 - 1$ broken generators corresponding to the $\mathrm{SU}(N_f)_A$ factor. Excitations around the vev (5.51) are parametrized by the field $\Sigma_{ij}(x)$ defined by

$$\overline{q}_i(x)q_j(x) = \Lambda^3_{\chi\mathrm{SB}}\Sigma_{ij}(x). \tag{5.53}$$

This in turn can be written in terms of the NG matrix field $\boldsymbol{\pi}(x) \equiv \pi^A(x)t^A_{\mathbf{f}}$ as

$$\boldsymbol{\Sigma}(x) \equiv e^{\frac{i\sqrt{2}}{f_\pi}\boldsymbol{\pi}(x)}, \tag{5.54}$$

with $f_\pi$ a constant with dimensions of energy called the pion decay constant for reasons that will

eventually become clear. Mathematically speaking, the field $\mathbf{\Sigma}$ parametrizes the coset

$$\frac{\mathrm{SU}(N_f)_R \times \mathrm{SU}(N_f)_L}{\mathrm{SU}(N_f)_V}, \tag{5.55}$$

leading to the following transformation under $\mathrm{SU}(N_f)_R \times \mathrm{SU}(N_f)_L$:

$$\mathbf{\Sigma} \longrightarrow U_R \mathbf{\Sigma} U_L^\dagger. \tag{5.56}$$

We specialize the analysis now to the case $N_f = 2$, with only the $u$ and $d$ quarks. The unbroken $\mathrm{SU}(2)_V$ symmetry is just the good old isospin interchanging both quarks, while the NG bosons are the three pions $\pi^\pm$ and $\pi^0$

$$\boldsymbol{\pi} = \frac{1}{\sqrt{2}} \begin{pmatrix} \pi^0 & \sqrt{2}\pi^+ \\ \sqrt{2}\pi^- & -\pi^0 \end{pmatrix}. \tag{5.57}$$

The objection might be raised that pions are not massless particles as the Goldstone theorem requires. Our analysis has ignored the nonvanishing quark masses, explicitly breaking the $\mathrm{SU}(2)_R \times \mathrm{SU}(2)_L$ global chiral symmetry. Since the $u$ and $d$ quarks are relatively light, we have instead three *pseudo*-NG bosons whose masses are not zero but still lighter than other states in the theory. It is precisely the strong mass hierarchy between the pions and the remaining hadrons what identifies them as the pseudo-NG bosons associated with chiral symmetry breaking. In the $N_f = 3$ case, where we add the strange quark to the two lightest ones, $\mathrm{SU}(3)_V$ is Gell-Mann's eightfold way discussed on page 54 and the set of pseudo-NG bosons is enriched by the four kaons and the $\eta$-meson in the octet appearing on the right-hand side of Eq. (4.72).

As mentioned in the introduction, quarks and gluons do not exist as asymptotic states and QCD at low energies is a theory of hadrons. The lowest lying particles are the pion triplet, whose interactions can be obtained from symmetry considerations alone playing the EFT game. The question is how to write the simplest action for NG bosons containing operators with the lowest energy dimension and compatible at the same time with all the symmetries of the theory. For terms with just two derivatives, the solution is

$$\begin{aligned} S_{\mathrm{NG}} &= \frac{f_\pi^2}{4} \int d^4x \, \mathrm{tr}\left(\partial_\mu \mathbf{\Sigma}^\dagger \partial^\mu \mathbf{\Sigma}\right) \\ &= \int d^4x \left[\frac{1}{2}\mathrm{tr}\left(\partial_\mu \boldsymbol{\pi} \partial^\mu \boldsymbol{\pi}\right) - \frac{1}{3f_\pi^2}\mathrm{tr}\left(\partial_\mu \boldsymbol{\pi}[\boldsymbol{\pi}, [\boldsymbol{\pi}, \partial^\mu \boldsymbol{\pi}]\right) + \ldots\right]. \end{aligned} \tag{5.58}$$

This *chiral effective action* contains an infinite sequence of higher-dimensional operators suppressed by increasing powers of the dimensionful constant $f_\pi$. It determines how pions couple among themselves at low energies. Its coupling to the electromagnetic field is obtained by replacing $\partial_\mu \mathbf{\Sigma}$ by the adjoint covariant derivative $D_\mu \mathbf{\Sigma} = \partial_\mu \mathbf{\Sigma} - iA_\mu[Q, \mathbf{\Sigma}]$ where the charge matrix is given by $Q = e\sigma^3$. This, however, does not exhaust all their electromagnetic interactions. Neutral pions couple to photons as a consequence of the anomalous realization of the $\mathrm{U}(1)_A$ symmetry, resulting in the $\pi^0 \to 2\gamma$

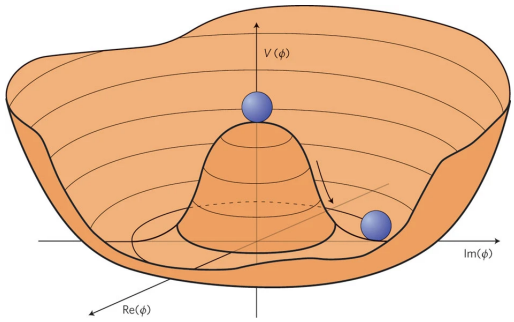



**Fig. 9:** Illustration from Ref. [88] depicting the celebrated Mexican hat potential shown in Eq. (5.61).

decay (see Section 7).

In our analysis of chiral symmetry breaking we encountered two energy scales: $\Lambda_{\chi\text{SSB}}$ appearing in (5.51) as a consequence of the quark condensate having dimensions of (energy)$^3$, and $f_\pi$ needed to give the pion fields their proper dimensions in Eq. (5.54). Both of them have to be experimentally measured. In the pion EFT it is $f_\pi$ that determines the relative size of the infinite terms in the effective action (5.58). Operators weighted by $f_\pi^{-n}$ typically give contributions or order $(E/f_\pi)^n$ with $E$ the characteristic energy of the process under study. In the spirit of EFT, working at a given experimental precision, only a finite number of terms in the chiral Lagrangian have to be retained, making the theory fully predictive (see Refs. [85, 86] for comprehensive reviews of chiral perturbation theory).

### 5.4 The Brout–Englert–Higgs mechanism

Besides the ones already discussed, a further instance of SSB in condensed matter connecting with one of the key concepts in the formulation of the SM is the Brout–Englert–Higgs (BEH) mechanism. In the Bardeen–Cooper–Schrieffer (BCS) theory of superconductivity the transition from the normal to the superconductor phase is triggered by the condensation of Cooper pairs, collective excitations of two electrons bound together by phonon exchange. Having net electric charge, the Cooper pair wave function transforms under electromagnetic U(1) phase rotations and their condensation spontaneously breaks this invariance. The physical consequence of this is a screening of magnetic fields inside the superconductor, the Meissner effect, physically equivalent to the electromagnetic vector potential $\mathbf{A}(t, \mathbf{r})$ acquiring an effective nonzero mass [87].

The main difference between the BCS example and the ones discussed above is that this is not about spontaneously breaking some global symmetry, but gauge invariance itself. This might look like risky business, since we know that preserving gauge invariance is crucial to get rid of unwanted physical states that otherwise would pop up in the theory's physical spectrum destroying its consistency. As we will see, due to the magic of SSB gauge invariance is in fact not lost, only hidden. That is why, even if not manifest, it still protects the theory.

Let us analyze spontaneous symmetry breaking triggered by a complex scalar coupled to the elec-

tromagnetic field. We start with the action

$$S = \int d^4r \left[ -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + (D_\mu \phi)^* (D^\mu \phi) - \frac{\lambda}{4} \left( \phi^* \phi - \frac{v^2}{2} \right)^2 \right], \tag{5.59}$$

where $D_\mu = \partial_\mu - ieA_\mu$ is the covariant derivative already introduced in the footnote of page 40. This action is invariant under U(1) gauge transformations acting as

$$\phi(x) \longrightarrow e^{ie\epsilon(x)} \phi(x), \qquad \phi(x)^* \longrightarrow e^{-ie\epsilon(x)} \phi(x)^*, \qquad A_\mu(x) \longrightarrow A_\mu(x) + \partial_\mu \epsilon(x). \tag{5.60}$$

As shown in Fig. 9, the scalar field potential

$$V(\phi^* \phi) = \frac{\lambda}{4} \left( \phi^* \phi - \frac{v^2}{2} \right)^2, \tag{5.61}$$

has the celebrated Mexican hat shape with a valley of minima located at $\phi^* \phi = \frac{v^2}{2}$. When the scalar field takes a nonzero vev

$$\langle \phi \rangle = \frac{v}{\sqrt{2}} e^{i\vartheta_0}, \tag{5.62}$$

U(1) invariance is spontaneously broken, since $\langle \phi \rangle$ does not remain invariant, $\langle \phi \rangle \to e^{ie\epsilon} \langle \phi \rangle$. The dynamics of the fluctuations around the vev (5.62) is obtained by plugging

$$\phi(x) = \frac{1}{\sqrt{2}} \big[ v + h(x) \big] e^{i\vartheta(x)} \tag{5.63}$$

into (5.59). The resulting action is

$$\begin{aligned} S = \int d^4x \Big[ & -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{e^2 v^2}{2} \left( A_\mu + \frac{1}{e} \partial_\mu \vartheta \right) \left( A^\mu + \frac{1}{e} \partial^\mu \vartheta \right) + \frac{1}{2} \partial_\mu h \partial^\mu h - \frac{\lambda v^2}{4} h^2 \\ & - \frac{\lambda v}{4} h^3 - \frac{\lambda}{16} h^4 + \frac{e^2}{2} \left( A_\mu + \frac{1}{e} \partial_\mu \vartheta \right) \left( A^\mu + \frac{1}{e} \partial^\mu \vartheta \right) (2vh + h^2) \Big], \end{aligned} \tag{5.64}$$

which remains invariant under U(1) gauge transformations, now acting as

$$A_\mu \longrightarrow A_\mu + \partial_\mu \epsilon, \qquad \vartheta \longrightarrow \vartheta - e\epsilon, \qquad h \longrightarrow h. \tag{5.65}$$

In fact, the phase field $\vartheta(x)$ is the NG boson resulting from the spontaneous breaking of the U(1) symmetry by the vev in Eq. (5.62).

At this stage, we still keep a photon with two polarizations while the two real degrees of freedom of the complex field $\phi$ have been recast in terms of the field $h$ and the NG boson $\vartheta$. We can fix the gauge freedom (5.65) by setting $\vartheta = 0$. In doing so, the disappearing NG boson transmutes into the longitudinal component of $A_\mu$, as befits a massive gauge field (see the footnote on page 32). We then

arrive at the gauge-fixed action

$$S = \int d^4x \left( -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{e^2 v^2}{2} A_\mu A^\mu + \frac{1}{2} \partial_\mu h \partial^\mu h - \frac{\lambda v^2}{4} h^2 \right.$$
$$\left. - \frac{\lambda v}{4} h^3 - \frac{\lambda}{16} h^4 + e^2 v A_\mu A^\mu h + \frac{e^2}{2} A_\mu A^\mu h^2 \right), \tag{5.66}$$

where the photon has acquired a nonzero mass[15]

$$m_\gamma = ev. \tag{5.67}$$

The real scalar field $h$ gets massive as well,

$$m_h = v\sqrt{\frac{\lambda}{2}}, \tag{5.68}$$

and has cubic and quartic self-interactions terms, besides coupling to the photon through terms involving two gauge fields and one scalar and two gauge fields and two scalars. As we see, no degree of freedom has gone amiss. We ended up with a massive photon with three physical polarizations and a real scalar, making up for the four real degrees of freedom we started with. SSB has just rearranged the theory's degrees of freedom.

Here we have been only concerned with giving mass to the photon. Imagine now that we would have two chiral fermions $\psi_R$, $\psi_L$ such that they transform differently under U(1)

$$\psi_L(x) \longrightarrow e^{ie\epsilon(x)} \psi_L(x), \qquad \psi_R(x) \longrightarrow \psi_R(x). \tag{5.69}$$

Due to the theory's chiral nature, a mass term of the form $\overline{\psi}_L \psi_R + \overline{\psi}_R \psi_L$ would not be gauge invariant, so it seems that we need to keep our fermions massless for the sake of consistency. Using the Higgs field, however, there is a way to construct an action where the fermions couple to the complex scalar field in a gauge invariant way,

$$S_{\text{fermion}} = \int d^4x \left( i\overline{\psi}_R \not{D} \psi_R + i\overline{\psi}_L \not{D} \psi_+ - c\phi \overline{\psi}_L \psi_R - c\phi^* \overline{\psi}_R \psi_L \right), \tag{5.70}$$

where $c$ is some dimensionless constant. This particular form of the coupling between $\phi$ and the fermions is called a Yukawa coupling, since it is similar to the one introduced by Hideki Yukawa in his 1935 theory of nuclear interactions between nucleons and mesons [89]. The interest of this construction is that once the field $\phi$ acquires the vev (5.62), and after gauging away the field $\vartheta$, the fermion action takes the form

$$S_{\text{fermion}} = \int d^4x \left[ i\overline{\psi}_R \not{D} \psi_R + i\overline{\psi}_L \not{D} \psi_L - \frac{cv}{\sqrt{2}} (\overline{\psi}_L \psi_R - \overline{\psi}_R \psi_L) \right.$$
$$\left. - \frac{c}{\sqrt{2}} h \overline{\psi}_L \psi_R - \frac{c}{\sqrt{2}} h \overline{\psi}_R \psi_L \right]. \tag{5.71}$$

[15] The same result can be obtained noticing that the action (5.64) contains a term $ev^2 A^\mu \partial_\mu \vartheta$ mixing the NG boson and the gauge field. Physically, this means that as the photon propagates it transmutes into the NG boson and vice versa. Resumming these transmutations results in the mass term for $A^\mu$.

Thus, the same mechanism giving mass to the photon also results in a mass for the fermion field,

$$m_f = \frac{cv}{\sqrt{2}}, \tag{5.72}$$

also generated without an explicit breaking of gauge invariance, hidden due to the choice of vacuum of the complex scalar field. Notice that, owing to symmetry breaking, the now massive Dirac fermion couples to the remaining scalar degree of freedom $h$ with a strength controlled by the dimensionless constant $\frac{c}{\sqrt{2}} = \frac{m_f}{v}$. This indicates that the higher the mass of the fermion, the stronger it couples to the Higgs field. This feature, as we will see, has important experimental consequences for the SM.

This Abelian Higgs model illustrates the basic features of the BEH mechanism responsible for giving masses to the SM particles, with the scalar field $h$ corresponding to the Higgs boson discovered at CERN in 2012 [19, 20]. In its nonrelativistic version it also provides the basis for the Ginzburg–Landau analysis of the BCS theory of superconductivity, where the free energy in the broken phase has the same structure as the potential terms in the action (5.59)

$$\mathscr{F}_{\rm BCS} = \int d^3r \left\{ \frac{1}{2\mu}(\boldsymbol{\nabla}\times\mathbf{A})^2 + \frac{1}{2m_*}\left|\boldsymbol{\nabla}\phi - ie_*\mathbf{A}\phi\right|^2 + \frac{\lambda(T)}{4}\left[\phi^*\phi - \frac{v(T)}{2}^2\right]^2 \right\}. \tag{5.73}$$

Here $\phi(\mathbf{r})$ is the Cooper pair condensate, $\mu$ the magnetic permeability of the medium, and $m_*$ and $e_*$ the effective mass and charge of the quasiparticles. For $T > T_c$ we have $v(T) = 0$, so at temperatures above the critical one, the only minimum of the free energy is at $\langle\phi\rangle = 0$. When $T < T_c$, on the other hand, $v(T) \neq 0$ and the U(1) invariance of the theory is spontaneously broken at the $|\langle\phi\rangle| = v(T)$ minima, while the former one at $\langle\phi\rangle = 0$ becomes a local maximum. As in the case studied earlier, this results in a nonzero mass for the vector potential $\mathbf{A}(\mathbf{r})$ given by $m(T) = e_*v(T)$. This provides the order parameter of the transition and physically accounts for the Meissner effect inside the superconductor [83]. The system also contains a scalar massive excitation, the condensed matter equivalent of the Higgs boson [90, 91].

**Box 9. "Large" vs. "small" gauge transformations**

We return briefly to the discussion of Noether's second theorem on page 58. There we paid attention to gauge transformations in the connected component of the identity and made an important distinction among those approaching the identity at the spacetime boundary ($\epsilon_A \to 0$) and those that do not. Let us call them "small" and "large" gauge transformations, respectively. To understand the physical difference between them, we compare (5.17) with (5.16) to see that $j^\mu - S^\mu$ is conserved even off-shell, namely that $\partial_\mu(j^\mu - S^\mu)$ is *identically zero*. This means that we can write

$$j^\mu = S^\mu + \partial_\nu k^{\mu\nu} \approx \partial_\nu k^{\mu\nu}, \tag{5.74}$$

where $k^{\mu\nu}$ is an antisymmetric tensor and we have applied that $S^\mu$ vanishes on-shell. This peculiar structure of the gauge theory current implies that the gauge charge is determined by an integral over

the *boundary* of the spatial sections

$$Q \approx \int_\Sigma dV\, \partial_i k^{0i} = \int_{\partial\Sigma} dS_i\, k^{0i}. \tag{5.75}$$

Since the current, and therefore also $k^{\mu\nu}$, is linear in the gauge functions $\epsilon_A(x)$, we conclude that the charge vanishes for "small" gauge transformations

$$Q_{\text{small}} \approx 0. \tag{5.76}$$

This is not the case of "large" transformations, the ones determining the value of $Q$.

A very important fact to remember about "small" gauge transformations is that they are the ones leading to the Noether identities (5.14) that, as we indicated, express the redundancy intrinsic to gauge theories. Quantum mechanically, invariance under these transformations is mandatory in order to get rid of the spurious states that we introduced as the price of maintaining locality and Lorentz covariance. They cannot be spontaneously broken or affected by anomalies without rendering the theory inconsistent. However, no such restriction exists for "large" transformations, that can be broken without disastrous consequences.

To connect with the discussion of the Abelian Higgs model, let us look at the case of Maxwell's electrodynamics in the temporal gauge $A_0 = 0$. In the quantum theory, the vacuum Gauss law constraint $\boldsymbol{\nabla}\cdot\mathbf{E} = 0$ is implemented by the corresponding operator annihilating physical states, namely (to keep notation simple, we drop hats to denote operators)

$$\boldsymbol{\nabla}\cdot\mathbf{E}|\text{phys}\rangle = 0. \tag{5.77}$$

Finite gauge transformations preserving the temporal gauge condition $A_0 = 0$ are generated by time-independent gauge functions and implemented in the space of states by the operator

$$\mathscr{U}_\epsilon = \exp\left[i\int d^3r\, \mathbf{E}(t,\mathbf{r})\cdot\boldsymbol{\nabla}\epsilon(\mathbf{r})\right]. \tag{5.78}$$

Using the canonical commutation relations (3.68), we readily compute

$$\begin{aligned} \mathscr{U}_\epsilon A_0(t,\mathbf{r})\mathscr{U}_\epsilon^{-1} &= 0, \\ \mathscr{U}_\epsilon \mathbf{A}(t,\mathbf{r})\mathscr{U}_\epsilon^{-1} &= \mathbf{A}(t,\mathbf{r}) + \boldsymbol{\nabla}\epsilon(\mathbf{r}). \end{aligned} \tag{5.79}$$

At the same time, the operator $\mathscr{U}_\epsilon$ leaves the physical states invariant

$$\begin{aligned} \mathscr{U}_\epsilon|\text{phys}\rangle &= \exp\left[i\int d^3x\, \mathbf{E}(t,\mathbf{r})\cdot\boldsymbol{\nabla}\epsilon(\mathbf{r})\right]|\text{phys}\rangle \\ &= \exp\left[-i\int d^3x\, \epsilon(\mathbf{r})\boldsymbol{\nabla}\cdot\mathbf{E}(t,\mathbf{r})\right]|\text{phys}\rangle = |\text{phys}\rangle, \end{aligned} \tag{5.80}$$

where in the second line it is crucial that the gauge function $\epsilon(\mathbf{r})$ *vanishes at infinity* so that after

integrating by parts we do not pick up a boundary term. This means that $\mathscr{U}_\epsilon \to \mathbb{1}$ as $|\mathbf{r}| \to \infty$.

We have shown that invariance of the physical states under "small" gauge transformations follows from Gauss' law (5.77) annihilating them, precisely the condition that factors out the spurious degrees of freedom. The conclusion is that "large" gauge transformations are not necessary to eliminate the gauge redundancy and can be broken without jeopardizing the consistency of the theory. This is precisely how the BEH mechanism works. The nonvanishing vacuum expectation value of the complex scalar field breaks "large" gauge transformations without spoiling Gauss' law. This is the reason why we need to qualify our statement in pages 20 and 56 that gauge invariance is just a redundancy in state labelling: "small" gauge transformations are indeed redundancies, but "large" gauge transformations are *bona fide* symmetries.

## 6 Some more gauge invariances

So far the only gauge theory we dealt with was Maxwell's electrodynamics, although here and there we hinted at its non-Abelian generalizations. It is about time to introduce these in a more systematic fashion. We start with a set of fermions $\boldsymbol{\psi}^T = (\psi_1, \ldots, \psi_N)$ transforming in some representation $\mathbf{R}$ of the gauge group $G$

$$\boldsymbol{\psi} \longrightarrow e^{i\alpha^a T^a_{\mathbf{R}}} \boldsymbol{\psi} \equiv g(\alpha)\boldsymbol{\psi}. \tag{6.1}$$

By now, we know very well how to construct an action that has this symmetry,

$$S = \int d^4x\, \overline{\boldsymbol{\psi}}\big(i\partial\!\!\!/ - m\big)\boldsymbol{\psi}. \tag{6.2}$$

The problem arises when we want to make $G$ a local invariance. In this case, the action we just wrote fails to be invariant due to the nonvanishing derivatives of $\alpha^a(x)$,

$$\partial_\mu \boldsymbol{\psi} \longrightarrow g\partial_\mu \boldsymbol{\psi} + i\partial_\mu g \boldsymbol{\psi} = g\big(\partial_\mu \boldsymbol{\psi} + ig^{-1}\partial_\mu g\big)\boldsymbol{\psi}, \tag{6.3}$$

where, to avoid cluttering expressions, we have omitted the dependence of the group element $g$ on the parameters $\alpha^a$.

To overcome this problem we have to find a covariant derivative $D_\mu$, similarly to the one we introduced for Maxwell's theory, with the transformation

$$D_\mu \boldsymbol{\psi} \longrightarrow g D_\mu \boldsymbol{\psi}. \tag{6.4}$$

A reasonable Ansatz turns out to be

$$D_\mu \boldsymbol{\psi} = \big(\partial_\mu - iA_\mu\big)\boldsymbol{\psi}, \tag{6.5}$$

where we omitted the identity multiplying $\partial_\mu$ and $A_\mu \equiv A^a_\mu T^a_{\mathbf{R}}$ is a field taking values in the algebra of

generators of $G$. In order to get the transformations (6.4), $A_\mu$ has to transform according to

$$A_\mu \longrightarrow A'_\mu = ig^{-1}\partial_\mu g + g^{-1}A_\mu g. \tag{6.6}$$

With this we can turn (6.2) into a locally invariant action by replacing $\partial_\mu$ with $D_\mu$ defined in Eq. (6.5). In addition, we must include the dynamics of the new field $A_\mu$ adding a suitable kinetic term that preserves the gauge invariance of the fermionic action. The Abelian-informed choice $\partial_\mu A_\nu - \partial_\nu A_\mu$ for the gauge field strength will not do, since it does not transform covariantly

$$\begin{aligned}\partial_\mu A_\nu - \partial_\nu A_\mu \longrightarrow\ & g^{-1}\big(\partial_\mu A_\nu - \partial_\nu A_\mu\big)g + i\big[g^{-1}\partial_\mu g, g^{-1}\partial_\nu g\big] \\ & + \big[g^{-1}A_\mu g, g^{-1}\partial_\nu g\big] + \big[g^{-1}\partial_\mu g, g^{-1}A_\nu g\big].\end{aligned} \tag{6.7}$$

This however suggests a wiser choice,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + i\big[A_\mu, A_\nu\big], \tag{6.8}$$

with the much nicer (i.e., covariant) transformation

$$F_{\mu\nu} \longrightarrow F'_{\mu\nu} = g^{-1}F_{\mu\nu}g. \tag{6.9}$$

Notice that, similar to $A_\mu$, the field strenght $F_{\mu\nu}$ takes values in the algebra of generators, so we can write $F_{\mu\nu} = F^a_{\mu\nu}T^a_{\mathbf{R}}$, with the components given by

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc}A^b_\mu A^c_\nu, \tag{6.10}$$

where $f^{abd}$ are the structure constants of the Lie algebra of generators, $[T^a_{\mathbf{R}}, T^b_{\mathbf{R}}] = if^{abc}T^c_{\mathbf{R}}$.

We denote by $\mathscr{G}$ the set of gauge transformations acting on the fields. Although to fix ideas here, we have considered transformations (6.1) in the connected component of the identity $\mathscr{G}_0$, the derived expressions remain valid for all transformations in $\mathscr{G}$, even if they lie in disconnected components (we saw an example of this in the case of the Lorentz group studied in page 41). For transformations in $\mathscr{G}_0$, we can write their infinitesimal form,

$$g(\alpha) \simeq \mathbb{1} + i\alpha^a T^a_{\mathbf{R}}, \tag{6.11}$$

to write the first order transformation of both the gauge field and its field strength

$$\begin{aligned}\delta_\alpha A^a_\mu &= \partial_\mu \alpha^a + if^{abc}\alpha^b A^c_\mu \equiv (D_\mu \alpha)^a, \\ \delta_\alpha F^a_{\mu\nu} &= if^{abc}\alpha^b F^c_{\mu\nu},\end{aligned} \tag{6.12}$$

where in the first line we expressed the variation of the gauge field in terms of the (adjoint) covariant derivative of the gauge function. The field strength, in turn, can be also recast as the commutator of two covariant derivatives, $F_{\mu\nu} = [D_\mu, D_\nu]$.

After all these preliminaries, we can write a gauge invariant action for fermions coupled to non-Abelian gauge fields,

$$\begin{aligned} S_{\text{YM}} &= \int d^4x \left[ -\frac{1}{2g_{\text{YM}}^2} \text{tr}\left(F_{\mu\nu}F^{\mu\nu}\right) + \overline{\psi}\left(i\not{D} - m\right)\psi \right] \\ &= \int d^4x \left[ -\frac{1}{4g_{\text{YM}}^2} F^a_{\mu\nu}F^{a\mu\nu} + \overline{\psi}\left(i\not{\partial} - m\right)\psi + A^a_\mu \overline{\psi}\gamma^\mu T^a_{\mathbf{R}}\psi \right], \end{aligned} \tag{6.13}$$

where $g_{\text{YM}}$ is the only coupling constant of the theory[16]. This non-Abelian generalization of QED was first formulated by C. N. Yang and Robert L. Mills [92]. Yang–Mills (YM) theories are the backbone of our understanding of elementary particle physics. Although the action $S_{\text{YM}}$ reduces to that of QED in Eq. (4.58) for $G = U(1)$, it displays a much richer structure for non-Abelian gauge groups. For starters, the commutator in the field strength (6.8) is nonzero and the $F^a_{\mu\nu}F^{a\mu\nu}$ term in Eq. (6.13) contains cubic and quartic gauge field self-interaction terms. This indicates that, unlike the photon, non-Abelian gauge bosons are never free particles even if uncoupled to matter.

The general analysis of gauge invariance follows in many aspects the Abelian case. The corresponding electric and magnetic fields are defined in terms of the gauge potential $A^a_\mu \equiv (A^a_0, -\mathbf{A}^a)$ by

$$\begin{aligned} \mathbf{E}^a &= -\boldsymbol{\nabla} A^a_0 - \frac{\partial \mathbf{A}^a}{\partial t} + f^{abc} A^a_0 \mathbf{A}^b, \\ \mathbf{B}^a &= \boldsymbol{\nabla} \times \mathbf{A}^a + f^{abc} \mathbf{A}^b \times \mathbf{A}^c, \end{aligned} \tag{6.14}$$

and, unlike their Abelian counterparts, they are not gauge invariant. The electric field $\mathbf{E}^a$ is in fact the momentum canonically conjugate to $\mathbf{A}^a$,

$$\left\{ A^a_i(t,\mathbf{r}), E^b_j(t,\mathbf{r}') \right\}_{\text{PB}} = \delta_{ij}\delta^{ab}\delta^{(3)}(\mathbf{r}-\mathbf{r}'), \tag{6.15}$$

and the Hamiltonian reads

$$H = \int d^3x \left[ \frac{1}{2}\mathbf{E}^a \cdot \mathbf{E}^a + \frac{1}{2}\mathbf{B}^a \cdot \mathbf{B}^a + A^a_0 \left(\mathbf{D}\cdot\mathbf{E}\right)^a \right]. \tag{6.16}$$

Similarly to Maxwell's electrodynamics, $A^a_0$ plays the role of a Lagrange multiplier enforcing the Gauss law constraint, now reading

$$(\mathbf{D}\cdot\mathbf{E})^a \equiv \boldsymbol{\nabla}\cdot\mathbf{E}^a + f^{abc}\mathbf{A}^b \times \mathbf{E}^c = 0. \tag{6.17}$$

In the quantum theory, classical fields are replaced by operators. Using the non-Abelian version of the temporal gauge, $A^a_0 = 0$, residual gauge transformations correspond to time-independent gauge

[16] The factors of $g_{\text{YM}}$ in front of the first term in the action can be removed by a rescale $A_\mu \to g_{\text{YM}}A_\mu$. In doing so, an inverse power of the coupling constant appears in the derivative terms in Eq. (6.6) and the first identity in Eq. (6.12), while the commutator in Eq. (6.8) acquires a power of $g_{\text{YM}}$, as well as the structure constant term in Eq. (6.10).

functions $\alpha^a(\mathbf{r})$ and are generated by $\mathbf{D}\cdot\mathbf{E}$,

$$\begin{aligned}\delta_\alpha \mathbf{A}(t,\mathbf{r}) &= i\left[\int d^3r\,\alpha^a(\mathbf{r})(\mathbf{D}\cdot\mathbf{E})^a, \mathbf{A}(t,\mathbf{r})\right]\\ &= \boldsymbol{\nabla}\alpha^a + if^{abc}\alpha^b\mathbf{A}^c \equiv (\mathbf{D}\alpha)^a,\end{aligned} \tag{6.18}$$

where we have used the canonical commutation relations derived from Eq. (6.15) and to avoid boundary terms after integration by parts we need to restrict to "small" gauge transformations where $\alpha^a(\mathbf{r})$ vanishes when $|\mathbf{r}|\to\infty$. Those in the connected component of the identity $\mathscr{G}_0$ are therefore implemented on the space of physical states by the operator

$$\mathscr{U}(\alpha) = \exp\left[i\int d^3r\,\alpha^a(\mathbf{r})(\mathbf{D}\cdot\mathbf{E})^a\right]. \tag{6.19}$$

As in the Abelian case discussed in Box 9 (see page 71), the invariance under these "small" gauge transformations has to be preserved at all expenses to avoid unphysical states entering the theory's spectrum. To achieve this, we require that the Gauss law annihilates physical states:

$$(\mathbf{D}\cdot\mathbf{E})^a|\text{phys}\rangle = 0. \tag{6.20}$$

In the presence of non-Abelian sources, $(\mathbf{D}\cdot\mathbf{E})^a$ gets replaced by $(\mathbf{D}\cdot\mathbf{E})^a - \rho^a$, with $\rho^a$ the matter charge density operator.

We should not forget about "large" gauge transformations whose gauge parameter $\alpha^a(\mathbf{r})$ does not vanish when $|\mathbf{r}|\to\infty$. Notice that any transformation of this kind can be written as

$$g(\mathbf{r})_{\text{large}} = hg(\mathbf{r})_{\text{small}}, \tag{6.21}$$

where $h\neq\mathbb{1}$ is a rigid transformation such that $g(\mathbf{r})_{\text{large}}\to h$ as $|\mathbf{r}|\to\infty$. They build up what can be called a copy of the group at infinity, $G_\infty$, the global invariance leading to charge conservation by the first Noether theorem. This is a real symmetry that quantum mechanically can be realized either à la Wigner–Weyl or à la Nambu–Goldstone. For the SM gauge group SU(3) × SU(2) × U(1), the color $\text{SU(3)}_\infty$ symmetry remains unbroken by the vacuum, whereas due to the BEH mechanism the electroweak factor $[\text{SU(2)}\times\text{U(1)}]_\infty$ is partially realized à la Nambu–Goldstone, with a preserved $\text{U(1)}_\infty$ corresponding to the global invariance of electromagnetism[17].

## 7 Anomalous symmetries

In Section 5, we mentioned the possibility that classical symmetries or invariances could somehow turn out to be incompatible with the process of quantization but so far did not elaborate any further. Since anomalous symmetries are crucial in our understanding of a number of physical phenomena, it is about time to look into anomalies in some detail (see Refs. [93–96] for some reviews on the topic).

[17] As we will see shortly, the unbroken U(1) generator is a mixture of the two generators of the Cartan subalgebra of the electroweak SU(2) × U(1) gauge group factor.

### 7.1 Symmetry vs. the quantum

Let us go back to the QED action Eq. (4.58). We have already discussed the global phase invariance leading by the first Noether theorem to the conserved current (4.57). In addition, we can also consider the transformations

$$\psi \longrightarrow e^{i\alpha\gamma_5}\psi, \qquad \overline{\psi} \longrightarrow \overline{\psi}e^{i\alpha\gamma_5}, \tag{7.1}$$

where $\gamma_5$ is the chirality matrix defined in Eq. (4.53). Unlike the transformation $\psi \to e^{i\vartheta}\psi$ rotating the positive and negative chirality components of the Dirac spinor by the same phase, in Eq. (7.1) they change by opposite phases. In what follows, we refer to the first type as *vector* transformations, while the second we dub as *axial-vector*. The latter, however, are not a symmetry of the QED action for $m \neq 0$, since $\overline{\psi}\psi \to \overline{\psi}e^{2i\alpha\gamma_5}\psi \neq \overline{\psi}\psi$, whereas $\overline{\psi}\gamma^\mu\partial_\mu\psi$ is invariant. In fact, using the Dirac field equations it can be shown that the axial-vector current

$$j_5^\mu = \overline{\psi}\gamma_5\gamma^\mu\psi \tag{7.2}$$

satisfies the relation

$$\partial_\mu j_5^\mu = 2im\overline{\psi}\gamma_5\psi \tag{7.3}$$

and for $m = 0$ gives the conservation equation associated with the invariance of massless QED under axial-vector transformations. Similar to what we found on Box 8 for the flavor symmetry of QCD, in this limit the global $\mathrm{U}(1)_V$ symmetry of QED gets enhanced to $\mathrm{U}(1)_V \times \mathrm{U}(1)_A$.

In the quantum theory, Noether currents are constructed as products of field operators evaluated at the same spacetime point. These quantities are typically divergent and it is necessary to introduce some regularization in order to make sense of them. In the case of QED one way to handle the vector current $j^\mu(x) = \overline{\psi}(x)\gamma^\mu\psi(x)$ is by using point splitting

$$j^\mu(x,\epsilon)_{\mathrm{reg}} \equiv \overline{\psi}\left(x - \frac{1}{2}\epsilon\right)\gamma^\mu\psi\left(x + \frac{1}{2}\epsilon\right)\exp\left(ie\int_{x-\frac{1}{2}\epsilon}^{x+\frac{1}{2}\epsilon} dx^\mu A_\mu\right), \tag{7.4}$$

where the divergences appear as poles in $\epsilon = 0$. Notice that since the phases introduced by the gauge transformations of the two fields are evaluated at different points, an extra Wilson line term is needed to restore gauge invariance of the regularized current. Alternatively, we can use Pauli–Villars (PV) regularization, where a number of spurious fermion fields of masses $M_i$ are added to the action

$$S_{\mathrm{reg}} = \int d^4x \left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \overline{\psi}(i\not{D} - m)\psi + \sum_{i=1}^{n} c_k\overline{\Psi}_k(i\not{D} - M_k)\Psi_k\right], \tag{7.5}$$

with $n$ and $c_k$ chosen so that the limit

$$j^\mu(x)_{\mathrm{reg}} \equiv \lim_{x'\to x}\left[\overline{\psi}(x')\gamma^\mu\psi(x) + \sum_{k=1}^{n} c_k\overline{\Psi}_k(x')\gamma^\mu\Psi_k(x)\right] \tag{7.6}$$

remains finite (i.e., all poles at $x - x' = 0$ cancel). An important feature of the PV regularization is that it explicitly preserves gauge invariance. The masses $M_k$ act as regulators, since in the limit $M_k \to \infty$ the PV fermions decouple and the original divergences reapear.

The need to make sense of composite operators is at the core of the potential problems with current conservation in the quantum domain. The regularization procedure might collide with some of the classical symmetries of the theory, resulting in its breaking after divergences are properly handled. This is why our discussion of the regularization of the current operator in QED has been conspicuously concerned with the issue of gauge invariance of the vector current. The existence of gauge invariant regularization schemes guarantees that the current coupling to the gauge field can be defined in the quantum theory without spoiling its conservation $\partial_\mu j^\mu = 0$ at operator level. Otherwise, we would be in serious trouble, as we can see by applying the quantization prescription Eq. (3.66) to the stability condition of the Gauss law Eq. (3.54),

$$[G, H] = -i\partial_\mu j^\mu, \tag{7.7}$$

where we have defined $G \equiv \boldsymbol{\nabla} \cdot \mathbf{E} - j^0$. If $\partial_\mu j^\mu \neq 0$, the Gauss law condition ensuring the factorization of redundant states would not be preserved by time evolution. Indeed, imposing the constraint at $t = 0$ on some state, $G|\Psi(0)\rangle = 0$, we would have at first order in $\delta t$

$$G|\Psi(\delta t)\rangle = -i\delta t G H|\Psi(0)\rangle = -\delta t \partial_\mu j^\mu |\Psi(0)\rangle \neq 0, \tag{7.8}$$

so the constraint is no longer satisfied and unphysical states enter the spectrum. Another sign that something goes wrong when implementing the Gauss law constraint in theories with gauge anomalies appears when computing the commutator of two $G$'s evaluated at different points. In the presence of a gauge anomaly, it is no longer zero [97–99], but

$$[G(\mathbf{r}), G(\mathbf{r}')] = c\mathbf{B}(\mathbf{r}) \cdot \boldsymbol{\nabla} \delta^{(3)}(\mathbf{r} - \mathbf{r}'), \tag{7.9}$$

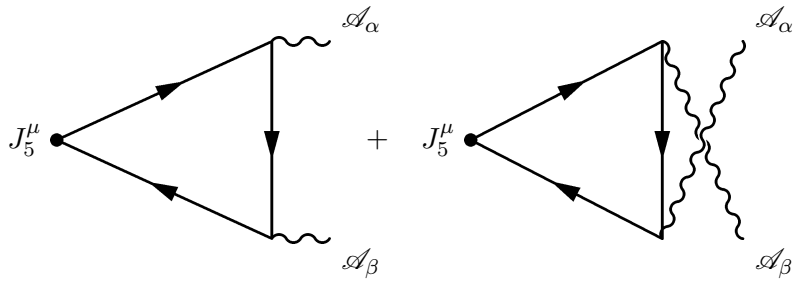
where $c \neq 0$ is a constant determined by the value of $\partial_\mu j^\mu$. This result implies that $G(\mathbf{r})|\text{phys}\rangle = 0$ cannot be consistently imposed, since this condition would imply $[G(\mathbf{r}), G(\mathbf{r}')]|\text{phys}\rangle = 0$ whereas the right-hand side of Eq. (7.9) gives a nonzero result when acting on the state[18]. This being the case, spurious states cannot be factored out from the spectrum, with the upshot that the theory becomes inconsistent.

This shows that in constructing QFTs, gauge anomalies cannot emerge. This condition is a very powerful constraint in model building, since it limits both the type of fields that can be allowed in the actions and also their couplings. As we will see in Box 13 in page 104, in the SM this requirement completely fixes the hypercharges of quarks and leptons, up to a global normalization (see Ref. [96] for examples of anomaly cancellation in the SM and beyond).

After this digression, we go back to the quantum mechanical definition of the axial-vector current Eq. (7.2) and the fate of its (pseudo)conservation Eq. (7.3). To simplify things, we consider the massless

[18] Something similar happens in the case of non-Abelian gauge theories that we will discuss in the next section. There, the commutator of two Gauss law operators acquires a central extension, $[G^a(\mathbf{r}), G^b(\mathbf{r}')] = if^{abc}G^c(\mathbf{r})\delta^{(3)}(\mathbf{r}-\mathbf{r}') + \mathcal{A}^{ab}(\mathbf{r}, \mathbf{r}')$, with $G^a \equiv (\mathbf{D} \cdot \mathbf{E})^a - j^{a0}$ in this case.

case where axial-vector transformations Eq. (7.1) are a symmetry of the classical action. A very convenient way to study this problem is to treat the gauge field as a classical external source coupling to the quantum Dirac field. This is made clear by denoting gauge fields and field strengths using calligraphic fonts as $\mathscr{A}_\mu$ and $\mathscr{F}_{\mu\nu}$, respectively. Instead of working with operators, we deal with their vacuum expectation values in the presence of the background field and compute $\langle J_5^\mu \rangle_{\mathscr{A}} \equiv \langle 0|J_5^\mu|0\rangle$ together with its divergence. This can be done using either the regularized operators introduced above (see, for example, Ref. [100] for a calculation using point-splitting regularization) or diagrammatic techniques. In the latter case, we need to compute the celebrated triangle diagrams




(7.10)

where in the left vertex of both diagrams (indicated by a dot) an axial-vector current is inserted, whereas the other two are coupled to the external gauge field through the vector gauge currents. Since in these lectures we are not entering into the computation of Feynman graphs, we will not elaborate on how to calculate these ones. Details can be found in Chapter 9 of Refs. [14] or in [94]. Here we just give the final result for the anomaly of the axial-vector current,

$$\partial_\mu \langle J_5^\mu \rangle_{\mathscr{A}} = -\frac{e^2\hbar}{16\pi^2}\epsilon^{\mu\nu\alpha\beta}\mathscr{F}_{\mu\nu}\mathscr{F}_{\alpha\beta}. \tag{7.11}$$

Despite having used all the time natural units with $\hbar = 1$, in this expression we have restored the powers of the Planck constant to make explicit the fact that the anomaly is a pure quantum effect.

This crucial result has a long history. The diagrams in Eq. (7.10) were computed in 1949 by Jack Steinberger [101] and later in 1951 by Julian Schwinger [102], in both cases in the context of the electromagnetic decay of neutral mesons[19]. Almost two decades later, the consequences of the triangle diagram for the quantum realization of the axial-vector symmetry of QED were pointed out by Stephen Adler [105], and John S. Bell and Roman Jackiw [106] in what are considered today the foundational papers of the subject of quantum anomalies.

There are some very important issues that should be mentioned concerning the calculation of the axial anomaly Eq. (7.11). We have stressed how the anomaly could be seen as originated by the need to regularize UV (i.e., short distance) divergences in the definition of the current or, alternatively, in the computation of the triangle diagrams. Nevertheless, using either method, we find a regular result in the limit in which the regulator is removed. In the language of QFT, we do not need to subtract and renormalize divergences to find the anomaly of the axial current. At the level of diagrams, what happens is that, although the integrals are linearly divergent, this only results in an ambiguity in their

[19] Other early calculations of the triangle diagrams were carried out in 1949 by Hiroshi Fukuda and Yoneji Miyamoto [103], and by S. Ozaki, S. Oneda, and S. Sasaki [104].

value that is fixed by requiring the gauge (vector) current to be conserved. In the case of the point splitting calculation, introducing a Wilson line similar to the one inserted in Eq. (7.4) in the regularized definition of the axial-vector current to preserve gauge invariance we are led to the axial anomaly after taking the $\epsilon \to 0$ limit.

Another important point to be stressed is a *tension* between the conservation of the gauge and the axial-vector currents: we can impose the conservation of either of the two, *but not of both simultaneously*. After the above discussion of the dire consequences of violating gauge current conservation, the choice is clear enough.

### 7.2 The physical power of the anomaly

When studying the global symmetries of QCD, we have also encountered axial transformations [see Box 8 and in particular Eq. (5.48)] and mentioned that they are anomalous. Now we can be more explicit. The axial-vector current of interest in this case is given by

$$J_5^\mu = \overline{\boldsymbol{q}}\gamma_5\gamma^\mu \boldsymbol{q}, \tag{7.12}$$

where a sum over color indices should be understood. Its anomaly comes from triangle diagrams similar to the ones shown in Diagram (7.10), this time with quarks running in the loop. But, together with the triangles coupling to the electromagnetic external potential $\mathscr{A}_\mu$, we also have a pair of triangles where the vertices on the right couple to an external gluon field $\mathcal{A}_\mu^a$ (for this, we also use calligraphic fonts to indicate that we are dealing with classical sources). This results in the anomaly

$$\partial_\mu \langle J_5^\mu \rangle_{\mathscr{A},\mathcal{A}} = -\frac{N_c}{16\pi^2}\left(\sum_{f=1}^{N_f} q_f^2\right)\epsilon^{\mu\nu\alpha\beta}\mathscr{F}_{\mu\nu}\mathscr{F}_{\alpha\beta} - \frac{N_f}{16\pi^2}\epsilon^{\mu\nu\alpha\beta}\mathcal{F}^a_{\mu\nu}\mathcal{F}^a_{\alpha\beta}, \tag{7.13}$$

where $\mathcal{F}^a_{\mu\nu}$ is the non-Abelian field strength associated with the external gluon field and $N_c$ is the number of colors. The coefficient of the first term is obtained by summing the expression of the axial anomaly given in (7.11) to all quarks running in the loop. As for the second, the quarks couple to the gluon fields through the gauge current

$$J^{\mu a} = \overline{\boldsymbol{q}}\gamma^\mu \tau^a \boldsymbol{q}, \tag{7.14}$$

where $\tau^a$ are the generators of the fundamental representation of SU(3) acting on the color indices of each component of $\boldsymbol{q}$. Since the axial current does not act on color indices, the prefactor is proportional to $(\mathrm{tr}\,\mathbb{1})(\mathrm{tr}\,\{\tau^a,\tau^b\}) = N_f\delta^{ab}$, with $\mathbb{1}$ the identity in flavor space.

Anomalies can also affect the global non-Abelian $\mathrm{SU}(N_f)_L \times \mathrm{SU}(N_f)_R$ symmetry defined in (5.49). This global symmetry group can be rearranged in terms of vector and axial transformations $\mathrm{SU}(N_f)_L \times \mathrm{SU}(N_f)_R = \mathrm{SU}(N_f)_V \times \mathrm{SU}(N_f)_A$ acting on the quark fields as

$$\mathrm{SU}(N_f)_V : \boldsymbol{q} \to e^{i\alpha_V^I t_{\mathbf{f}}^I}\boldsymbol{q}, \qquad \mathrm{SU}(N_f)_A : \boldsymbol{q} \to e^{i\alpha_A^I t_{\mathbf{f}}^I \gamma_5}\boldsymbol{q}, \tag{7.15}$$

with $\boldsymbol{q}_R$ and $\boldsymbol{q}_L$ transforming respectively with the same or opposite SU($N_f$) parameters[20]. Vector currents, however, are always anomaly-free. A simple way to come to this conclusion is to notice that the PV regularization method introduced above preserved all vector symmetries, since these remain unbroken by fermion mass terms[21]. We thus focus on the chiral SU$(N_f)_A$ factor, whose associated axial-vector current is

$$J_5^{I\mu} = \overline{\boldsymbol{q}}\gamma_5\gamma^\mu t_{\mathbf{f}}^I \boldsymbol{q}, \tag{7.16}$$

where, again, there is a tacit sum over the quark color index. As in the case of the singlet current (7.12), there are contributions coming from the photon and gluon couplings of the quarks. Taking into account that, unlike photons, gluons are flavor-blind, we find

$$\partial_\mu \langle J_5^{I\mu}\rangle_{\mathscr{A},\mathcal{A}} = -\frac{N_c}{16\pi^2}\left[\sum_{f=1}^{N_f} q_f^2 (t_{\mathbf{f}}^I)_{ff}\right]\epsilon^{\mu\nu\alpha\beta}\mathscr{F}_{\mu\nu}\mathscr{F}_{\alpha\beta} - \frac{N_f}{16\pi^2}\left(\operatorname{tr} t_{\mathbf{f}}^I\right)\epsilon^{\mu\nu\alpha\beta}\mathcal{F}^a_{\mu\nu}\mathcal{F}^a_{\alpha\beta}. \tag{7.17}$$

Since all generators of SU($N_f$) are traceless, the second term is zero but the first one does not necessarily vanish.

Let be focus on the dynamics of the two lightest quarks $u$ and $d$, where $q_u = \frac{2}{3}e$ and $q_d = -\frac{1}{3}e$. In this case $N_f = 2$ and the flavor group is generated by $t_{\mathbf{f}}^I = \frac{1}{2}\sigma^I$, with $\sigma^I$ the Pauli matrices. We have then

$$\sum_{f=1}^{2} q_f^2 (t_{\mathbf{f}}^1)_{ff} = \sum_{f=1}^{2} q_f^2 (t_{\mathbf{f}}^2)_{ff} = 0, \qquad \sum_{f=1}^{2} q_f^2 (t_{\mathbf{f}}^3)_{ff} = \frac{e^2}{6}, \tag{7.18}$$

where $N_c$ is the number of quark colors. This means that $J_5^{3\mu}$ is anomalous,

$$\partial_\mu \langle J_5^{3\mu}\rangle_{\mathscr{A},\mathcal{A}} = -\frac{e^2 N_c}{48\pi^2}\epsilon^{\mu\nu\alpha\beta}\mathscr{F}_{\mu\nu}\mathscr{F}_{\alpha\beta}. \tag{7.19}$$

The physical importance of this result lies in that after chiral symmetry breaking (see Box 8 in page 65), the operator $\partial_\mu J_5^{a\mu}$ becomes the interpolating field for pions, creating them out of the vacuum[22]

$$\langle \pi^a(p)|\partial_\mu J_5^{a\mu}(x)|0\rangle = f_\pi m_\pi \delta^{ab} e^{-ip\cdot x} \qquad \Longrightarrow \qquad \pi^a(x) = \frac{1}{f_\pi m_\pi}\partial_\mu J_5^{a\mu}(x), \tag{7.20}$$

where $m_\pi$ is the pion mass and $f_\pi$ the pion decay constant introduced in Eq. (5.54) to parametrize the matrix of NG bosons resulting from chiral symmetry breaking. Although to compute the anomaly (7.19) we took the electromagnetic field to be a classical source, the corresponding operator identity implies the

[20] A warning note here. Unlike the Abelian U(1)$_A$, transformations in SU$(N_f)_A$ do not close and therefore do not form a group. This can be checked by composing two of them and applying the Baker–Campbell–Hausdorff formula. Our notation has to be understood in a formal sense.

[21] This argument also applies to the SU(3) gauge invariance of QCD, which cannot be anomalous since it acts in the same way on quarks of both chiralities. As a consequence, the theory can be regularized in a gauge invariant way.

[22] The first identity follows from $\langle \pi^a(p)|J_5^{b\mu}(x)|0\rangle \sim p^\mu \delta^{ab} e^{-ip\cdot x}$, a direct consequence of the Goldstone theorem [79].

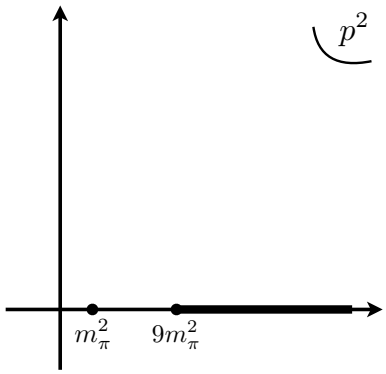



**Fig. 10:** Complex $p^2$-plane showing the structure of singularities of the function $f(p^2)$ in Eq. (7.24): a pole at $p^2 = m_\pi^2$ and a branch cut beginning at $p^2 = 9m_\pi^2$.

existence of a nontrivial overlap between the neutral pion state and the state with two photons,

$$\langle \mathbf{k}_1, \lambda_1; \mathbf{k}_2, \lambda_2 | \pi^0(p) \rangle = \frac{e^2 N_c}{12\pi^2 f_\pi} (2\pi)^4 \delta^{(4)}(p - k_1 - k_2) \epsilon_{\mu\nu\alpha\beta} k_1^\mu k_2^\nu \epsilon^\alpha(\mathbf{k}_1) \epsilon^\beta(\mathbf{k}_2). \tag{7.21}$$

The width of the process can be computed from this result to be

$$\Gamma(\pi^0 \to 2\gamma) = \frac{\alpha^2 N_c^2 m_\pi^3}{576\pi^3 f_\pi^2} = 7.73\,\text{eV}, \tag{7.22}$$

which is perfectly consistent with experimental measurements [107]

$$\Gamma(\pi^0 \to 2\gamma)_{\text{exp}} = 7.798 \pm 0.056\ (\text{stat.}) \pm 0.109\ (\text{syst.})\ \text{eV}. \tag{7.23}$$

Incidentally, the presence of $f_\pi = 93\,\text{MeV}$ in Eq. (7.22) gives a rationale for it being called the pion decay constant.

The electromagnetic decay of the neutral pion is a direct consequence of the existence of the axial anomaly. On general grounds, it can be argued that the amplitude for the decay process of the $\pi^0$ into two photons has the structure

$$\langle \mathbf{k}_1, \lambda_1; \mathbf{k}_2, \lambda_2 | \pi^0(p) \rangle = i \frac{p^2 - m_\pi^2}{f_\pi m_\pi^2} p^2 f(p^2) (2\pi)^4 \delta^{(4)}(p - k_1 - k_2) \epsilon_{\mu\nu\alpha\beta} k_1^\mu k_2^\nu \epsilon^\alpha(\mathbf{k}_1) \epsilon^\beta(\mathbf{k}_2), \tag{7.24}$$

with $f(p^2)$ a function of the pion squared momentum. We could naively assume $f(p^2)$ to be well-behaved, with a pole singularity at $p^2 = m_\pi^2$ and a branch cut starting at $9m_\pi^2$ signalling multi-pion production (see Fig. 10). Were this the case, the amplitude would be suppressed in the $p^2 \to 0$ limit. Historically, this result was known as the Sutherland–Veltman theorem [108, 109] and essentially ruled out the existence of the process $\pi^0 \to 2\gamma$, that was nevertheless observed. The catch lies in that the regularity hypothesis concerning $f(p^2)$, called partial conservation of the axial current (PCAC), is wrong due to the axial anomaly. The calculation of the triangle diagrams (7.10) shows that this function is not

regular at zero momentum, but actually has a pole

$$f(p^2) \sim \frac{ie^2 N_c}{12\pi}\frac{1}{p^2} \qquad \text{as} \qquad p \to 0. \tag{7.25}$$

This singularity is precisely responsible for compensating the low-momentum suppression of the amplitude (7.24), giving the nonzero result accounting for the $\pi^0 \to 2\gamma$ decay. It is somewhat fascinating that the anomaly, that we identified from the start as resulting from UV ambiguities in the definition of the current, is also associated with an IR pole and determined by its residue. This reflects the profound topological connections of QFT anomalies [93–96].

**Box 10. The path integral way to the anomaly**

There are many different roads leading to the chiral anomaly. For our presentation above we have chosen the perturbative approach, involving the computation of the two one-loop triangle diagrams shown in Eq. (7.10). But the anomaly can also be computed using path integrals, where it appears as a result of the noninvariance of the functional measure under chiral rotations of the Dirac fermions.

To see how this comes about, let us consider again a Dirac fermion coupled to an external electromagnetic field $\mathscr{A}_\mu$ that we treat as a classical source. Its action is given by

$$\begin{aligned} S[\psi,\overline{\psi},\mathscr{A}_\mu] &= \int d^4x \overline{\psi}\gamma^\mu\big(i\partial_\mu + e\mathscr{A}_\mu\big)\psi \\ &= \int d^4x \Big[\overline{\psi}_R\big(i\partial_\mu + e\mathscr{A}_\mu\big)\psi_R + \overline{\psi}_L\big(i\partial_\mu + e\mathscr{A}_\mu\big)\psi_L\Big], \end{aligned} \tag{7.26}$$

where in the second line we split the Dirac fermion into its two chiralities. A quantum effective action $\Gamma[\mathscr{A}_\mu]$ for the external field can be defined by integrating out the fermions

$$e^{i\Gamma[\mathscr{A}_\mu]} = \int \mathscr{D}\overline{\psi}\mathscr{D}\psi\, e^{iS[\psi,\overline{\psi},\mathscr{A}_\mu]}. \tag{7.27}$$

The important point in this expression is that the Dirac fields are dummy variables that can be modified without changing the value of the functional integral. In particular, we can implement the following "change of variables":

$$\psi = e^{i\alpha\gamma_5}\psi' \qquad \Longrightarrow \qquad \psi_{R,L} = e^{\pm i\alpha}\psi'_{R,L}, \tag{7.28}$$

writing the original Dirac field in terms of its chiral-transform [see Eq. (7.1)]. As we know, in the absence of a Dirac mass term the fermion action does not change

$$S[\psi,\overline{\psi},\mathscr{A}_\mu] = S[\psi',\overline{\psi}',\mathscr{A}_\mu], \tag{7.29}$$

reflecting the classical chiral invariance of the massless theory.

However, we have to be careful when implementing this change in the integral (7.27). The reason is that we have to properly transform the fermion integration measure, which in principle

might pick up a nontrivial Jacobian. Since the transformation is linear in the fermions, this Jacobian can only depend on the external sources, as well as on the transformations parameter $\alpha$,

$$\mathscr{D}\overline{\psi}\mathscr{D}\psi = J[\mathscr{A}_\mu]\mathscr{D}\overline{\psi}'\mathscr{D}\psi'. \tag{7.30}$$

Taking this into account, we go back to (7.27) that now reads

$$e^{i\Gamma[\mathscr{A}_\mu]} = \int \mathscr{D}\overline{\psi}'\mathscr{D}\psi'\, e^{iS[\psi',\overline{\psi}',\mathscr{A}_\mu]+\log J[\mathscr{A}_\mu]} \equiv \int \mathscr{D}\overline{\psi}'\mathscr{D}\psi'\, e^{iS'[\psi',\overline{\psi}',\mathscr{A}_\mu]}. \tag{7.31}$$

Thus, the effective action can be computed in the new variables, provided we use the new fermion action $S'[\psi',\overline{\psi}',\mathscr{A}_\mu]$ including an additional term,

$$S'[\psi',\overline{\psi}',\mathscr{A}_\mu] = \int d^4x\, \overline{\psi}'\gamma^\mu\big(i\partial_\mu + e\mathscr{A}_\mu\big)\psi' - i\log J[\mathscr{A}_\mu], \tag{7.32}$$

that, coming from the functional measure, is obviously a pure quantum effect. A convenient way to compute the Jacobian is by expanding the Dirac fermions in a basis of Dirac operator $\not{D}(\mathscr{A}) \equiv \gamma^\mu(\partial_\mu - ie\mathscr{A}_\mu)$ eigenstates. Using a regularization method preserving gauge invariance, a finite result is obtained [95, 110, 111]:

$$-i\log J[\mathscr{A}_\mu] = \frac{e^2\alpha}{16\pi^2}\int d^4x\, \epsilon^{\mu\nu\alpha\beta}\mathscr{F}_{\mu\nu}\mathscr{F}_{\alpha\beta}. \tag{7.33}$$

Notice that in the case of massive fermions the change (7.28) also introduces, besides the quantum anomalous term, a complex phase in the mass, which has a classical origin:

$$\begin{aligned} S'[\psi',\overline{\psi}',\mathscr{A}_\mu] = \int d^4x \Big[&\overline{\psi}'_R\gamma^\mu\big(i\partial_\mu + e\mathscr{A}_\mu\big)\psi'_R + \overline{\psi}'_L\gamma^\mu\big(i\partial_\mu + e\mathscr{A}_\mu\big)\psi'_L \\ &+ me^{2i\alpha}\big(\overline{\psi}'_R\psi'_L + \overline{\psi}'_L\psi'_R\big)\Big] + \frac{e^2\alpha}{16\pi^2}\int d^4x\, \epsilon^{\mu\nu\alpha\beta}\mathscr{F}_{\mu\nu}\mathscr{F}_{\alpha\beta}. \end{aligned} \tag{7.34}$$

The last term associated to the nonzero Jacobian is just the integrated form of the chiral anomaly found in (7.11). The analysis just presented will be useful in analyzing the strong CP problem in the next section.

## 8 The strong CP problem and axions

When studying magnetic monopoles in Box 5 (see page 27), we discussed the possibility of having non-trivial gauge field topologies. In this section, we are going to look deeper into the role played by topology in non-Abelian gauge field theories and study how nonequivalent topological gauge field configurations define different vacua of the theory.

### 8.1 The (infinitely) many vacua of QCD

To fix ideas, let us consider pure YM theory in the temporal gauge $A^a_0 = 0$, preserved by the set $\mathscr{G}$ of time-independent gauge transformations $g(\mathbf{r})$. Adding to the Euclidean space $\mathbb{R}^3$ the point at infinity,

it gets compatified to a three-sphere, $\mathbb{R}^3 \cup \{\infty\} \simeq S^3$. Thus, the residual gauge transformations in $\mathscr{G}$ define maps from $S^3$ onto the gauge group[23]:

$$\mathscr{G} : S^3 \longrightarrow G. \tag{8.1}$$

The space $\mathscr{G}$ consists of infinitely topological nonequivalent sectors classified by the third-homotopy group $\pi_3(G)$ [57–60]. As an example, let us consider a gauge theory with group $G = \mathrm{SU}(2)$. This Lie group is topologically equivalent to a three-dimensional sphere $S^3$, as can be seen by writing

$$g = n^0 \mathbb{1} + i\mathbf{n} \cdot \boldsymbol{\sigma}, \tag{8.2}$$

with $n^0$ and $\mathbf{n} = (n^1, n^2, n^3)$ real. Both unitarity

$$g^\dagger g = g g^\dagger = \left[(n^0) + \mathbf{n}^2\right] \mathbb{1} = \mathbb{1}, \tag{8.3}$$

and the requirement of unit determinant

$$\det g = (n^0)^2 + \mathbf{n}^2 = 1, \tag{8.4}$$

lead to the condition

$$(n^0)^2 + \mathbf{n}^2 = 1, \tag{8.5}$$

so $(n^0, \mathbf{n})$ parametrizes the unit three-sphere $S^3$. Since $\pi_3(S^3) = \mathbb{Z}$, the set of time-independent SU(2) gauge transformations decomposes into topological nonequivalent sectors

$$\mathscr{G} = \bigcup_{n \in \mathbb{Z}} \mathscr{G}_n, \tag{8.6}$$

where $n$ is the winding number of the map $S^3 \to S^3$. For a gauge transformation $g(\mathbf{r})$, its winding number can be shown to be

$$n = \frac{1}{24\pi^2} \int_{S^3} d^3r \, \epsilon_{ijk} \mathrm{tr} \left[ (g^{-1}\partial_i g)(g^{-1}\partial_j g)(g^{-1}\partial_k g) \right]. \tag{8.7}$$

Moreover, two gauge transformations can be continuously deformed into one another only when they share the same winding number, with $\mathscr{G}_0$ the identity's connected component. Additivity is an important property of the winding number. Given $g \in \mathscr{G}_n$ and $g' \in \mathscr{G}_{n'}$, their product $gg'$ has winding number

$$n_{gg'} = n_g + n_{g'}, \tag{8.8}$$

and in particular $n_{g^{-1}} = -n_g$. This, together with the fact that $\mathbb{1} \in \mathscr{G}_0$, shows that $\mathscr{G}_0$ is the only sector forming a subgroup.

From the discussion in Section 6, we learn that physical states are preserved by "small" gauge

[23] At a more physical level, the compactification of $\mathbb{R}^3$ to $S^3$ amounts to requiring that all fields, as well as gauge transformations, have well-defined limits as $|\mathbf{r}| \to \infty$, independent of the direction along which the limit is taken.

transformations in $\mathscr{G}_0$ provided they satisfy the Gauss law (6.20). As for transformations in $\mathscr{G}_n$ with $n \neq 0$, keeping in mind that quantum states are rays in a Hilbert space defined up to a global complex phase, we conclude that physical invariance under a transformation $g_1 \in \mathscr{G}_1$ requires

$$g_1|\text{phys}\rangle = e^{i\theta}|\text{phys}\rangle, \tag{8.9}$$

for some $\theta \in \mathbb{R}$. This number should be independent of the state, since otherwise gauge transformations would give rise to observable interference. Another relevant fact to notice is that the value of $\theta$ is also independent of the transformation in $\mathscr{G}_1$. To see this, let us consider $g_1, g_1' \in \mathscr{G}_1$ and assume that

$$g_1|\text{phys}\rangle = e^{i\theta}|\text{phys}\rangle, \qquad g_1'|\text{phys}\rangle = e^{i\theta'}|\text{phys}\rangle. \tag{8.10}$$

Since by additivity of the winding number $g_1' g_1^{-1} \in \mathscr{G}_0$, and transformations in the connected component of the identity leave the physical states invariant without any complex phase, we immediately conclude that $\theta' = \theta$. Using a similar argument it is straightforward to show that for $g_n \in \mathscr{G}_n$

$$g_n|\text{phys}\rangle = e^{in\theta}|\text{phys}\rangle. \tag{8.11}$$

The conclusion is that a single actual number $\theta$ determines the action of all gauge transformations on physical states.

We can reach the same conclusion about the vacuum structure of YM theories in a different way. Besides the gauge kinetic term in the action (6.13), there is also a second admissible gauge invariant term

$$\begin{aligned} S_\theta &= -\frac{\theta}{32\pi^2}\int d^4x\, F^a_{\mu\nu}\widetilde{F}^{a\mu\nu} \\ &= -\frac{\theta}{8\pi^2}\int d^4x\, \mathbf{E}^a\cdot\mathbf{B}^a, \end{aligned} \tag{8.12}$$

where $\widetilde{F}^a_{\mu\nu}$ is the non-Abelian analog of the dual tensor field introduced in Eq. (3.45), defined as

$$\widetilde{F}^a_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}F^{a\alpha\beta}. \tag{8.13}$$

What makes the $\theta$-term (8.12) interesting is that it is the integral of a total derivative

$$\epsilon^{\mu\nu\alpha\beta}F^a_{\mu\nu}F^a_{\alpha\beta} = \partial_\mu \mathscr{J}^\mu, \tag{8.14}$$

and therefore does not contribute to the field equations. The current on the right-hand side of the previous equation takes the form (see Box 11 below for a rather simple derivation of this result)

$$\mathscr{J}^\mu = 4\epsilon^{\mu\nu\alpha\beta}\left(A^a_\nu\partial_\alpha A^a_\beta + \frac{1}{3}f^{abc}A^a_\nu A^b_\alpha A^c_\beta\right). \tag{8.15}$$

In the $A_0^a = 0$ gauge, we have

$$\epsilon^{\mu\nu\alpha\beta} F^a_{\mu\nu} F^a_{\alpha\beta} = 4 \frac{\partial}{\partial t} \left[ \mathbf{A}^a \cdot \left( \boldsymbol{\nabla} \times \mathbf{A}^a \right) + \frac{1}{3} f^{abc} \mathbf{A}^a \cdot \left( \mathbf{A}^b \times \mathbf{A}^c \right) \right], \tag{8.16}$$

which, once integrated and with the proper normalization, gives the following expression of the $\theta$-term

$$\begin{aligned} S_\theta = -\frac{\theta}{8\pi^2} \Bigg\{ \int d^3 r \left[ \mathbf{A}^a \cdot \left( \boldsymbol{\nabla} \times \mathbf{A}^a \right) + \frac{1}{3} f^{abc} \mathbf{A}^a \cdot \left( \mathbf{A}^b \times \mathbf{A}^c \right) \right] \Bigg|_{t=\infty} \\ - \int d^3 r \left[ \mathbf{A}^a \cdot \left( \boldsymbol{\nabla} \times \mathbf{A}^a \right) + \frac{1}{3} f^{abc} \mathbf{A}^a \cdot \left( \mathbf{A}^b \times \mathbf{A}^c \right) \right] \Bigg|_{t=-\infty} \Bigg\}. \end{aligned} \tag{8.17}$$

To ensure finiteness, we take the gauge field $\mathbf{A} = \mathbf{A}^a T^a_{\mathbf{R}}$ to approach pure-gauge configurations $\mathbf{A}_\pm = g_\pm^{-1} \boldsymbol{\nabla} g_\pm$ at $t = \pm\infty$ (see Fig. 11). It is easy to see that the integrands in Eq. (8.17) are not gauge invariant and therefore the $\theta$-term is nonzero (again, a derivation is outlined in Box 11),

$$\begin{aligned} S_\theta = \frac{\theta}{24\pi^2} \int d^3 r \, \mathrm{tr} \left\{ (g_+^{-1} \boldsymbol{\nabla} g_+) \cdot \left[ (g_+^{-1} \boldsymbol{\nabla} g_+) \times (g_+^{-1} \boldsymbol{\nabla} g_+) \right] \right\} \\ - \frac{\theta}{24\pi^2} \int d^3 r \, \mathrm{tr} \left\{ (g_-^{-1} \boldsymbol{\nabla} g_-) \cdot \left[ (g_-^{-1} \boldsymbol{\nabla} g_-) \times (g_-^{-1} \boldsymbol{\nabla} g_-) \right] \right\}. \end{aligned} \tag{8.18}$$

Comparing with Eq. (8.7), we identify the winding numbers $n_\pm$ of the asymptotic gauge transformations $g_\pm$, to write

$$S_\theta = (n_+ - n_-)\theta. \tag{8.19}$$

Thus, non-Abelian gauge field configurations are classified into topological sectors interpolating between early and late time configurations of definite winding number $n_\pm$. These sectors are labelled by the integer $n = n_+ - n_-$, and when summing in the Feynman path integral over all gauge configurations we also have to include all possible sectors. Each one is weighted by the same phase,

$$e^{iS_\theta} = e^{in\theta}, \tag{8.20}$$

that we encountered in Eq. (8.11).

**Box 11. Gauge fields and differential forms**

The analysis of YM theories gets very much simplified in the language of differential forms [57–60]. The gauge field $A_\mu = A^a_\mu T^a_{\mathbf{R}}$ can be recast as the Lie algebra valued one-form

$$A = -iA_\mu dx^\mu, \tag{8.21}$$

while the two-form field strength is given by

$$F \equiv -\frac{i}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu = dA + A \wedge A, \tag{8.22}$$

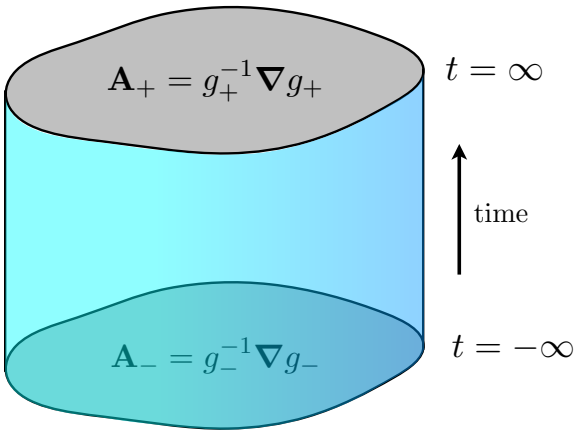



**Fig. 11:** Representation of the spacetime interpolating between two pure gauge configurations $\mathbf{A}_\pm = g_\pm \boldsymbol{\nabla} g_\pm$ at $t = \pm\infty$, in the $A_0 = 0$ gauge.

where in the second term on the right-hand side a matrix multiplication of the one-forms is also understood (in the Abelian case the matrices commute and the term vanishes due to the anticommutativity of the wedge product). The factor of $-i$ in both eqs. (8.21) and (8.22) is introduced to avoid cluttering expressions with powers of $i$.

Gauge transformations are determined by a zero-form $g \in \mathscr{G}$ acting on the gauge field one-form as [cf. (6.6)]

$$A \longrightarrow A' = g^{-1}dg + g^{-1}Ag. \tag{8.23}$$

This leads to the corresponding transformation of the field strength

$$\begin{aligned} F \longrightarrow F' &= dA' + A' \wedge A' \\ &= g^{-1}Fg, \end{aligned} \tag{8.24}$$

that once written in components agrees with the one given in Eq. (6.9). In fact, given an adjoint $p$-form field

$$\Phi_p = -\frac{i}{p!}\Phi_{\mu_1\ldots\mu_p}dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_p} \quad \Longrightarrow \quad \Phi_p \to \Phi_p' = g^{-1}\Phi_p g, \tag{8.25}$$

a covariant exterior derivative is defined acting as

$$D\Phi_p \equiv d\Phi_p + A \wedge \Phi_p - (-1)^p \Phi_p \wedge A \quad \Longrightarrow \quad (D\Phi_p)' = g^{-1}(D\Phi_p)g, \tag{8.26}$$

satisfying the Leibniz rule

$$D(\Phi_p \wedge \Psi_q) = (D\Phi_p) \wedge \Psi_q + (-1)^p \Phi_p \wedge (D\Psi_q). \tag{8.27}$$

Using these definitions and properties, it is easy to check that the field strength two-form (8.22)

verifies the Bianchi identity $DF = 0$.

In four dimensions there are two gauge invariant four-forms that can be constructed from the field-strength two-form. The first one is

$$\mathrm{tr}\,(F \wedge \star F), \tag{8.28}$$

where $\star$ denotes the Hodge dual, acting on a $p$-form field as [58]

$$\star\Phi_p = -\frac{i}{p!(4-p)!}\epsilon^{\mu_1\ldots\mu_p}{}_{\nu_1\ldots\nu_{4-p}}\Phi_{\mu_1\ldots\mu_p}dx^{\nu_1}\wedge\ldots\wedge dx^{\nu_{4-p}}. \tag{8.29}$$

Since this operation commutes with the multiplication by a zero-form, the gauge invariance of (8.28) follows directly from applying the cyclic property of the trace. In addition, we can also construct a second gauge invariant four-form

$$\mathrm{tr}\,(F \wedge F), \tag{8.30}$$

so the action of pure YM theory without matter couplings can be written as

$$S_{\text{YM}} = \frac{1}{2g^2_{\text{YM}}}\int_{\mathcal{M}_4}\mathrm{tr}\,(F\wedge\star F) + \frac{\theta}{8\pi^2}\int_{\mathcal{M}_4}\mathrm{tr}\,(F\wedge F), \tag{8.31}$$

where $\mathcal{M}_4$ represents the four-dimensional spacetime. The two terms correspond respectively to the kinetic and $\theta$ terms given in components in eqs. (6.13) and (8.12). Incidentally, notice that while the term inside the first integral is always a maximal form in any dimension, the one in the second term is only maximal in $D = 4$. In fact, no analog of the $\theta$-term exits in odd-dimensional spacetimes.

Although in these lectures we are restricting our attention to (flat) Minkowski spacetime, QFTs can also be defined in curved spacetimes. In this respect, the action (8.31) written in terms of differential forms is also valid for non-flat metrics. An interesting difference between the two terms is that, while the first one depends on the spacetime metric the $\theta$-term does not and is therefore topological. Metric dependence is actually signaled by the presence of the Hodge dual in the action.

Another relevant fact that can be easily shown using differential forms is that the $\theta$-term is a total derivative, as we saw in Eq. (8.16). Indeed, Eq. (8.30) can be explicitly written in terms of the gauge field one-form as

$$\begin{aligned}\mathrm{tr}\,(F\wedge F) &= \mathrm{tr}\,(dA\wedge dA + 2dA\wedge A\wedge A + A\wedge A\wedge A\wedge A)\\ &= d\,\mathrm{tr}\left(A\wedge dA + \frac{2}{3}A\wedge A\wedge A\right),\end{aligned} \tag{8.32}$$

where we have used that $\mathrm{tr}\,(A\wedge A\wedge A\wedge A) = 0$, as a result of the anticommutativity of one-forms and the trace's cyclic property. Using the properties of the Hodge dual operator, we finally write

$$\star\mathrm{tr}\,(F\wedge F) = d^\dagger J, \tag{8.33}$$

where $d^\dagger \equiv \star d \star$ is the adjoint exterior derivative [58] and $J$ is the current one form

$$J = \star \mathrm{tr}\left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A\right). \tag{8.34}$$

Once expressed in components we retrieve Eq. (8.16).

The trace on the right-hand side of (8.34) defines the *Chern–Simons form*. Applying (8.23) and after some algebra we obtain its gauge transformation

$$\begin{aligned} \omega_3(A) \equiv \mathrm{tr}\left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A\right) \\ \longrightarrow \quad \omega_3(A) - \frac{1}{3}\mathrm{tr}\left[(g^{-1}dg) \wedge (g^{-1}dg) \wedge (g^{-1}dg)\right]. \end{aligned} \tag{8.35}$$

The Chern–Simons form is a very interesting object for many reasons. One is that it gives rise to the action

$$S_{\mathrm{CS}} = -\frac{k}{4\pi} \int_{\mathcal{M}_3} \mathrm{tr}\left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A\right), \tag{8.36}$$

where $\mathcal{M}_3$ is a three-dimensional spacetime and $k$ is a constant known as the Chern–Simons level. Although (8.35) implies that the action is not gauge invariant

$$S_{\mathrm{CS}} \longrightarrow S_{\mathrm{CS}} + \frac{k}{12\pi} \int_{\mathcal{M}_3} \mathrm{tr}\left[(g^{-1}dg) \wedge (g^{-1}dg) \wedge (g^{-1}dg)\right], \tag{8.37}$$

the extra term equals $2\pi n k$, with $n$ the winding number of the gauge transformation defined in Eq. (8.7). Since the quantum theory can be formulated using functional integrals involving $\exp(iS_{\mathrm{CS}})$, this gauge variance is not a problem provided the Chern–Simons level $k$ is an integer. The action (8.36) defines a topological field theory appearing in many contexts in physics, ranging from quantum gravity [112, 113] to condensed matter, where it has found important applications in the theory of the quantum Hall effect [84, 114].

To conclude this discussion, let us also mention that the four-form (8.30) is also related to the axial anomaly studied in Section 7. Defined on a Euclidean spacetime, the integrated anomaly of the axial-vector current can be shown to be [95, 110, 111]

$$\int_{\mathcal{M}_4} d^4x\, \partial_\mu \langle J^\mu_A(x) \rangle = -2i\big(N_+ - N_-\big), \tag{8.38}$$

and $N_\pm$ are the number of positive/negative chirality solutions to the equation $\not{D}(A)\psi = 0$, with $\not{D}(A) \equiv \gamma^\mu(\partial_\mu - iA_\mu)$ the Dirac operator on the Euclidean manifold $\mathcal{M}_4$. The difference $N_+ - N_-$ appearing on the right-hand side of Eq. (8.38) is in fact a topological invariant called the *index* of the Dirac operator. This quantity can be computed using the Atiyah–Singer index

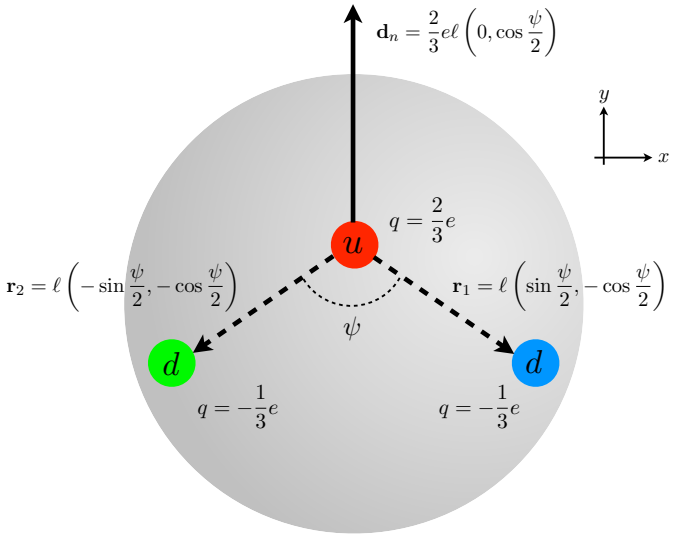



**Fig. 12:** Classical depiction of the neutron and its electric dipolar moment $\mathbf{d}_n$. The components of the $d$ quarks position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are written using the coordinate axes shown in the picture, with origin on the position of the $u$ quark.

theorem [57–60] and in four dimensions it is given by the integral of the four-form (8.30)

$$\text{ind}\,\not{D} = -\frac{1}{8\pi^2}\int_{\mathcal{M}_4} F\wedge F, \tag{8.39}$$

which, as explained above, is itself a topological quantity. By substituting this result into (8.38), we retrieve the known form of the anomaly, apart from a global factor of $i$ that is the consequence of working in Euclidean signature.

### 8.2 Breaking CP strongly

A significant feature of the $\theta$-term (8.12) is that it violates both parity and CP, the combination of parity and charge conjugation,

$$\text{CP}: \begin{cases} \mathbf{E}^a(t,\mathbf{r}) & \longrightarrow & \mathbf{E}^a(t,-\mathbf{r}) \\ \mathbf{B}^a(t,\mathbf{r}) & \longrightarrow & -\mathbf{B}^a(t,-\mathbf{r}) \end{cases} \qquad \Longrightarrow \qquad \text{CP}: S_\theta \longrightarrow -S_\theta. \tag{8.40}$$

To understand these transformations heuristically, we can use the analogy with Maxwell's electric and magnetic fields to conclude that $\mathbf{E}^a$ is reversed by both parity and charge conjugation, whereas the pseudovector $\mathbf{B}^a$ is preserved by the former and reversed by the latter. Notice that since CPT is a symmetry of QFT, a breaking of CP is equivalent to a violation of time reversal T.

Among the phenomena where CP (or T) violation can manifest in QCD is the existence of a nonvanishing electric dipole moment of the neutron (see, for example, Refs. [115, 116] for reviews). To be clear, were neutrons elementary, we would not expect them to have an electric dipolar moment. But being composed of three valence quarks with different charges, a nonvanishing value may appear depending on the quark distribution. To estimate its size, let us consider a classical picture of the neutron

assuming a structure similar to the water molecule (see Fig. 12): the two $d$ quarks are located at a distance $\ell$ of the $u$ quark and their position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ span an angle $\psi$ with each other. Taking coordinates on the plane defined by the three quarks, the modulus of the electric dipole moment $\mathbf{d}_n$ is readily computed to be

$$|\mathbf{d}_n| = \frac{2}{3} e\ell \cos\frac{\psi}{2} \equiv \frac{2}{3} e\ell \sin\frac{\theta}{2}, \tag{8.41}$$

where we have introduced the angle $\theta \equiv \pi - \psi$, controlling the amount of CP violation. To estimate the prefactor in Eq. (8.41), we recall that the distance $\ell$ between the quarks is of the order of the pion's Compton wavelength

$$\ell \simeq \frac{\hbar}{m_\pi c}, \tag{8.42}$$

where for computational purposes we have restored powers of $\hbar$ and $c$. Noticing that $\hbar c \simeq 200\,\text{MeV}\cdot\text{fm}$ and $m_\pi c^2 \simeq 135\,\text{MeV}$, we find

$$|\mathbf{d}_n| \simeq 10^{-13} \sin\frac{\theta}{2}\, e\cdot\text{cm}. \tag{8.43}$$

A comparison with experimental measurements of the neutron electric dipole [117, 118]

$$|\mathbf{d}_n|_{\text{exp}} \lesssim 10^{-26}\, e\cdot\text{cm}, \tag{8.44}$$

leads then to the bound

$$\theta \lesssim 10^{-13}. \tag{8.45}$$

This means that the angle $\psi = \pi - \theta$ in Fig. 12 is extremely close to $\pi$, making the quark configuration inside the neutron look like a $CO_2$ rather than a water molecule.

This cartoon calculation exhibits the basic feature of the so-called *strong CP problem*: the stringent experimental bound for the neutron electric dipole moment implies the existence of a dimensionless parameter that is extremely small without any dynamical reason. Once we rephrase the problem in the correct language of QCD, we will see that this parameter is precisely the $\theta$ coupling introduced in Eq. (8.12).

From a QFT point of view the neutron electric dipole emerges from the dimension-five nonminimal coupling of the neutron to the electromagnetic field

$$S \supset -\frac{i}{2}|\mathbf{d}_n| \int d^4x\, \overline{n}\sigma^{\mu\nu}\gamma_5 n F_{\mu\nu}, \tag{8.46}$$

where $n$ is the neutron field and $\sigma^{\mu\nu}$ has been defined in Eq. (4.49). This term is explicitly gauge invariant but breaks parity, as follows from the presence of $\gamma_5$. It is, however, invariant under charge conjugation, which preserves the neutron and gauge fields, and therefore it breaks CP. The operator (8.46) is in fact an effective interaction emerging from loop diagrams in the EFT of pions and nucleons described by an extension of the action (5.58). To construct this theory, let us consider QCD with the two light flavors $u$

and $d$. Written in terms of the chiral isospin doublets

$$\boldsymbol{q}_{R,L} = \begin{pmatrix} u_{R,L} \\ d_{R,L} \end{pmatrix}, \tag{8.47}$$

the microscopic action takes the form

$$S = \int d^4x \left( i\overline{\boldsymbol{q}}_R \not{D} \boldsymbol{q}_R + i\overline{\boldsymbol{q}}_L \not{D} \boldsymbol{q}_L + \overline{\boldsymbol{q}}_L M \boldsymbol{q}_R + \overline{\boldsymbol{q}}_R M^T \boldsymbol{q}_L - \frac{\theta}{32\pi^2} \epsilon^{\mu\nu\alpha\beta} F^a_{\mu\nu} F^a_{\alpha\beta} + \dots \right), \tag{8.48}$$

where $D_\mu = \partial_\mu - iA^a_\mu T^a$ denotes the gauge covariant derivative and the mass matrix is given by

$$M = \begin{pmatrix} m_u & 0 \\ 0 & m_d \end{pmatrix}. \tag{8.49}$$

We have included the $\theta$-term, while the ellipsis indicates other terms not important for the argument. In writing the action (5.58) we assumed that quarks are massless, and also the NG bosons associated with chiral SSB, but we now relax this condition. Although the chiral $\mathrm{SU}(2)_R \times \mathrm{SU}(2)_L$ transformations

$$\boldsymbol{q}_{R,L} \longrightarrow U_{R,L} \boldsymbol{q}_{R,L}, \tag{8.50}$$

do not leave the quark action (8.48) invariant, we can restore the symmetry promoting the mass matrix $M$ to a spurion field transforming as

$$M \longrightarrow U_L M U_R^\dagger. \tag{8.51}$$

Thus, the original action can be seen as one where chiral symmetry is spontaneously broken by $M$ taking the value in Eq. (8.49). The transformation of $M$, together with Eq. (5.56), provides the basic clue to incorporate masses into the NG action (5.58). An invariant mass term can be built by taking the trace of the product of the mass and the NG boson matrices

$$S_{\mathrm{NG}} = \int d^4x \left[ \frac{f_\pi^2}{4} \mathrm{tr} \left( D_\mu \boldsymbol{\Sigma}^\dagger D^\mu \boldsymbol{\Sigma} \right) + f_\pi^3 B_0 \mathrm{tr} \left( M^\dagger \boldsymbol{\Sigma} + \boldsymbol{\Sigma}^\dagger M \right) \right]. \tag{8.52}$$

Here $D_\mu \boldsymbol{\Sigma} = \partial_\mu \boldsymbol{\Sigma} - iA_\mu [Q, \boldsymbol{\Sigma}]$, with $Q = e\sigma^3$ the pion charge matrix, is the electromagnetic covariant derivative and $B_0$ is a numerical constant that cannot be determined within the EFT framework[24]. Substituting the explicit expressions of $M$ and $\boldsymbol{\Sigma}$, and expanding in powers of the pion fields, we find the mass term

$$\Delta S_{\mathrm{NG}} = -f_\pi B_0 (m_u + m_d) \int d^4x \Big[ (\pi^0)^2 + 2\pi^+ \pi^- \Big], \tag{8.53}$$

[24] The pion effective action $S_{\mathrm{NG}}$ also contains terms induced by the anomalous global symmetries of QCD, which are fully determined by the mathematical structure of the anomaly (see, for example, Ref. [93]). An example is the term proportional to $\left( \mathrm{tr}\, \log \boldsymbol{\Sigma} - \mathrm{tr}\, \log \boldsymbol{\Sigma}^\dagger \right) F_{\mu\nu} \widetilde{F}^{\mu\nu}$, accounting for the electromagnetic decay of the neutral pion discussed in page 82.

from where we read off the pion mass

$$m_\pi^2 = 2f_\pi B_0(m_u + m_d) \qquad \Longrightarrow \qquad B_0 = \frac{m_\pi^2}{2f_\pi(m_u+m_d)}. \tag{8.54}$$

Within this approximation, neutral and charged pions have the same mass.

Nucleons can also be added to the chiral Lagrangian (see Refs. [119, 120] for reviews). They are introduced through the isospin doublet

$$N = \begin{pmatrix} p \\ n \end{pmatrix}, \tag{8.55}$$

transforming under $\mathrm{SU}(2)_R \times \mathrm{SU}(2)_L$ as outlined in Refs. [121–123]

$$N \longrightarrow K(U_R, U_L, \boldsymbol{\Sigma})N. \tag{8.56}$$

The so-called compensating field $K(U_R, U_L, \boldsymbol{\Sigma})$ is a SU(2)-valued matrix depending on the NG boson matrix $\boldsymbol{\Sigma}(x)$, and through it on the spacetime point. It is defined by $K(U_R, U_L, \boldsymbol{\Sigma}) = \boldsymbol{u}'(x)^{-1} U_R \boldsymbol{u}(x)$, where $\boldsymbol{u}(x)^2 \equiv \boldsymbol{\Sigma}(x)$ and $\boldsymbol{u}'(x)^2 \equiv \boldsymbol{\Sigma}'(x) = U_R \boldsymbol{\Sigma}(x) U_L^\dagger$, thus providing a nonlinear realization of the $\mathrm{SU}(2)_R \times \mathrm{SU}(2)_L$ global chiral symmetry acting on the nucleon isospin doublet.

Having established the transformation of nucleons, we add to the effective action the term

$$\Delta S_{\pi N} = \int d^4x\, \overline{N}\Big[i\not{\mathcal{D}} - f(\boldsymbol{\Sigma})\Big]N, \tag{8.57}$$

with $f(\boldsymbol{\Sigma})$ a matrix-valued function depending on the NG boson matrix and such that $\not{\mathcal{D}} \equiv \not{\mathbb{D}} + if(\boldsymbol{\Sigma})$ defines a covariant derivative with respect to the local transformation (8.56), $\not{\mathcal{D}} \to K \not{\mathcal{D}} K^\dagger$. At linear order in the pion fields, it includes the pion–nucleon vertices

$$\begin{aligned} f(\boldsymbol{\Sigma}) &= m_N \mathbb{1} + \frac{g_A}{2f_\pi}\gamma^\mu\gamma_5 \partial_\mu \boldsymbol{\pi} + \mathcal{O}(\boldsymbol{\pi}^2) \\ &= m_N \mathbb{1} + \frac{g_A}{2\sqrt{2}f_\pi}\big(\overline{n}\gamma^\mu\gamma_5 n - \overline{p}\gamma^\mu\gamma_5 p\big)\partial_\mu\pi^0 + \frac{g_A}{2f_\pi}\big(\overline{n}\gamma^\mu\gamma_5 p\, \partial_\mu\pi^- + \overline{p}\gamma^\mu\gamma_5 n\, \partial_\mu\pi^+\big), \end{aligned} \tag{8.58}$$

where $m_N$ is the nucleon mass. Incidentally, substituting this expression of $f(\boldsymbol{\Sigma})$ into the action (8.57) we can integrate by parts and move the derivative from $\boldsymbol{\pi}$ to $N$ and $\overline{N}$. For scattering processes with on-shell nucleons the Dirac equation $i\not{\partial}N = m_N N$ can be implemented to write the nucleon–pion interaction term as $ig_{\pi NN}\overline{N}t_{\mathbf{f}}^I N\pi^I$, with $t_{\mathbf{f}}^I$ the generators in the fundamental representation of SU(2). Furthermore, the coupling constant $g_{\pi NN}$ satisfies by the Goldberger–Treiman relation [124]

$$f_\pi g_{\pi NN} = g_A m_N. \tag{8.59}$$

Notice that, since $g_A$ is real, the couplings in Eq. (8.58) preserve CP.

We would like to study the effects in the chiral Lagrangian of adding the $\theta$-term to the quark action. At this point we should invoke the analysis presented in Box 10 (see page 83) where we saw how, due to the chiral anomaly, implementing a chiral rotation of the fermions induces a $\theta$-term in the action. More

precisely, performing a chiral rotation of the $u$-quark

$$u_{R,L} \longrightarrow e^{\pm i\alpha} u_{R,L}, \tag{8.60}$$

results in shifting the value of the theta angle

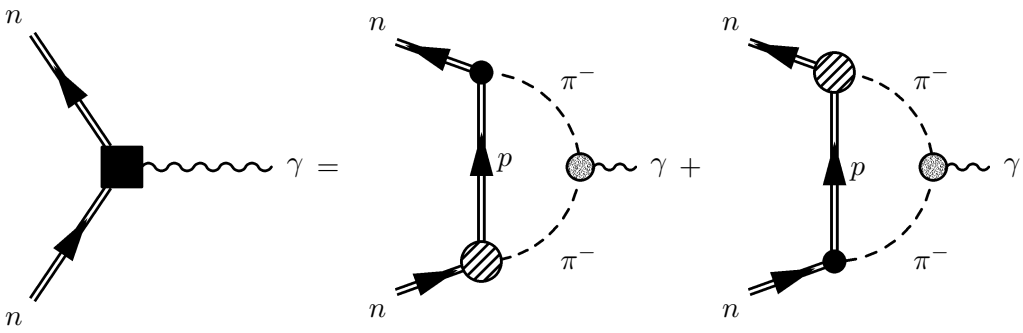
$$S = \int d^4x \left( i\overline{\boldsymbol{q}}_R \not{D} \boldsymbol{q}_R + i\overline{\boldsymbol{q}}_L \not{D} \boldsymbol{q}_L + \overline{\boldsymbol{q}}_L M \boldsymbol{q}_R + \overline{\boldsymbol{q}}_R M^\dagger \boldsymbol{q}_L - \frac{\theta - 2\alpha}{32\pi^2} \epsilon^{\mu\nu\alpha\beta} F^a_{\mu\nu} F^a_{\alpha\beta} + \dots \right), \tag{8.61}$$

and a complex mass matrix

$$M = \begin{pmatrix} e^{2i\alpha} m_u & 0 \\ 0 & m_d \end{pmatrix}. \tag{8.62}$$

In particular, setting $\alpha = \frac{1}{2}\theta$ the $\theta$-term cancels and all dependence on $\theta$ is shifted to a phase in the mass matrix $M$. In more physical terms, we have transferred the source of CP violation in the quark action from the $\theta$-term to a complex coupling[25].

It might seem that, at the level of the chiral effective field theory, the phase in the mass matrix $M = \text{diag}(e^{i\theta} m_u, m_d)$ could be removed by an appropriate chiral transformation of the NG field $\boldsymbol{\Sigma}(x)$. In doing so, however, we introduce a $\theta$-dependence in $f(\boldsymbol{\Sigma}, \theta)$ defined in (8.57), inducing additional nucleon–pion couplings. In particular, besides the neutron–proton–pion vertex in Eq. (8.58), there is a new CP violating vertex contributing to the dimension-five non-minimal electromagnetic coupling in Eq. (8.46)




(8.63)

The black dots in the diagrams on the right-hand side represent the CP-violating vertex, whereas the lined blobs indicate the neutron–pion coupling in (8.58). The chiral loop integrals are logarithmically divergent and once evaluated give the following contribution to the neutron electric dipole moment [125]

$$|\mathbf{d}_n| = \frac{1}{4\pi^2} \frac{|g_{\pi NN} \overline{g}_{\pi NN}|}{m_N} \log\left(\frac{m_N}{m_\pi}\right), \tag{8.64}$$

where

$$|\overline{g}_{\pi NN}| \approx 0.027 |\theta| \tag{8.65}$$

is the coupling of the CP-violating vertex and, in the spirit of EFT, integrals have been cut off at $\Lambda = m_\pi$.

[25] In fact, it is easy to prove that the quantity $\overline{\theta} \equiv \theta + \arg\det M$ remains invariant under chiral transformations of the quarks.

Substituting the value for the CP-preserving pion–nucleon coupling and implementing the experimental bound (8.44), we find

$$|\theta| \lesssim 10^{-11}. \tag{8.66}$$

We see that the amount of fine tuning in the $\theta$ parameter needed to explain experiments is not very far off the one obtained for the angle $\theta$ in (8.45) in the classical toy model of the neutron (not by accident both quantities were denoted by the same Greek letter).

**Box 12. A "potential" for $\theta$**

We would like to understand how the energy of the ground state of QCD depends on the parameter $\theta$. There are a number of things that can be said about this quantity, that we denote by $V(\theta)$. As we learned above [see Eq. (8.19)], the $\theta$-term is a topological object and any physical quantity depending on it like $V(\theta)$ should be periodic in $\theta$ with period equal to $2\pi$,

$$V(\theta + 2\pi) = V(\theta). \tag{8.67}$$

Moreover, there exists a very elegant argument showing that energy is minimized for $\theta = 0$ [126]

$$V(0) \leq V(\theta). \tag{8.68}$$

To go beyond these general considerations and find an explicit expression of $V(\theta)$ in QCD, we consider the potential energy in the pion effective action (8.52),

$$\mathcal{V}(\mathbf{\Sigma}) = -\frac{m_\pi^2 f_\pi^2}{2(m_u + m_d)} \mathrm{tr}\left(M^\dagger \mathbf{\Sigma} + M \mathbf{\Sigma}^\dagger\right), \tag{8.69}$$

where $M$ is given by

$$M = \begin{pmatrix} e^{i\theta} m_u & 0 \\ 0 & m_d \end{pmatrix}. \tag{8.70}$$

To find the vacuum energy, we look for a NG boson matrix configuration minimizing $\mathcal{V}(\mathbf{\Sigma})$.

In fact, since the mass matrix is diagonal it can be seen that the trace in (8.69) only depends on the diagonal components of $\mathbf{\Sigma}$. This means that, in order to minimize the potential, it is enough consider to NG matrices of the form $\mathbf{\Sigma} = \mathrm{diag}(e^{i\varphi_1}, e^{i\varphi_2})$. Furthermore, the dependence on $\theta$ in the mass matrix can be shifted to the NG boson matrix by the field redefinition

$$\mathbf{\Sigma} \longrightarrow \widetilde{\mathbf{\Sigma}} \equiv \begin{pmatrix} e^{-\frac{i\theta}{2}} & 0 \\ 0 & 1 \end{pmatrix} \mathbf{\Sigma} \begin{pmatrix} e^{-\frac{i\theta}{2}} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{i(\varphi_1 - \theta)} & 0 \\ 0 & e^{i\varphi_2} \end{pmatrix}. \tag{8.71}$$

Imposing the condition $\det \widetilde{\mathbf{\Sigma}} = 1$, we have $\varphi_1 + \varphi_2 = \theta \bmod 2\pi$.

Substituting the redefined NG matrix field $\widetilde{\Sigma}$ into (8.69) with $M = \mathrm{diag}(m_u, m_d)$, we arrive

at the potential

$$\mathcal{V}(\varphi_1,\varphi_2) = -\frac{m_\pi^2 f_\pi^2}{m_u+m_d}\big(m_u\cos\varphi_1 + m_d\cos\varphi_2\big), \tag{8.72}$$

that has to be minimized subject to the constraint $\varphi_1+\varphi_2=\theta$. The equation to be solved is

$$m_u\sin\varphi_1 = m_d\sin(\theta-\varphi_1), \tag{8.73}$$

that, after a bit of algebra, gives

$$\begin{aligned}\cos^2\varphi_1 &= \frac{(m_u+m_d\cos\theta)^2}{m_u^2+m_d^2+2m_um_d\cos\theta},\\ \cos^2\varphi_2 &= \frac{(m_d+m_u\cos\theta)^2}{m_u^2+m_d^2+2m_um_d\cos\theta}.\end{aligned} \tag{8.74}$$

Substituting these results into (8.72), we arrive at the expression of the QCD vacuum energy as a function of $\theta$

$$V(\theta) = -\frac{m_\pi^2 f_\pi^2}{m_u+m_d}\sqrt{m_u^2+m_d^2+2m_um_d\cos\theta}. \tag{8.75}$$

In Fig. 13 we have represented this function for various values of the ratio $m_d/m_u$, from where we see that, as announced, the minimum occurs at $\theta=0$. We also see that when $m_u=m_d$ there are cusps at the maxima located at $\theta=(2n+1)\pi$, that are smoothed out when the quarks have different masses. Being an experimental fact that $\theta$ is very small, we can expand $V(\theta)$ around $\theta=0$ to find

$$V(\theta) = -m_\pi^2 f_\pi^2 + \frac{1}{2}m_\pi^2 f_\pi^2\frac{m_um_d}{(m_u+m_d)^2}\theta^2. \tag{8.76}$$

This expression will become handy later on when it will be reinterpreted as the potential for the axion field.

Since $m_s \gg m_u, m_d$ we have restricted our attention to QCD with the two lightest flavors, although the analysis can be easily extended to any $N_f \geq 2$. The resulting expression of the ground state energy $V(\theta; m_1,\ldots,m_f)$ for small $\theta$ is symmetric under permutations of the quark masses and satisfies a recursion relation

$$V(\theta; m_1,\ldots,m_{f-1}) = \lim_{m_f\to\infty} V(\theta; m_1,\ldots,m_f), \tag{8.77}$$

implementing the decoupling of the $f$-th flavor.

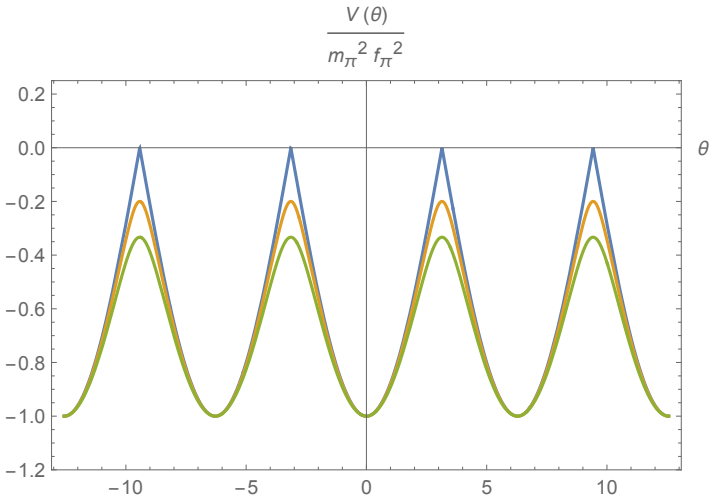



**Fig. 13:** Plot of $V(\theta)$ in Eq. (8.75) for three different values of the $\frac{m_u}{m_d}$ ratio: 1 (blue), 0.3 (orange), and 0.5 (green).

### 8.3 Enters the axion

We would like to understand the smallness of $\theta$ in a natural way, i.e., either as following from some symmetry principle or by finding out some dynamical reason for its value[26]. One possible explanation would be that $m_u = 0$, so a chiral rotation of the $u$-quark field would get rid of the $\theta$-term without introducing CP-violating phases in the chiral Lagrangian. This is however no good, since all experimental evidences indicate that the $u$-quark is not massless.

A very popular solution to the CP problem is the one proposed by Roberto Peccei and Helen Quinn [127, 128] consisting in making the $\theta$-parameter the vev of a pseudoscalar field $a(x)$, the *axion* [129, 130], whose potential would drive it to $\langle 0|a(x)|0\rangle = 0$. To be more precise, let us consider the action

$$S = \int d^4x \left( i\overline{\boldsymbol{q}}_R \not{D} \boldsymbol{q}_R + i\overline{\boldsymbol{q}}_L \not{D} \boldsymbol{q}_L + \overline{\boldsymbol{q}}_L M \boldsymbol{q}_R + \overline{\boldsymbol{q}}_R M^\dagger \boldsymbol{q}_L - \frac{1}{32\pi^2 f_a} a F^a_{\mu\nu} \widetilde{F}^{a\mu\nu} \right), \tag{8.78}$$

where $f_a$ is an energy scale introduced so the axion field has the canonical dimenension of energy. We can now play the old game of shifting the last term in the action (8.78) to a complex phase in the mass matrix. In the low-energy effective field theory, this phase can be absorbed into the NG bosons matrix by the field redefinition (cf. the analysis presented in Box 12)

$$\boldsymbol{\Sigma} \longrightarrow \begin{pmatrix} e^{-\frac{ia}{2f_a}} & 0 \\ 0 & 1 \end{pmatrix} \boldsymbol{\Sigma} \begin{pmatrix} e^{-\frac{ia}{2f_a}} & 0 \\ 0 & 1 \end{pmatrix}. \tag{8.79}$$

In the absence of a mass term for the NG bosons, $\boldsymbol{\Sigma}$ only has derivative couplings and the theory is invariant under constant shifts of the axion field, $a(x) \to a(x) + \text{constant}$. The presence of the term $f_\pi^3 B_0 \mathrm{tr}\left(M^\dagger \boldsymbol{\Sigma} + \boldsymbol{\Sigma}^\dagger M\right)$, however, induces a potential that can be read off Eq. (8.75) with $\theta$ re-

[26] The fact that in the $CO_2$ molecule the angle $\theta$ is zero is a consequence of the dynamics of the atomic orbitals and is therefore "natural".

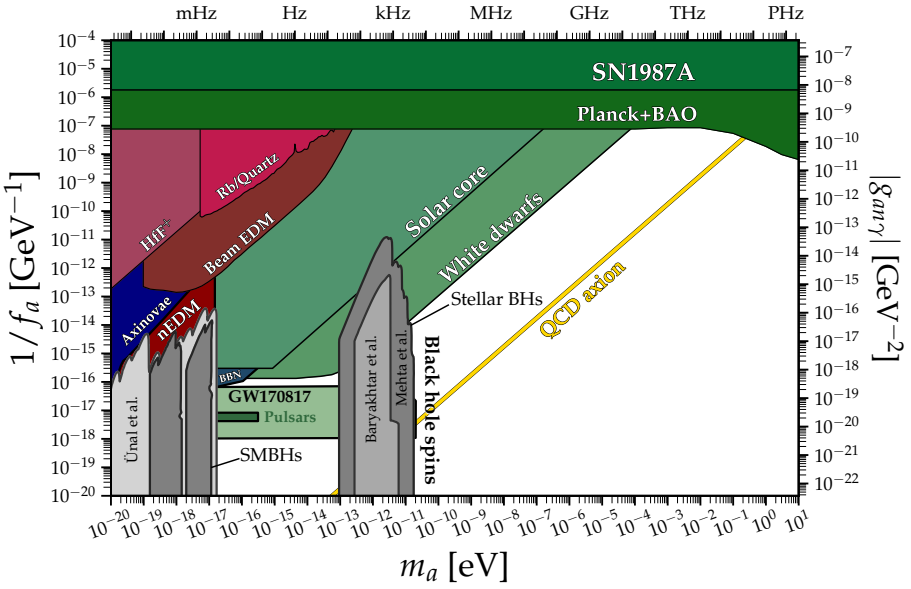



**Fig. 14:** Exclusion plot from Ref. [134] for the axion parameters $f_a$ (resp. $g_{an\gamma}$) and $m_a$. The yellow line represents the relation given in Eq. (8.81).

placed by $a/f_a$. Expanding around the minimum at $a = 0$, we find

$$V(a) = \frac{m_\pi^2 f_\pi^2}{2f_a^2} \frac{m_u m_d}{(m_u + m_d)^2} a^2 + \ldots, \tag{8.80}$$

where we have dropped constant terms and the ellipsis indicates higher-order axion self-interactions. This gives the axion mass

$$m_a = \frac{m_\pi f_\pi}{f_a} \frac{\sqrt{m_u m_d}}{m_u + m_d} = 5.7 \left( \frac{10^9 \text{ GeV}}{f_a} \right) \text{ meV}. \tag{8.81}$$

The field redefinition (8.79) also induces axion interactions with mesons, baryons, leptons, and photons. For example,

$$S_{\text{axion}} \supset - \int d^4x \left( \frac{i}{2} g_{ap\gamma} a \bar{p} \sigma^{\mu\nu} \gamma_5 p F_{\mu\nu} + \frac{i}{2} g_{an\gamma} a \bar{n} \sigma^{\mu\nu} \gamma_5 n F_{\mu\nu} + \frac{g_{a\gamma\gamma}}{4} a F_{\mu\nu} \widetilde{F}^{\mu\nu} \right), \tag{8.82}$$

where $g_{an\gamma} = -g_{ap\gamma} \sim f_a^{-2}$ and $g_{a\gamma\gamma} \sim f_a^{-1}$. The last non-minimal electromagnetic coupling of the axion comes from the anomaly-induced term in the chiral Lagrangian pointed out in the footnote on page 93. In a strong magnetic field, this term allows the conversion of a photon into an axion and vice versa, one of the main astrophysical signatures of the axion and also the target process of the light-shining-through-walls experiments [131].

Among other candidates for dark matter (sterile neutrinos, supersymmetric particles, etc.) axions are currently one of the most popular candidates to account for the missing matter in the universe [132, 133]. Cosmological and astrophysical phenomena provide a wide class of observational windows for these kind of particles, ranging from CMB physics to stellar astrophysics and black holes (see Fig. 14). Observations so far have been used to constrain the parameter space for axion-like particles (ALPs),

leaving a wide allowed region including most of the values of the QCD axion. A comprehensive overview of current axion experiments and the bounds on different parameters can be found in the review [116], as well as in Ref. [117] (see also Ref. [134] for a collection of exclusion plots for various parameters).

## 9 The electroweak theory

It is time we look into the electroweak sector of the SM. As already mentioned several times in these lectures, our current understanding of the electromagnetic and weak forces is based on a gauge theory with group $\mathrm{SU}(2)\times\mathrm{U}(1)_Y$. This theory has subtle differences with respect to the color SU(3) QCD gauge group used to describe strong interactions. The basic one is that it is a chiral theory in which left- and right-handed fermions transform in different representations of the gauge group. Closely related to this is that the $\mathrm{SU}(2)\times\mathrm{U}(1)_Y$ gauge invariance is spontaneously broken at low energies by an implementation of the BEH mechanism explained in Section 5. This feature, that for decades was the shakiest part of the electroweak theory, was finally confirmed in July 2012 when the detection of the Higgs boson was announced at CERN, thus fitting the final piece into the jigsaw puzzle.

Whereas only hadrons (i.e., quarks) partake of the strong interaction, the weak force affects both quarks and leptons. Its chiral character is reflected in that the weak interaction violate parity, a fact discovered in the late 1950s in the study of $\beta$-decay and other processes mediated by the weak force [135–138]. Unlike gluons, which couple to quarks through a vector current $J^\mu_{\mathrm{QCD}}=\overline{q}\gamma^\mu q$, the carriers of the weak force interact with matter via the V − A current $J^\mu_{\mathrm{weak}}=\overline{\psi}\gamma^\mu(\mathbb{1}-\gamma_5)\psi$, with $\psi$ either a lepton or a quark field [139, 140].

### 9.1 Implementing $\mathrm{SU}(2)\times\mathrm{U}(1)_Y$

To be more precise, $\beta$-decay transmutes left-handed electrons into left-handed electron neutrinos (and vice versa), while $u$-quarks (resp. $d$-quarks) transform into $d$ quarks (resp. $u$-quarks). This suggests grouping left-handed electrons/neutrinos and quarks into doublets

$$\mathbf{L}=\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_L, \qquad \mathbf{Q}=\begin{pmatrix}u\\ d\end{pmatrix}_L, \tag{9.1}$$

and assume they transform in the fundamental representation **2** of the SU(2) algebra. At the same time, since right-handed electrons and quarks do not undergo $\beta$-decay, their components are taken to be SU(2) singlets

$$\ell_R\equiv e^-_R, \qquad U_R\equiv u_R, \qquad D_R\equiv d_R. \tag{9.2}$$

Moreover, since there is no experimental evidence of the existence of right-handed neutrinos, we do not include them in the description (at least for now; we will return to this issue later).

The whole picture is complicated because the weak force mixes with the electromagnetic interaction. In fact, the $\mathrm{U}(1)_Y$ of the electroweak gauge group is not the U(1) of Maxwell's theory. The

**Leptons**

| | $i=1$ | $i=2$ | $i=3$ | $t^3_{\mathbf{R}}$ | $Y_{\mathbf{R}}$ |
|---|---|---|---|---|---|
| $\mathbf{L}^i$ | $\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_L$ | $\begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_L$ | $\begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_L$ | $\frac{1}{2}\sigma^3$ | $-\frac{1}{2}\mathbb{1}$ |
| $\ell^i_R$ | $e^-_R$ | $\mu^-_R$ | $\tau^-_R$ | $0$ | $-1$ |

**Quarks**

| | $i=1$ | $i=2$ | $i=3$ | $t^3_{\mathbf{R}}$ | $Y_{\mathbf{R}}$ |
|---|---|---|---|---|---|
| $\mathbf{Q}^i$ | $\begin{pmatrix}u\\ d\end{pmatrix}_L$ | $\begin{pmatrix}c\\ s\end{pmatrix}_L$ | $\begin{pmatrix}t\\ b\end{pmatrix}_L$ | $\frac{1}{2}\sigma^3$ | $\frac{1}{6}\mathbb{1}$ |
| $U^i_R$ | $u_R$ | $c_R$ | $t_R$ | $0$ | $\frac{2}{3}$ |
| $D^i_R$ | $d_R$ | $s_R$ | $b_R$ | $0$ | $-\frac{1}{3}$ |

**Table 2:** Transformation properties of leptons and quarks in the electroweak sector of the SM. In addition to the indicated representations of SU(2) × U(1)$_Y$, quarks transform in the fundamental **3** irrep of SU(3), whereas leptons are singled under this group.

generator $Y_{\mathbf{R}}$ of the former, called the *weak hypercharge*, satisfies the Gell-Mann–Nishijima relation

$$Q = Y_{\mathbf{R}} + t^3_{\mathbf{R}}, \tag{9.3}$$

where $Q$ is the charge of the field in units of $e$ and $t^3_{\mathbf{R}}$ is the Cartan generator of SU(2) in the representation $\mathbf{R}$. As an example, for $\mathbf{L}$ in Eq. (9.1) we have $t^3_{\mathbf{2}} \equiv \frac{1}{2}\sigma^2 = \text{diag}(\frac{1}{2}, -\frac{1}{2})$ and $Q = \text{diag}(0, -1)$, so we have $Y(\mathbf{L}) = -\frac{1}{2}\mathbb{1}$. Repeating this for all lepton and quark fields, we find

$$Y(\mathbf{L}) = -\frac{1}{2}\mathbb{1}, \quad Y(\ell) = -1, \quad Y(\mathbf{Q}) = -\frac{1}{6}\mathbb{1}, \quad Y(U_R) = \frac{2}{3}, \quad Y(D_R) = -\frac{1}{3}, \tag{9.4}$$

where for the SU(2) singlets we have $t^3_{\mathbf{1}} = 0$. Notice that for U(1)$_Y$ we have $Y_{\mathbf{R}} = Y\,\mathbb{1}$, so the representation of U(1)$_Y$ is fully determined by the *hypercharge* $Y$.

We might be tempted to believe that with this we have determined how *all* matter fields in the SM transform under the gauge group SU(2) × U(1)$_Y$. However, for reasons that we so far ignore, nature has decided to have three copies of the structure just described. In addition to the electron, its neutrino, and the $u$- and $d$-quarks there are two more replicas or *families*. The second family includes the muon ($\mu^-$)

and its neutrino ($\nu_\mu$), together with the charm ($c$) and strange ($s$) quarks. The third family, on the other hand, contains the $\tau^-$ lepton, its neutrino ($\nu_\tau$), and the top ($t$) and bottom ($b$) quarks. Apart from an increasing hierarchy of masses, each extra family exactly replicates the transformation properties of the fields in the first one. To include this feature in our description, we add an index $i = 1, 2, 3$ to the doublet $\{\mathbf{L}^i, \mathbf{Q}^i\}$ and singlet $\{\ell_R^i, U_R^i, D_R^i\}$ fields introduced above, summarizing in Table 2 the three-family structure with the corresponding representations of SU(2) $\times$ U(1)$_Y$. We should not forget that, besides the electroweak quantum numbers, leptons are singlets with respect to color SU(3), whereas quarks are triplets transforming in the fundamental representation of this group.

Once the matter content of the SM is determined, as well as how the fields transform under the electroweak gauge group, we fix our attention on the gauge bosons. In the case of SU(2), it is convenient to use the $\{t^\pm_{\mathbf{R}}, t^3_{\mathbf{R}}\}$ basis, so the corresponding gauge field is written as[27]

$$\mathbf{W}_\mu = W^+_\mu t^-_{\mathbf{R}} + W^-_\mu t^+_{\mathbf{R}} + W^3_\mu t^3_{\mathbf{R}}, \tag{9.5}$$

whereas for the Abelian gauge field associated with U(1)$_Y$, we have

$$\mathbf{B}_\mu = B_\mu Y \mathbb{1}. \tag{9.6}$$

The covariant derivative needed to construct the matter action is then given by

$$\begin{aligned} D_\mu &= \partial_\mu - ig\mathbf{W}_\mu - ig'\mathbf{B}_\mu \\ &= \partial_\mu - igW^+_\mu t^-_{\mathbf{R}} - igW^-_\mu t^+_{\mathbf{R}} - igW^3_\mu t^3_{\mathbf{R}} - ig'B_\mu Y\mathbb{1}, \end{aligned} \tag{9.7}$$

where $g$ and $g'$ are the coupling constants associated with the two factors of the electroweak gauge group.

We should not forget, however, that the electric charge $Q$, the hypercharge $Y\mathbb{1}$, and the SU(2) Cartan generator $t^3_{\mathbf{R}}$ are not independent, but connected by the Gell-Mann–Nishijima relation (9.3). It is therefore useful to consider the combinations

$$\begin{aligned} A_\mu &= B_\mu \cos\theta_w + W^3_\mu \sin\theta_w, \\ Z_\mu &= -B_\mu \sin\theta_w + W^3_\mu \cos\theta_w, \end{aligned} \tag{9.8}$$

where $A_\mu$ is to be identified with the electromagnetic field, whose gauge group will be denoted by U(1)$_{\text{em}}$ to distinguish it from the one associated with the gauge field $\mathbf{B}_\mu$. The parameter $\theta_w$ is called the *weak mixing angle* and sometimes also the Weinberg angle, although it was first introduced by Glashow in Ref. [37]. Expressing the covariant derivative (9.7) in terms of the $\{W^\pm_\mu, A_\mu, Z_\mu\}$ gauge fields, we find

$$\begin{aligned} D_\mu = \partial_\mu &- igW^+_\mu t^-_{\mathbf{R}} - igW^-_\mu t^+_{\mathbf{R}} - iA_\mu\big(g\sin\theta_w t^3_{\mathbf{R}} + g'\cos\theta_w Y\mathbb{1}\big) \\ &- iZ_\mu\big(g\sin\theta_w t^3_{\mathbf{R}} - g'\cos\theta_w Y\mathbb{1}\big). \end{aligned} \tag{9.9}$$

[27] In terms of the generators $t^\pm_{\mathbf{R}} \equiv t^1_{\mathbf{R}} \pm it^2_{\mathbf{R}}$, the SU(2) algebra reads $[t^3_{\mathbf{R}}, t^\pm_{\mathbf{R}}] = \pm t^\pm_{\mathbf{R}}$, $[t^+_{\mathbf{R}}, t^-_{\mathbf{R}}] = 2t^3_{\mathbf{R}}$. This is just the algebra of ladder operators familiar from the theory of angular momentum in quantum mechanics.

Now, if $A_\mu$ is to be identified with the electromagnetic field, it has to couple to the electric charge matrix $eQ$. Consistency with the Gell-Mann–Nishijima relation (9.3) implies then

$$g \sin\theta_w = g' \cos\theta_w = e \qquad \Longrightarrow \qquad \tan\theta_w = \frac{g}{g'}. \tag{9.10}$$

This relation shows that the weak mixing angle not only measures the mixing among the Abelian gauge fields associated with the $\mathrm{U}(1)_Y$ and the Cartan generator of SU(2), but also of the relative strength of the interactions associated with the two factors of the electroweak gauge group. Implementing all the previous relations, the covariant derivative reads

$$D_\mu = \partial_\mu - \frac{ie}{\sin\theta_w} W_\mu^+ t_{\mathbf{R}}^- - \frac{ie}{\sin\theta_w} W_\mu^- t_{\mathbf{R}}^+ - ieA_\mu Q - \frac{2ie}{\sin(2\theta_w)} Z_\mu \big(t_{\mathbf{R}}^3 - Q\sin^2\theta_w\big), \tag{9.11}$$

where we have eliminated $Y$, $g$, and $g'$ in favor of $Q$, $e$, and $\theta_w$. With this, the SM matter action reads

$$S_{\text{matter}} = \sum_{k=1}^{3} \int d^4x \left( i\overline{\mathbf{L}}^k \not{D} \mathbf{L}^k + i\overline{\ell}_R^k \not{D} \ell_R^k + i\overline{\mathbf{Q}}^k \not{D} \mathbf{Q}^k + i\overline{U}_R^k \not{D} U_R^k + i\overline{D}_R^k \not{D} D_R^k \right). \tag{9.12}$$

Next we look at the gauge action

$$S_{\text{gauge}} = -\frac{1}{2} \int d^4x \left[ \mathrm{tr}\left(\mathbf{W}_{\mu\nu}\mathbf{W}^{\mu\nu}\right) + \mathrm{tr}\left(\mathbf{B}_{\mu\nu}\mathbf{B}^{\mu\nu}\right) \right], \tag{9.13}$$

where $\mathbf{W}_{\mu\nu}$ and $\mathbf{B}_{\mu\nu}$ are the field strengths of $\mathbf{W}_\mu$ and $\mathbf{B}_\mu$ respectively. Recasting it in terms of the electromagnetic and $Z_\mu$ gauge fields defined in Eq. (9.8), we have

$$\begin{aligned} S_{\text{gauge}} = -\int d^4x \Big\{ &\frac{1}{4} W_{\mu\nu}^+ W^{-\mu\nu} + \frac{1}{4} Z_{\mu\nu} Z^{\mu\nu} + \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{ie}{2} \cot\theta_w W_\mu^+ W_\nu^- Z^{\mu\nu} \\ &- \frac{ie}{2} W_\mu^+ W_\nu^- F^{\mu\nu} + \frac{e^2}{2\sin\theta_w} \Big[ (W_\mu^+ W^{+\mu})(W_\mu^- W^{-\mu}) - (W_\mu^+ W^{-\mu})^2 \Big] \Big\}, \end{aligned} \tag{9.14}$$

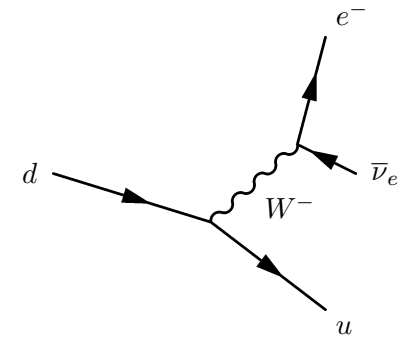
where $Z_{\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu$, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and we have defined

$$W_{\mu\nu}^\pm = \partial_\mu W_\nu^\pm - \partial_\nu W_\mu^\pm \mp e\big(W_\mu^\pm A_\nu - W_\nu^\pm A_\mu\big) \mp ie\cot\theta_w \big(W^\pm Z_\nu - W_\nu^\pm Z_\mu\big). \tag{9.15}$$

The SM gauge couplings can be now read off eqs. (9.11), (9.12), (9.14), and (9.15). The first thing to notice from the last two equations is that the $W_\mu^\pm$ gauge fields have electric charge $\pm e$ and also couple to the $Z_\mu$ gauge field, which has itself zero electric charge. A look at the matter action also shows that the two components of the SU(2) doublets are transmuted into one another by the emission/absorption of a $W$ boson. As to the $Z^0$, it can be emitted/absorbed by quarks and leptons with couplings that depend on their $\mathrm{SU}(2) \times \mathrm{U}(1)_Y$ quantum numbers (see Chapter 5 of Ref. [14] or any other SM textbook for the details). As a practical example, the neutron $\beta$-decay $n \to p^+ e^- \overline{\nu}_e$ proceeds by the emission of a $W^-$ by one of the neutron's $d$ quarks, turning itself into a $u$ quark (and the neutron into a proton). The $W^-$

then decays into an electron and an electronic antineutrino.

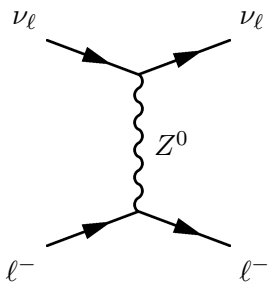
$$n[udd] \longrightarrow p^+[uud] + e^- + \overline{\nu}_e \qquad \Longrightarrow \qquad$$




(9.16)

As a second example, we also have lepton–neutrino scattering mediated by the interchange of a $Z^0$

$$\ell^- + \nu_\ell \longrightarrow \ell^- + \nu_\ell \qquad \Longrightarrow$$




(9.17)

where $\ell$ stands for $e$, $\mu$ or $\tau$. The existence of weak processes without transfer of electric charge is a distinctive prediction of the Glashow–Weinberg–Salam model. The discovery of these so-called neutral weak currents in the Gargamelle bubble chamber at CERN in 1973 [141] was solid experimental evidence in favor of the electroweak theory (see also Ref. [142] for a historical account). Let us also mention that $S_{\text{matter}} + S_{\text{gauge}}$ includes QED, and therefore describes all electromagnetic-mediated processes among leptons and quarks.

**Box 13. Hypercharges and anomaly cancellation**

Our discussion in Section 7 has very much stressed the need to eliminate anomalies affecting gauge invariance. Gauge anomalies come from the same triangle diagrams we encountered in our discussion of the chiral anomaly, namely those shown in Eq. (7.10). The only difference is that, instead of having an axial-vector current on the left and two vector currents on the right, now we have three gauge currents, one at each vertex.

Fortunately, to decide whether the SM is anomaly free we do not need to compute the diagrams themselves. It is enough to look at the group theory factor and check that the result is zero once we sum over all chiral fermions running in the loop. To compute this factor we consider the gauge generator at each vertex $(T^a_{\mathbf{R}})_{ij}$, where the indices $i, j$ are associated with the gauge index of the incoming/outgoing fermion entering/leaving the vertex, while $a$ is the index of the gauge field attached to it. Thus, for a given fermion species in the loop, the group theory factor multiplying the sum of the two triangles in (7.10) is given by

$$(T^a_{\mathbf{R}})_{ij}(T^b_{\mathbf{R}})_{jk}(T^c_{\mathbf{R}})_{ki} + (T^a_{\mathbf{R}})_{ij}(T^c_{\mathbf{R}})_{jk}(T^b_{\mathbf{R}})_{ki} = \mathrm{tr}\left(T^a_{\mathbf{R}}\{T^b_{\mathbf{R}}, T^c_{\mathbf{R}}\}\right). \tag{9.18}$$

Notice how the second term on the left-hand side is obtained from the first one by interchanging the two right vertices, as it happens in the second triangle diagram. Next, we have to sum over all fermion species, taking into account that left- and right-handed fermions contribute with opposite signs. Thus, the condition for anomaly cancellation is

$$\sum_L \mathrm{tr}\left(T^a_{\mathbf{R}}\{T^b_{\mathbf{R}}, T^c_{\mathbf{R}}\}\right)_L - \sum_R \mathrm{tr}\left(T^a_{\mathbf{R}}\{T^b_{\mathbf{R}}, T^c_{\mathbf{R}}\}\right)_R = 0, \tag{9.19}$$

where the sums are respectively over all left- and right-handed fermions in their corresponding representations. In checking anomaly cancellation it is important to keep in mind that if the gauge group has several semisimple factors, like the case of the SM, the generator $T^a_{\mathbf{R}}$ is the tensor product of the generators of each factor.

There is a simple way to summarize the group-theoretical information contained in Table 2 by just indicating the representations of the different fermion species with respect to SU(3) $\times$ SU(2) $\times$ U(1)$_Y$, including also now the gauge group factor associated with the strong force. Using the notation $(\mathbf{N}_c, \mathbf{N})_Y$, with $\mathbf{N}_c$, $\mathbf{N}$, and $Y$ the representations of SU(3), SU(2), and U(1)$_Y$, we write for a single family

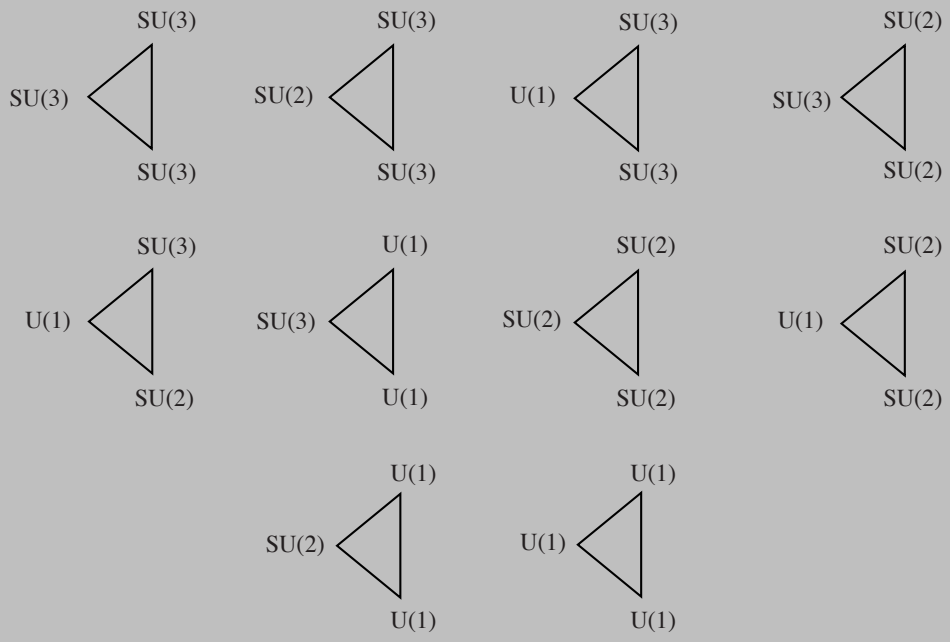
$$\begin{aligned} &\mathbf{L}^i : (\mathbf{1}, \mathbf{2})^L_{-\frac{1}{2}}, \qquad && \ell^i_R : (\mathbf{1}, \mathbf{1})^R_{-1}, \\ &\mathbf{Q}^i : (\mathbf{3}, \mathbf{2})^L_{\frac{1}{6}}, \qquad && U^i_R : (\mathbf{3}, \mathbf{1})^R_{\frac{2}{3}}, \qquad D^i_R : (\mathbf{3}, \mathbf{1})^R_{-\frac{1}{3}}, \end{aligned} \tag{9.20}$$

and we also introduced a superscript to remind ourselves whether they are left- or right-handed fermions (a useful information to decide what sign they come with in the anomaly cancellation condition). In this notation, the generators of the representation $(\mathbf{N}_c, \mathbf{N})_Y$ are given by

$$T^{(I,a)}_{(\mathbf{N}_c,\mathbf{N})_Y} = t^I_{\mathbf{N}_c} \otimes \mathbf{1} \otimes \mathbf{1} + \mathbf{1} \otimes t^a_{\mathbf{N}} \otimes 1 + \mathbf{1} \otimes \mathbf{1} \otimes Y, \tag{9.21}$$

where $I = 1, \ldots, 8$ and $a = 1, 2, 3$ respectively label the generators of SU(3) and SU(2). At a practical level, in order to check anomaly cancellation in the SM we attach a group factor to each vertex of the triangle and compute the left-hand side of (9.19) to check whether it vanishes. Since

we have three different factors and three vertices, there are ten inequivalent possibilities




Some of the possibilities are rather trivial. For example, the triangle with three SU(3) factors gives zero since the strong interaction does not distinguish left- from right-handed quarks and the two terms on the left-hand side of (9.19) are equal. The same happens whenever we have a single SU(3) or SU(2) factor, since the generators of these groups are traceless. At the end of the day, there are just four nontrivial cases. Using an obvious notation, they are: $SU(2)^3$, $SU(2)^2U(1)$, $SU(3)^2U(1)$, and $U(1)^3$. In the first case, since only left-handed fermions couple to SU(2), anomaly cancellation follows directly from the properties of the Pauli matrices

$$\mathrm{tr}\left(\sigma^i\{\sigma^j,\sigma^k\}\right) = 2\delta_{jk}\mathrm{tr}\,\sigma_i = 0. \tag{9.22}$$

For $SU(2)^2U(1)$, again the SU(2) factors only allow left-handed fermions in the loop, and the anomaly cancellation condition reads

$$\sum_L Y_L = 0, \tag{9.23}$$

while in the $SU(3)^2U(1)$ triangle the color factor rules out leptons, so we have

$$\sum_{\text{quarks},L} Y_L - \sum_{\text{quarks},R} Y_R = 0. \tag{9.24}$$

Finally, we are left with the triangle with one U(1) at each vertex, leading to the condition

$$\sum_L Y_L^3 - \sum_R Y_R^3 = 0, \tag{9.25}$$

where the sum in this case extends to all fermion species.

But this is not all. Since the SM model couples to gravity, it turns out that we might have gauge anomalies triggered by triangle diagrams with one gauge boson and two gravitons. The condition to avoid this is

$$\sum_L \mathrm{tr}\,(T^a_{\mathbf{R}})_L - \sum_R \mathrm{tr}\,(T^a_{\mathbf{R}})_R = 0. \tag{9.26}$$

In this case there are just three possibilities, corresponding to having a SU(3), SU(2) or U(1) factor in the non-graviton vertex. For the first two cases, the condition for anomaly cancellation is automatically satisfied, again because the generators of SU(3) and SU(2) are traceless. The third possibility, on the other hand, gives a nontrivial condition

$$\sum_L Y_L - \sum_R Y_R = 0, \tag{9.27}$$

where the sum runs over both leptons and quarks.

We have found the four conditions (9.23), (9.24), (9.25), and (9.27) to ensure the cancelation of anomalies, all of them involving the hypercharges of the chiral fermion fields in the SM. Now, instead of checking whether the hypercharges in Eq. (9.20) satisfy this condition, we are going to see to what extent anomaly cancellation determines the fermion hypercharges. Let us therefore write the representations of leptons and quarks in each family as $(\mathbf{1},\mathbf{2})^L_{Y_1}$, $(\mathbf{1},\mathbf{1})^R_{Y_2}$, $(\mathbf{3},\mathbf{2})^L_{Y_3}$, $U^i_R : (\mathbf{3},\mathbf{1})^R_{Y_4}$, and $D^i_R : (\mathbf{3},\mathbf{1})^R_{Y_5}$, reading now the anomaly cancellation conditions as equations to determine $Y_1,\ldots,Y_5$. These are

$$\begin{aligned} 2Y_1 + 6Y_3 &= 0, \\ 6Y_3 - 3Y_4 - 3Y_5 &= 0, \\ 2Y_1^3 + 6Y_3^3 - Y_2^3 - 3Y_4^3 - 3Y_5^3 &= 0, \\ 2Y_1 + 6Y_3 - Y_2 - 3Y_4 - 3Y_5 &= 0. \end{aligned} \tag{9.28}$$

Now, since these are homogeneous equations there exists the freedom to fix the overall normalization of the five hypercharges or, equivalently, to choose the value of one of them. Taking for example $Y_2 = -1$, we are left with four equations for the four remaining unknowns. They have a single solution given by

$$Y_1 = -\frac{1}{2}, \qquad Y_2 = -1, \qquad Y_3 = \frac{1}{6}, \qquad Y_4 = -\frac{1}{3}, \qquad Y_5 = \frac{2}{3}, \tag{9.29}$$

up to the interchange of $Y_4$ and $Y_5$ (notice that the associated fields $U^i_R$ and $D^i_R$ transform in the same representation with respect to the other two gauge group factors). This solution precisely reproduces the hypercharges shown in Eq. (9.20).

With this calculation we have learned two things. One is that all gauge anomalies (and also

the so-called mixed gauge-gravitational anomalies) cancel in the SM, and that they do so within each family. And second, that the anomaly cancellation condition is a very powerful way of constraining viable models in particle physics: in the SM it fixes, up to a global normalization, the $\mathrm{U}(1)_Y$ charges of all chiral fermions in the theory.

### 9.2 But, where are the masses?

Adding together eqs. (9.12) and (9.14), we still do not get the full action of the electroweak sector of the SM model. The reason is that all fermion species in the SM have nonvanishing masses and, therefore, we need to add the corresponding mass terms to the matter action. This is, however, a very risky business in a chiral theory like the electroweak model. As we learned in Box 7 (see page 48), fermion mass terms mix left- and right-handed components. In our case, since they transform in different representations of the $\mathrm{SU}(2) \times \mathrm{U}(1)_Y$ gauge group, adding such terms spoils gauge invariance and with that all hell breaks loose.

Fermion masses are not the only problem. Weak interactions are short ranged, something that can only be explained if the intermediate bosons $W^{\pm}$ and $Z^0$ have masses of the order of tens of GeV. Mass terms of the form $m_W^2 W_\mu^{\mp} W^{\pm\mu}$ and $m_Z^2 Z_\mu Z^\mu$ also violate gauge invariance, so it seems that we are facing double trouble.

The theory resulting from adding all needed mass terms to $S_{\mathrm{matter}} + S_{\mathrm{gauge}}$ is the original model proposed in 1961 by Glashow [37], where gauge invariance in *explicitly broken*. The inclusion of masses in the SM in a manner compatible with gauge invariance was achieved by Weinberg and Salam [38, 39] and requires the implementation of the BEH mechanism [34–36] studied in Section 5 in its Abelian version. In the case at hand, we need to introduce a SU(2) complex scalar doublet

$$\mathbf{H} = \begin{pmatrix} H^+ \\ H^0 \end{pmatrix}, \tag{9.30}$$

with $Y(\mathbf{H}) = \frac{1}{2}\mathbb{1}$, so using the Gell-Mann–Nishijima relation (9.3) we find that $H^+$ has charge $e$ and $H^0$ is neutral. We consider then the action

$$S_{\mathrm{Higgs}} = \int d^4x \left[ (D_\mu \mathbf{H})^\dagger D^\mu \mathbf{H} - \frac{\lambda}{4}\left( \mathbf{H}^\dagger \mathbf{H} - \frac{v^2}{2} \right)^2 \right], \tag{9.31}$$

where the covariant derivative is defined in (9.11). Although the action is fully $\mathrm{SU}(2) \times \mathrm{U}(1)_Y$ invariant, the potential has the Mexican hat shape shown in Fig. 9 and the field $\mathbf{H}$ gets a nonzero vev, that by a suitable gauge transformation can always be brought to the form

$$\langle \mathbf{H} \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{9.32}$$

This vev obviously breaks SU(2) and, having nonzero hypercharge, also $\mathrm{U}(1)_Y$. However, since $\langle H^+ \rangle = 0$ it nevertheless preserves the gauge invariance of electromagnetism. We have then the SSB pattern

$$\mathrm{SU}(2) \times \mathrm{U}(1)_Y \longrightarrow \mathrm{U}(1)_{\mathrm{em}}. \tag{9.33}$$

The masses of the gauge bosons are obtained by substituting the vev (9.32) into the action (9.31) and collecting the terms quadratic in the gauge fields. With this, we see that the $W$ and $Z$ bosons acquire nonzero masses given, respectively, by

$$m_W = \frac{ev}{2\sin\theta_w}, \qquad m_Z = \frac{ev}{\sin(2\theta_w)}, \tag{9.34}$$

and satisfying the custodial relation $m_W = m_Z \cos\theta_w$.

Interestingly, the scale $v$ is related to the Fermi constant $G_F$, a quantity that can be measured at low energies. Considering the neutron $\beta$-decay process in Eq. (9.16) at energies below the mass of the $W$ boson and comparing with the result obtained from the Fermi interaction

$$S_{\text{Fermi}} = \frac{G_F}{\sqrt{2}} \int d^4x\, \overline{\nu}_e \gamma_\mu (1-\gamma_5) e\, \overline{d}\gamma^\mu (1-\gamma_5) u, \tag{9.35}$$

we get the relation

$$G_F = \frac{\sqrt{2}}{8} \frac{e^2}{m_W^2 \sin^2\theta_w} = \frac{1}{\sqrt{2}v^2}, \tag{9.36}$$

where the expression of $m_W$ given in Eq. (9.34) has been used. Substituting now the experimental value of the Fermi constant $G_F = 1.166 \times 10^{-5}$ GeV$^2$ [117], we find

$$v \approx 246 \text{ GeV}. \tag{9.37}$$

In order to give mass to the fermions, we need to follow the strategy explained in page 70 and write the appropriate Yukawa couplings, which in this case read

$$\begin{aligned} S_{\text{Yukawa}} = -\sum_{i,j=1}^{3} \int d^4x \Big( & C_{ij}^{(\ell)} \overline{\mathbf{L}}^i \mathbf{H} \ell_R^j + C_{ji}^{(\ell)*} \overline{\ell}_R^i \mathbf{H}^\dagger \mathbf{L}^j + C_{ij}^{(q)} \overline{\mathbf{Q}}^i \mathbf{H} D_R^j + C_{ji}^{(q)*} \overline{D}_R^i \mathbf{H}^\dagger \mathbf{Q}^j \\ & + \widetilde{C}_{ij}^{(q)} \overline{\mathbf{Q}} \widetilde{\mathbf{H}} U_R^j + \widetilde{C}_{ji}^{(q)*} \overline{U}_R^i \widetilde{\mathbf{H}}^\dagger \mathbf{Q}^j \Big). \end{aligned} \tag{9.38}$$

The two terms in the second line involve the conjugate field

$$\widetilde{\mathbf{H}} \equiv i\sigma^2 \begin{pmatrix} H^{+*} \\ H^{0*} \end{pmatrix} = \begin{pmatrix} H^{0*} \\ -H^{+*} \end{pmatrix}, \tag{9.39}$$

which has $Y(\widetilde{\mathbf{H}}) = -\frac{1}{2}\mathbb{1}$ and can be seen to transform also as a $SU(2)$ doublet. Given the transformation properties of all fields involved, it is very easy to check that the action (9.38) is SU(2) $\times$ U(1)$_Y$ gauge invariant. Notice that here we are assuming that neutrino masses are not due to the BEH mechanism. This is the reason why lepton doublets only couple to the Higgs doublet $\mathbf{H}$, whose upper component has zero vev. In the case of quarks, however, we need to generate masses for both the upper and lower components of $\mathbf{Q}$. This is why they couple to the conjugate field $\widetilde{\mathbf{H}}$, whose upper component acquires a

nonzero vev

$$\langle \widetilde{\mathbf{H}} \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} v \\ 0 \end{pmatrix}. \tag{9.40}$$

To find the expression of the fermion masses generated by the BEH mechanism, we substitute in the Yukawa action the field $\mathbf{H}$ and its conjugate $\widetilde{\mathbf{H}}$ by their vevs (9.32) and (9.40). The resulting mass terms have the form

$$\begin{aligned} S_{\text{mass}} = -\int d^4x \Bigg[ & \left(\overline{e}_L, \overline{\mu}_L, \overline{\tau}_L\right) M^{(\ell)} \begin{pmatrix} e_R \\ \mu_R \\ \tau_R \end{pmatrix} + \left(\overline{d}_L, \overline{s}_L, \overline{b}_L\right) M^{(q)} \begin{pmatrix} d_R \\ s_R \\ b_R \end{pmatrix} \\ & + \left(\overline{u}_L, \overline{c}_L, \overline{t}_L\right) \widetilde{M}^{(q)} \begin{pmatrix} u_R \\ c_R \\ t_R \end{pmatrix} + \text{H.c.} \Bigg], \end{aligned} \tag{9.41}$$

where the mass matrices are given in term of the couplings in Eq. (9.38) by

$$M^{(\ell)}_{ij} = \frac{v}{\sqrt{2}} C^{(\ell)}_{ij}, \qquad M^{(q)}_{ij} = \frac{v}{\sqrt{2}} C^{(q)}_{ij}, \qquad \widetilde{M}^{(q)}_{ij} = \frac{v}{\sqrt{2}} \widetilde{C}^{(q)}_{ij}. \tag{9.42}$$

These complex matrices are however not necessarily diagonal, although they can be diagonalized through bi-unitary transformations

$$\begin{aligned} U^{(\ell)\dagger}_L M^{(\ell)} U^{(\ell)}_R &= \text{diag}(m_e, m_\mu, m_\tau), \\ V^{(q)\dagger}_L M^{(q)} V^{(q)}_R &= \text{diag}(m_d, m_s, m_b), \\ \widetilde{V}^{(q)\dagger}_L \widetilde{M}^{(q)} \widetilde{V}^{(q)}_R &= \text{diag}(m_u, m_c, m_t), \end{aligned} \tag{9.43}$$

where the eigenvalues are the leptons and quarks masses. Notice that fermion masses are determined by both the Higgs vev scale $v$ and the dimensionless Yukawa couplings $C^{(\ell)}_{ij}$, $C^{(q)}_{ij}$, and $\widetilde{C}^{(q)}_{ij}$, which are experimentally determined.

Let us focus for the time being on the quark sector (leptons will be dealt with below in section 9.4). Since $V^{(q)}_{L,R}$, $\widetilde{V}^{(q)}_{L,R}$ are constant unitary matrices we could use them to redefine the quark and lepton triplets in the total action

$$\begin{pmatrix} u'_{L,R} \\ c'_{L,R} \\ t'_{L,R} \end{pmatrix} = \widetilde{V}^{(q)\dagger}_{L,R} \begin{pmatrix} u_{L,R} \\ c_{L,R} \\ t_{L,R} \end{pmatrix}, \qquad \begin{pmatrix} d'_{L,R} \\ s'_{L,R} \\ b'_{L,R} \end{pmatrix} = V^{(q)\dagger}_{L,R} \begin{pmatrix} d_{L,R} \\ s_{L,R} \\ b_{L,R} \end{pmatrix}, \tag{9.44}$$

in such a way that the new fields are mass eigenstates, i.e., their free kinetic terms in the action have the standard diagonal form. A problem however arises when implementing this field redefinition in the interaction terms between the quarks and the $W^\pm$ gauge bosons, mixing the lower with upper components of the $SU(2)$ doublets. The issue is that, unlike in the kinetic terms, the matrices implementing the field

redefinition do not cancel

$$S \supset \int d^4x \, (\overline{u}_L, \overline{c}_L, \overline{t}_L)\gamma^\mu \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} W^+_\mu = \int d^4x \, (\overline{u}'_L, \overline{c}'_L, \overline{t}'_L)\widetilde{V}_L^{(q)\dagger} V_L^{(q)} \gamma^\mu \begin{pmatrix} d'_L \\ s'_L \\ b'_L \end{pmatrix} W^+_\mu, \quad (9.45)$$

where, to simplify the expression, the overall coupling is omitted and the corresponding coupling of the quarks to the $W^-$ boson is obtained by taking the Hermitian conjugate of this term. The combination

$$\widetilde{V}_L^{(q)\dagger} V_R^{(q)} \equiv V_{\text{CKM}} \quad (9.46)$$

defines the *Cabibbo–Kobayashi–Maskawa (CKM) matrix* [143] and determines the mixing among the quarks families. It is an experimental fact that this matrix is nondiagonal, so the emission/absorption of a $W^\pm$ boson does not merely transform the upper into the lower fields (or vice versa) *within* a single SU(2) quark doublet, but can also "jump" into another family. This gives rise to processes known as flavor changing charged currents. For example, there is a nonzero probability that a $u$ quark turns into a $s$ quark by the emission of a $W^+$, or vice versa with a $W^-$, accounting for decays like $\Lambda^0 \to p^+ e^- \overline{\nu}_e$. What happens inside the $\Lambda^0$ baryon ($uds$) is that the strange quark emits a $W^-$ and transforms into a $u$-quark, thus converting the $\Lambda^0$ into a proton ($uud$). The $W^-$ then decays into an electron and its antineutrino.

It is an interesting feature of the electroweak sector of the SM that there are no flavor changing *neutral* currents at tree level. In the case of electromagnetic-mediated processes, this follows from the fact that the field redefinitions induced by the matrices $V_{L,R}^{(q)}$ and $\widetilde{V}_{L,R}^{(q)}$ mix fields with the same electric charge, so they commute with the charge matrix $Q$ and cancel from the quark electromagnetic couplings. In the case of the weak neutral currents (mediated by the $Z^0$) the same happens, though maybe it is less obvious. Indeed, looking at the form of the covariant derivative (9.11) we find the following couplings between the quarks and the $Z^0$:

$$S \supset \int d^4x \left[ \left( \frac{1}{2} - \frac{2}{3}\sin^2\theta_w \right) (\overline{u}_L, \overline{c}_L, \overline{t}_L)\gamma^\mu \begin{pmatrix} u_L \\ c_L \\ t_L \end{pmatrix} - \left( \frac{1}{2} - \frac{1}{3}\sin^2\theta_w \right) (\overline{d}_L, \overline{s}_L, \overline{b}_L)\gamma^\mu \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} \right.$$
$$\left. + \frac{2}{3}\sin^2\theta_w (\overline{u}_R, \overline{c}_R, \overline{t}_R)\gamma^\mu \begin{pmatrix} u_R \\ c_R \\ t_R \end{pmatrix} - \frac{1}{3}\sin^2\theta_w (\overline{d}_R, \overline{s}_R, \overline{b}_R)\gamma^\mu \begin{pmatrix} d_R \\ s_R \\ b_R \end{pmatrix} \right], \quad (9.47)$$

where again we have dropped an overall constant which is irrelevant for the argument. What matters for our discussion is that, after the field redefinition, we get the combinations $V_{L,R}^{(q)\dagger} V_{L,R}^{(q)} = \mathbb{1} = \widetilde{V}_{L,R}^{(q)\dagger} \widetilde{V}_{L,R}^{(q)}$ and no mixing matrix is left behind. This shows that there are no flavor changing neutral currents at tree level[28].

[28] Once quantum effects are included, flavor changing neutral currents are suppressed due to the flavor mixing brought about by the Cabibbo–Kobayashi–Maskawa matrix, via the so-called GIM (Glashow–Iliopoulos–Maiani) mechanism [144].

**Box 14. SSB or QCD?**

We have seen how the BEH mechanism provides the rationale to understand how the particles in the SM acquire their masses, a scenario ultimately confirmed by the experimental detection of the Higgs boson. But, does the BEH mechanism really explains the mass of everything we see around us, from the paper in our hands to the sun over our heads? The answer is no. As we will see, the fraction of the mass of macroscopic objects that we can assign to the Higgs boson acquiring a vev is really tiny.

We know that the masses of protons and neutrons are very similar to one another, and much larger than the mass of the electron

$$m_p \simeq m_n \simeq 1836\, m_e. \tag{9.48}$$

In turn, the mass of a $(A, Z)$ nucleus is

$$M(A,Z) = Zm_p + (A-Z)m_n + \Delta M(A,Z), \tag{9.49}$$

with $\Delta M(A,Z)$ the binding energy, which varies from a bit over $1\%$ for deuterium to around $10\%$ for $^{62}_{28}$Ni. Taking Eq. (9.48) into account and to a fairly good approximation, the mass of an atom can be written in terms of its mass number alone

$$m(A,Z) \simeq Am_p. \tag{9.50}$$

The point of this argument is to show that in order to explain the mass around us we essentially need to explain the mass of the proton. But here we run into trouble if we want to trace back $m_p$ to the BEH mechanism. The values of the masses of the $u$ and $d$ quarks accounted for by the BEH mechanism (the so-called current algebra masses) are

$$m_u \simeq 2.2\ \text{MeV}, \qquad m_d = 4.7\ \text{MeV}. \tag{9.51}$$

Comparing with $m_p[uud] \simeq 938.3$ MeV and $m_d[udd] = 939.6$ MeV, we see that quark masses only explain about $1\%$ of the nucleon mass. Thus, close to $99\%$ of the mass in atomic form in the universe is not due to the BEH mechanism.

Where does this mass/energy come from? Actually, from QCD effects. Protons and neutrons are not only made out of their three valence quarks, but they are filled with a plethora of virtual quarks and gluons fluctuating in and out of existence whose energy make up the missing $99\%$. These effects can be computed numerically using lattice field theory [145, 146]. Here, however, we just want to offer some general arguments pointing to the origin of the difficulties in describing protons and neutrons in terms of their constituent quarks.

Let us begin with a very simple argument. We know that because of the strong dynamics of QCD at low energies quarks get confined into hadrons in a region whose linear size is of the

order $\Lambda_{\rm QCD}^{-1}$. Applying Heisenberg's uncertainty principle, we can estimate the size of their momentum fluctuations to be about

$$\Delta p \sim \Lambda_{\rm QCD}. \tag{9.52}$$

If fluctuations are isotropic the statistical average of the quark momentum vanishes, $\langle \mathbf{p} \rangle = 0$. Since $(\Delta p)^2 \equiv \langle \mathbf{p}^2 \rangle - \langle \mathbf{p} \rangle^2$, we determine the averaged quark momentum squared to be

$$\langle \mathbf{p}^2 \rangle \sim \Lambda_{\rm QCD}^2. \tag{9.53}$$

Now, $\Lambda_{\rm QCD}$ is of the order of a few hundred MeV, so the masses of the $u$ and $d$ quarks satisfy $m_u, m_d \ll \Lambda_{\rm QCD}$. This means that the linear momenta of the valence quarks inside protons and neutrons is much larger than their masses, so they are relativistic particles. Moreover, since their typical energy is of order $\Lambda_{\rm QCD}$, they are in the low energy regime of QCD where the dynamics is strongly coupled.

What we said about the $u$ and $d$ quarks does not apply however to the top ($m_t \simeq 173.7$ GeV), bottom ($m_b \simeq 4.6$ GeV), and charm ($m_c \simeq 1.3$ GeV) quarks, which under the same conditions would behave as nonrelativistic particles. Besides, since their energies are dominated by their masses, which are well above $\Lambda_{\rm QCD}$, their QCD interactions are weakly coupled. This is why heavy quark bounds states (quarkonium) can be analytically studied using perturbation theory, unlike the bound states of light quarks ($u$, $d$, and $s$) that have to be treated numerically. The difficulties in describing quarks inside protons and neutrons boils down to them being ultrarelativistic particles.

The moral of the story is that the popular line that the BEH mechanism "explains" mass is simply not correct. Most of our own mass and the mass of every object we see around us (and this includes the Earth, the Sun, the Moon, and the stars in the sky) has nothing to do with the Higgs field and is the result of the quantum behavior of the strong interaction. Even in a universe where the up and down quarks were massless, the proton and the neutron would still have nonzero masses and moreover very similar to the ones in our world.

### 9.3 The Higgs boson

In order to analyze mass generation in the electroweak sector of the SM, it was enough to replace the scalar doublet $\mathbf{H}$ by its vev. However, as we learned in Section 5.4 for the Abelian case, the system has excitations around the minimum of the potential corresponding to a propagating scalar degree of freedom. To analyze the dynamics of this field, the *Higgs boson*, we write the Higgs doublet $\mathbf{H}$ as

$$\mathbf{H}(x) = \frac{1}{\sqrt{2}} e^{i a^I(x) t^I_{\mathbf{2}}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \tag{9.54}$$

where $a^I(x)$ and $h(x)$ are the four real degrees of freedom encoding the two complex components in (9.30). In fact, as in the Abelian case of Section 5.4, we can use the gauge invariance of $S_{\rm Higgs} + S_{\rm Yukawa}$ to eliminate the global SU(2) global factor, after which we are left with a single real degree of

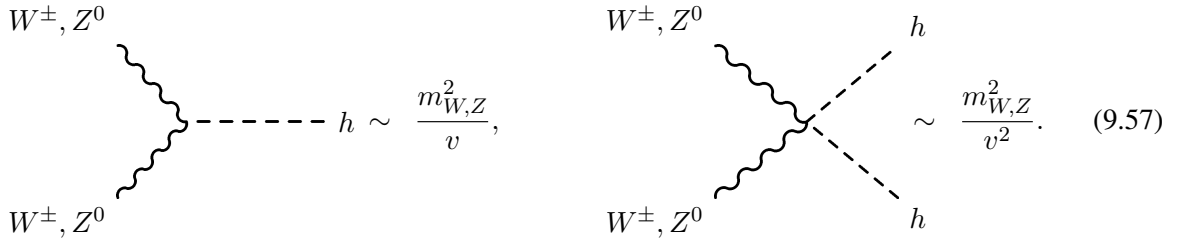
freedom representing the Higgs boson [36]. Substituting into (9.31) and expanding, we get

$$S_{\text{Higgs}} = \int d^4x \left[\frac{1}{2}\partial_\mu h \partial^\mu h - \frac{\lambda v^2}{4}h^2 - \frac{\lambda v}{4}h^3 - \frac{\lambda}{16}h^4 + \frac{2m_W^2}{v}W_\mu^- W^{+\mu} h \right. \tag{9.55}$$
$$\left. + \frac{m_W^2}{v^2}W_\mu^- W^{+\mu}h^2 + \frac{m_Z^2}{v}Z_\mu Z^\mu h + \frac{m_Z^2}{2v^2}Z_\mu Z^\mu h^2 + m_W^2 W_\mu^+ W^{-\mu} + \frac{m_Z^2}{2}Z_\mu Z^\mu\right],$$

where in the last two terms we recognize the masses for the $W^\pm$ and $Z^0$ gauge bosons. The first thing to be noticed is that the mass of the Higgs boson is determined by the vev $v$ and the strength $\lambda$ of the Higgs quartic self-couplings,

$$m_H = v\sqrt{\frac{\lambda}{2}} = (125.25 \pm 0.17)\ \text{GeV}, \tag{9.56}$$

where the current average experimental value is quoted [117]. The action (9.55) also contains the coupling between the Higgs boson and the $W^\pm$ and $Z^0$ intermediate bosons, giving rise to the interaction vertices



(9.57)

In both cases, the strength of the coupling is proportional to the mass squared of the corresponding intermediate bosons.

As to the coupling of the Higgs boson to fermions, this is obtained by replacing (9.54) into the Yukawa action (9.38),

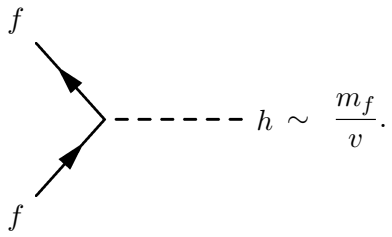
$$S_{\text{Yukawa}} = -\int d^4x \left[(\overline{e}_L, \overline{\mu}_L, \overline{\tau}_L)\left(\frac{1}{v}M^{(\ell)}\right)\begin{pmatrix} e_R \\ \mu_R \\ \tau_R \end{pmatrix} h \right. \tag{9.58}$$
$$\left. + (\overline{d}_L, \overline{s}_L, \overline{b}_L)\left(\frac{1}{v}M^{(q)}\right)\begin{pmatrix} d_R \\ s_R \\ b_R \end{pmatrix} h + (\overline{u}_L, \overline{c}_L, \overline{t}_L)\left(\frac{1}{v}\widetilde{M}^{(q)}\right)\begin{pmatrix} u_R \\ c_R \\ t_R \end{pmatrix} h + \text{H.c.}\right].$$

This, upon switching to mass eigenstates, takes the general form

$$S_{\text{Yukawa}} = -\sum_f \frac{m_f}{v}\int d^4x\, \overline{f} f h, \tag{9.59}$$

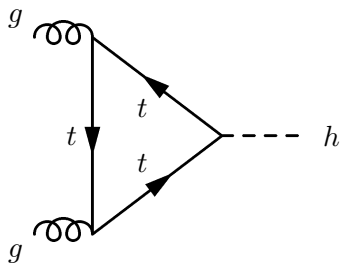
where $f = (e', \mu', \tau', u', d', c', s', t', b')$ runs over all the fermion mass eigenstates, apart from the three

neutrinos that we will treat separately. The corresponding interaction vertices are



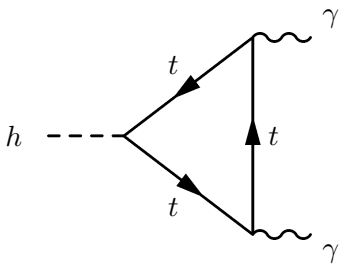

$$\sim \frac{m_f}{v}. \tag{9.60}$$

That the coupling of the Higgs boson to the fermions is proportional to their masses has important experimental consequences. Given the value of the Higgs vev energy scale found in (9.37), only the heaviest fermions have sizeable Higgs couplings, in particular the top quark with mass $m_t = 173.3$ GeV [117]. This fact is at the heart of the experimental strategy that culminated with the observation of the Higgs boson at CERN. In a hadron collider such as the LHC, there are plenty of gluons produced during the collision that can fuse through a top quark loop to produce a Higgs boson



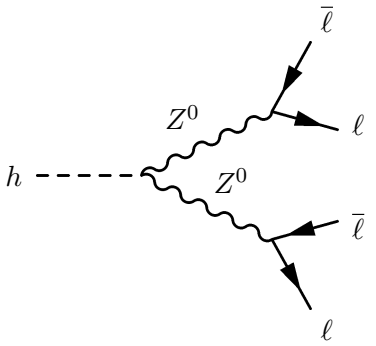

(9.61)

The Higgs boson produced in the gluon fusion process can decay in various distinctive ways. One of them is by a second top loop with emission of two photons




(9.62)

Alternatively, the Higgs boson may produce a pair of $Z^0$ bosons that in turn decay into two lepton–antilepton pairs




(9.63)

These were precisely the decay channels that led to the discovery of the Higgs boson by the ATLAS and CMS collaborations at the LHC [19, 20].

### 9.4 Neutrino masses

We have been postponing the issue of neutrinos masses. It is however an experimental fact that neutrinos have nonzero masses and this is something we have to incorporate in the SM action. One way to do it is to extend the SM to include right-handed *sterile* neutrinos $\nu_R^i$ transforming as $(\mathbf{1},\mathbf{1})_0$ under $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)_Y$ (see the notation introduced on page 105), adding then the following terms to the Yukawa action

$$\Delta S_{\text{Yukawa}} = -\sum_{i=1}^{3}\int d^4x\left(\widetilde{C}^{(\nu)}\overline{\mathbf{L}}^i\widetilde{\mathbf{H}}\nu_R^i + \widetilde{C}^{(\nu)*}_{ji}\overline{\nu}_R^i\widetilde{\mathbf{H}}\mathbf{L}^j\right). \tag{9.64}$$

Once the Higgs field gets a vev, this term generates a mass term of the form

$$\Delta S_{\text{Yukawa}} = -\int d^4x\left[(\overline{\nu}_{eL},\overline{\nu}_{\mu L},\overline{\nu}_{\tau L})\widetilde{M}^{(\nu)}\begin{pmatrix}\nu_{1R}\\ \nu_{2R}\\ \nu_{3R}\end{pmatrix} + \text{H.c.}\right], \tag{9.65}$$

with

$$M_{ij}^{(\nu)} = \frac{v}{\sqrt{2}}\widetilde{C}_{ij}^{(\nu)}. \tag{9.66}$$

Being singlets under all SM gauge groups, the sterile neutrinos only interact gravitationally with other particles.

**Box 15. Dirac vs. Majorana fermions**

In previous sections, we have shown how antiparticles in QFT are somehow related to complex fields, for example in the complex scalar field discussed in Box 6 (see page 37). In this case, particles are interchanged with antiparticles by replacing the field $\varphi(x)$ with its complex conjugate $\varphi(x)^*$. To make things more elegant, we may call this operation *charge conjugation* and the result the *charge conjugated field*

$$\mathrm{C}:\varphi(x)\longrightarrow\eta_C\varphi(x)^*\equiv\varphi^c(x), \tag{9.67}$$

where $\eta_C$ is some phase that we are always free to add while keeping the action (3.86) invariant. At the quantum level, C does indeed interchange particles and antiparticles

$$\mathrm{C}|\mathbf{p};0\rangle = \eta_C^*|0;\mathbf{p}\rangle, \qquad \mathrm{C}|0;\mathbf{p}\rangle = \eta_C|\mathbf{p};0\rangle. \tag{9.68}$$

From this perspective, a *real* scalar field is one identical to its charge conjugate, $\varphi(x)=\varphi^c(x)$. After quantization, its elementary excitations are their own antiparticles.

Let us try to make something similar with the Dirac field. In the scalar field case, replacing $\varphi(x)$ by $\varphi(x)^*$ does not change the field's Lorentz transformation properties, after all, complex conjugate or not, both fields are *scalars*. Not so for a Dirac fermion. The spinor $\psi(x)$ and its complex conjugate $\psi(x)^*$ do not transform the same way under the Lorentz group and neither satisfy the same Dirac equation. This means that we cannot define a "real" Dirac spinor just requiring $\psi(x) = \psi(x)^*$. We have to work a little bit more and consider

$$\mathrm{C} : \psi(x) \longrightarrow \eta_C(-i\gamma^2)\psi(x)^* \equiv \psi^c(x), \tag{9.69}$$

where again $\eta_C$ is a complex phase. This charge conjugate spinor transforms in the same way as the original field and also satisfies the same free Dirac equation. Moreover, its action on the multiparticle states generated by the creation operators $\widehat{b}(\mathbf{k}, s)^\dagger$ and $\widehat{d}(\mathbf{k}, s)^\dagger$ in Eq. (4.56) is given by

$$\mathrm{C}|\mathbf{k}, s; 0\rangle = \eta_C^*|0; \mathbf{k}, s\rangle, \qquad \mathrm{C}|0; \mathbf{k}, s\rangle = \eta_C|\mathbf{k}, s; 0\rangle, \tag{9.70}$$

and interchanges particles and antiparticles.

The spinor analog of the real scalar field is a *Majorana spinor*, which equals its charge conjugate

$$\psi(x) = \psi^c(x). \tag{9.71}$$

Upon quantization, this identifies particles and antiparticles, as follows from Eq. (9.70). It is interesting to implement the Majorana condition expressing the Dirac fermion in terms of its chiral components and using the representation (4.47) of the Dirac matrices

$$\begin{pmatrix} \chi_+ \\ \chi_- \end{pmatrix} = \eta_C \begin{pmatrix} i\sigma^2\chi_-^* \\ -i\sigma^2\chi_+ \end{pmatrix} \qquad \Longrightarrow \qquad \psi = \frac{1}{\sqrt{2}} \begin{pmatrix} \chi_+ \\ -i\eta_C\sigma^2\chi_+^* \end{pmatrix}. \tag{9.72}$$

In the second identity we wrote a solution to (9.71), and a similar expression can be written in terms of the negative chirality component $\chi_-$. Here we see how the Majorana condition halves the four complex components of a Dirac field down to two. In fact, the Majorana spinor can be written as the sum of a Weyl fermion and its charge conjugate as

$$\psi = \frac{1}{\sqrt{2}} \begin{pmatrix} \chi_+ \\ 0 \end{pmatrix} + \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ -i\eta_C\sigma^2\chi_+ \end{pmatrix} \equiv \frac{1}{\sqrt{2}}\left(\psi_+ + \psi_+^c\right). \tag{9.73}$$

Using this expression, we write the Dirac action for a Majorana fermion

$$S = \int d^4x \left[ i\overline{\psi}_+ \partial\!\!\!/ \psi_+ - \frac{m}{2}\left(\overline{\psi_+^c}\psi_+ + \overline{\psi}_+\psi_+^c\right)\right]. \tag{9.74}$$

Unlike Weyl fermions, Majorana spinors admit a mass term without doubling the number of degrees of freedom.

An important point concerning Majorana fermions is that they cannot be coupled to the elec-

tromagnetic field. This is to be expected, since the Majorana condition identifies particles with antiparticles that, as we saw in Box 7, have opposite electric charge. In more precise terms what happens is that the associated Noether current vanishes

$$j^\mu = \overline{\psi}\gamma^\mu\psi = \frac{1}{2}\left(\chi_+^\dagger\sigma_+^\mu\chi_+ + \chi_+^T\sigma_+^{\mu T}\chi_+^*\right) = 0. \tag{9.75}$$

This can be also seen as a consequence of the incompatibility of the Majorana condition (9.71) with a global U(1) phase rotation of the spinor $\psi \to e^{i\vartheta}\psi$. In particular, the Majorana mass term in (9.74) does not conserve the U(1) charge

$$\overline{\psi_+^c}\psi_+ + \overline{\psi}_+\psi_+^c \longrightarrow e^{2i\theta}\overline{\psi_+^c}\psi_+ + e^{-2i\theta}\overline{\psi}_+\psi_+^c, \tag{9.76}$$

a very important feature for the accidental symmetries of the SM such as lepton number.

The addition of sterile neutrinos to generate neutrino masses is only partly satisfactory. One obvious problem is its lack of economy, since it requires the addition of extra species to the SM that nevertheless do not partake in its interactions. But the solution is also unnatural. Due to the smallness of the neutrino masses, the new Yukawa couplings have to be many orders of magnitude smaller than the ones for charged leptons.

Generating a Dirac mass term is not the only possibility of accounting for neutrino masses. Having zero electric charge, they are the only fermions in the SM that can be of Majorana type. If this were the case, their mass terms in the action would be build from the left components alone, as we saw in Box 15

$$\Delta S = -\sum_{i,j=1}^{3}\int d^4x\left(\frac{1}{2}M_{ij}\overline{\nu^{ic}_L}\nu^j_L + \text{H.c.}\right), \tag{9.77}$$

where because of Fermi statistics $\overline{\nu^{ic}_L}\nu^j_L = \overline{\nu^{jc}_L}\nu^i_L$ and the mass matrix $M^{(\nu)}_{ij}$ can be taken to be symmetric. The problem now lies in how to generate a Majorana mass from a coupling of the neutrinos to the Higgs field, since both $\mathbf{L}^i$ and its charge conjugate are SU(2) doublets and there is no way to construct a gauge invariant *dimension four* operator involving $\mathbf{L}^i$, $\mathbf{L}^{ic}$, and $\mathbf{H}$ (or $\widetilde{\mathbf{H}}$). A group-theoretical way to see this is by noticing that the product representation $\mathbf{2}\otimes\mathbf{2}\otimes\mathbf{2} = \mathbf{4}\oplus\mathbf{2}\oplus\mathbf{2}$ does not contain any SU(2) singlet. This changes if we admit a dimension-five operator with two Higgs doublets, a left-handed fermion and its charge conjugate. Now it is possible to construct a gauge invariant term since $\mathbf{2}\otimes\mathbf{2}\otimes\mathbf{2}\otimes\mathbf{2} = \mathbf{5}\oplus\mathbf{3}\oplus\mathbf{3}\oplus\mathbf{3}\oplus\mathbf{1}\oplus\mathbf{1}$. For example,

$$\Delta S = -\frac{1}{M}\sum_{i,j=1}^{3}\int d^4x\left[C^{(\nu)}_{ij}\left(\overline{\mathbf{L}^{ic}}\,\widetilde{\mathbf{H}}^*\right)\left(\widetilde{\mathbf{H}}^\dagger\mathbf{L}^j\right) + \text{H.c.}\right] \tag{9.78}$$

is invariant under SU(2) × U(1)$_Y$. This operator in the action has to be understood, in the spirit of EFT, as the result of some new physics appearing at the energy scale $M \gg v$, with $v$ the Higgs vev.

When the Higgs field acquires its vev, the coupling (9.78) generates a Majorana mass term for the

neutrinos,

$$\Delta S = -\frac{1}{2}\sum_{i,j=1}^{3}\int d^4x\Big(M_{ij}^{(\nu)}\overline{\nu^{ic}}_L\nu_L^j + \text{H.c.}\Big), \tag{9.79}$$

where the neutrino mass matrix is given by

$$M_{ij}^{(\nu)} = \frac{v^2}{M}C_{ij}^{(\nu)}. \tag{9.80}$$

The entries of this matrix are suppressed by the factor $v/M \ll 1$, naturally producing neutrinos with masses well below the ones of the charged leptons. Thus, Majorana neutrinos not only are the most economical solution, making unnecessary adding new fermion species, but also avoids the unnaturalness of the neutrino Yukawa couplings. Incidentally, the Majorana mass term (9.79) violates lepton number, since $\nu_L^j$ and $\overline{\nu^{ic}}_L$ transform with the same phase [cf. (9.76)].

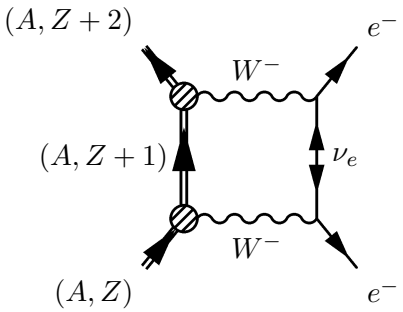
Neutrinos are regarded as one of the most promising windows to physics beyond the SM, being the main reason why neutrino physics has remained for decades one of the most exciting fields in (astro)particle physics and cosmology [147–149]. As to the question of whether the neutrino is a Dirac or a Majorana particle, however, the jury is still out. Some processes can only take place if the neutrino is its own antiparticle, most notably neutrinoless double $\beta$ decay [150, 151]. A nucleus with mass and atomic numbers $(A, Z)$ can undergo double $\beta$-decay and transmute into the nucleus $(A, Z+2)$ with emission of two electrons and two antineutrinos:

$$\begin{array}{rcl}(A,Z) & \longrightarrow & (A,Z+1) + e^- + \overline{\nu}_e \\ & & \hookrightarrow \quad (A,Z+2) + e^- + \overline{\nu}_e\end{array}. \tag{9.81}$$

If the neutrino is a Majorana particle there is an alternative. The neutrino produced in the first decay may interact with a neutron in the nucleus, turning it into a proton with the emission of an electron,

$$\overline{\nu}_e(\equiv \nu_e) + n \longrightarrow p^+ + e^-, \tag{9.82}$$

so no neutrino is emitted in the process $(A, Z) \to (A, Z+2) + 2e^-$. This is described by the diagram




(9.83)

where the double-arrowed line represents the Majorana neutrino. The detection of neutrinoless double $\beta$-decay would decide the question of the Dirac or Majorana character of the neutrino. A lot of experimental effort is being dedicated to this problem, so far without definite results (see Ref. [152] for an updated overview of past, present, and future experiments).

**Box 16. CP violation and the CKM and PMNS matrices**

When studying the strong CP problem in Section 8.2, we hinted at the fact that CP violation is associated with the existence of complex couplings in the action. This is shown easily, taking into account that the CP transformation acting on an operator $\mathscr{O}$ transforms it into its Hermitian conjugate, $\mathrm{CP}\mathscr{O}(\mathrm{CP})^{-1} = \mathscr{O}^\dagger$. Hence, a term in the Hamiltonian of the form $g\mathscr{O} + g^*\mathscr{O}^\dagger$, although being Hermitian, leads to CP violation unless the coupling is real, $g = g^*$. This is why when exploiting the axial anomaly to move the $\theta$ dependence in the QCD action from the $\theta$-term into a complex phase in the fermion mass matrix we said that we were *shifting* the source of CP-violation to a complex coupling.

Besides the $\theta$-term in the QCD action, it is a fact that CP symmetry is broken in the electroweak sector of the SM, for example in neutral kaon decays. Its origin is found in the unitary CKM matrix

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{9.84}$$

introduced in (9.46) since, as we will see now, it contains a complex phase that cannot be removed by redefinition of the quark fields. Let us be general and analyze the case of a SM with $n$ families. An $n \times n$ unitary matrix depends on $n^2$ real parameters (the $2n^2$ real parameter of a general complex matrix reduced by the $n^2$ conditions imposed by unitarity). In addition to this, we can play with the phases of the $2n$ quarks, keeping in mind the invariance of the action under a common phase redefinition of all quark fields leading to (perturbative) baryon number conservation. This means that $2n-1$ of the $n^2$ real parameters can be absorbed in the phases of the quark fields, and we are left with $n^2 - 2n + 1 = (n-1)^2$ independent ones. The question is how many of them correspond to complex phases. To decide this, let us recall that were the CKM matrix real it would be an SO($N$) matrix depending on $\frac{1}{2}n(n-1)$ real angles. Subtracting this number from the total number of independent real parameters computed above, we get the final number of complex phases in the CKM matrix to be

$$n^2 - 2n + 1 - \frac{1}{2}n(n-1) = \frac{1}{2}(n-1)(n-2). \tag{9.85}$$

For three families ($n = 3$) the matrix depends on a single complex phase $e^{i\delta}$ and three real angles $\theta_{12}$, $\theta_{13}$, and $\theta_{23}$. In terms of them, the CKM matrix is usually parametrized as

$$V_{\mathrm{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{9.86}$$

where $s_{ij} \equiv \sin\theta_{ij}$ and $c_{ij} \equiv \cos\theta_{ij}$. The modulus of the entries can be measured through the observation of various weak interaction mediated decays and scattering processes (see for example

Ref. [153]), with the result [117]

$$|V_{\rm CKM}| = \begin{pmatrix} 0.97435 \pm 0.00016 & 0.22500 \pm 0.00067 & 0.00369 \pm 0.00011 \\ 0.22486 \pm 0.00067 & 0.97349 \pm 0.00016 & 0.04182^{+0.00085}_{-0.00074} \\ 0.00857^{+0.00020}_{-0.00018} & 0.04110^{+0.00083}_{-0.00072} & 0.999118^{+0.000031}_{-0.000036} \end{pmatrix}, \tag{9.87}$$

while the value of the CP-violating phase is $\delta = 1.144 \pm 0.027$. The experimental measurement of $|V_{\rm CKM}|$ exhibits a clear hierarchy among its entries, derived from $s_{13} \ll s_{23} \ll s_{12} \ll 1$. This is manifest in the so-called Wolfenstein parametrization [154]

$$V_{\rm CKM} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \tag{9.88}$$

where $\lambda \equiv s_{12}$. The diagonal elements are all of order one, whereas the size of the other entries decreases as we move away from it.

A look at (9.85) shows that with just two families the corresponding flavor mixing matrix would contain no complex phases and depend on a single real parameter, the Cabibbo angle $\theta_C \equiv \theta_{12}$ [155]. Thus, CP violation in the electroweak sector, like the one showing up in for example kaon decays, requires the existence of at least three SM families.

CP-violation in the SM is of major importance, since it is a basic ingredient to explain why there is such a tiny amount of antimatter in our universe. However, the amount of CP violation produced by the single complex phase of the CKM matrix is far too small to account for the observed matter–antimatter asymmetry [156]. Finding additional sources in or beyond the SM is one of the big open problems in contemporary high energy physics.

Maybe the lepton sector is a good place to look for more CP violation. As with quarks, lepton masses appear when switching from interaction to mass eigenstates by diagonalizing the lepton mass matrix. Redefining the massive lepton fields

$$\begin{pmatrix} e'_{L,R} \\ \mu'_{L,R} \\ \tau'_{L,R} \end{pmatrix} = U^{(\ell)}_{L,R} \begin{pmatrix} e_{L,R} \\ \mu_{L,R} \\ \tau_{L,R} \end{pmatrix} \tag{9.89}$$

with $U^{(\ell)}_{L,R}$ defined in Eq. (9.43), the interaction terms with the $W^\pm$ bosons take the form

$$S \supset \int d^4x \left[ (\overline{e}'_L, \overline{\mu}'_L, \overline{\tau}'_L) U^{(\ell)\dagger}_L \gamma^\mu \begin{pmatrix} \nu_{eL} \\ \nu_{\mu L} \\ \nu_{\tau L} \end{pmatrix} W^+_\mu + \text{H.c.} \right]. \tag{9.90}$$

Here, the Hermitian conjugate term contains the interaction with the $W^-$ and we have dropped the global normalization. In the original version of the SM there are no right-handed neutrinos and therefore we can reabsorb the matrix $U^{(\ell)\dagger}_L$ in a redefinition of the left-handed neutrino fields,

without it appearing elsewhere in the SM action. As a result, if the neutrino were massless there would be no flavor mixing in the lepton sector.

Things are drastically different once we add the neutrino mass terms. Let us consider first the case of Dirac masses. As with quarks and charged leptons, the mass matrix in Eq. (9.66) can be diagonalized by a bi-unitary transformation

$$U_L^{(\nu)\dagger} M^{(\nu)} U_R^{(\nu)} = \text{diag}(m_1, m_2, m_3), \tag{9.91}$$

and the interaction term (9.90) is recast in terms of neutrino mass eigenstates as

$$S \supset \int d^4x \left[ (\overline{e}'_L, \overline{\mu}'_L, \overline{\tau}'_L) U_L^{(\ell)\dagger} U_L^{(\nu)} \gamma^\mu \begin{pmatrix} \nu_{1L} \\ \nu_{2L} \\ \nu_{3L} \end{pmatrix} W_\mu^+ + \text{H.c.} \right], \tag{9.92}$$

where

$$U \equiv U_L^{(\ell)\dagger} U_L^{(\nu)} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix}, \tag{9.93}$$

is the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) unitary matrix [157, 158]. Similarly to what the CKM matrix does for quarks, the PMNS matrix introduces flavor mixing in the leptonic sector. Moreover, following the same reasoning as with the CKM matrix, we see that for three families the PMNS matrix also depends on three real angles and a single complex phase, representing an additional source of CP violation. It also admits a parametrization similar to the one shown in Eq. (9.86) for the CKM matrix, where the phase is denoted by $\delta_{\text{CP}}$.

For Majorana neutrinos, however, the mass matrix (9.80) is symmetric and can be diagonalized by a *unitary* transformation

$$U_L^{(\nu)T} M U_L^{(\nu)} = \text{diag}(m_1, m_2, m_3), \tag{9.94}$$

so switching to neutrino mass eigenstates we find again an interaction term of the form (9.92). The big difference with respect to the Dirac case is that, since the Majorana mass term (9.79) is not invariant under phase rotations of the neutrino fields, we cannot get rid of two of three phases in the PMNS matrix. As a consequence, besides the three angles $\theta_{12}$, $\theta_{13}$, $\theta_{23}$ and the phase $e^{i\delta_{\text{CP}}}$ of the Dirac case, the matrix depends now on two additional complex phases $e^{i\lambda_1}$ and $e^{i\lambda_2}$, known as Majorana phases. The three angles and $\delta_{\text{CP}}$ can be measured from the neutrino oscillations, whereas the measurement of the two Majorana phases would be possible through the observation of neutrinoless double $\beta$ decay [152]. Fits of neutrino data (including the Super-Kamiokande atmospheric neutrino

data) give the following $3\sigma$ ranges for the absolute values of the entries of the PMNS matrix [159]

$$|U| = \begin{pmatrix} 0.801 \to 0.845 & 0.513 \to 0.579 & 0.143 \to 0.155 \\ 0.234 \to 0.500 & 0.471 \to 0.689 & 0.637 \to 0.776 \\ 0.271 \to 0.525 & 0.477 \to 0.694 & 0.613 \to 0.756 \end{pmatrix}. \tag{9.95}$$

It is interesting to compare the textures of the matrices (9.88) and (9.95). As already mentioned, for quarks the matrix is of order 1 at the diagonal, $\lambda$ for the second diagonal, and $\lambda^2$ in the upper right and lower left corners. There seems to be a hierarchical pattern (this is a bit of wishful thinking, clearly). In the case of neutrinos, however, it seems that there is democracy in all its entries, and a crude approximation to (9.95) would be to set all its entries to 1. This is a matrix with a single nonzero eigenvalue and two degenerate zeros, reminiscent of the normal or inverted hierarchies in the fit of the neutrino masses. Both textures are so different that it is difficult to imagine that they have a common origin. A major mystery, whose clarification is beyond the SM.

## 10 Scale invariance and renormalization

Renormalization appeared in physics as a way to make sense of the divergent results in QFT. In quantum mechanics, infinities are usually handled by invoking a normal ordering prescription, and even in QFT, they are absent when computing semiclassical contributions to processes in perturbation theory[29]. The trouble comes when calculating quantum corrections, associated in the perturbative expansion to Feynman diagrams with closed loops. These contain integrals over all independent momenta running in the loops that are frequently divergent.

We will not enter into the many details and subtleties involved in the study of divergences in QFT and the philosophy and practicalities of renormalization. They are explained in all major textbooks on the subject and a concise and not too technical overview can be found in Chapter 8 of Ref. [14]. The first step is to make the divergent integrals finite in order to handle them mathematically. This is done by introducing a proper regulator, that can either be a scale where loop momenta are cut off or a more abstract procedure to render the integrals finite, such as playing with the dimension of spacetime or introducing PV fermions. In any case, regularization implies the introduction of an energy scale $\Lambda$, called the cutoff for short. The basic point is that this cutoff is an artefact of the calculation and cannot appear in any *physical* quantity that we compute.

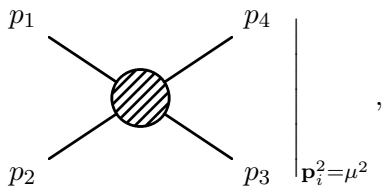
Roughly speaking, renormalization consists on getting rid of the cutoff. The key point to do this is the realization that the masses, couplings, and the fields themselves appearing in the classical action are not physical quantities. Therefore, there is nothing wrong with them depending on $\Lambda$. What must be cutoff independent are the physical quantities that we compute and can (and will) be compared with experiments. These quantities are *operationally defined*, in the sense that their definition within the theory's framework is given in terms of the process to be used to measure them. An example is the

[29] Here we are going to be concerned with UV divergences associated with the high energy regime of the theory. IR divergences, which appear in the limit of low momenta, cancel once the physical question is properly posed and all contributions to the given process are taken into account.

self-interacting scalar theory

$$S = \int d^4x \left( \frac{1}{2} \partial_\mu \varphi \partial^\mu \varphi - \frac{m^2}{2} \varphi^2 - \frac{\lambda}{4!} \varphi^4 \right), \tag{10.1}$$

where we would like to define the physical coupling $\lambda_{\text{phys}}$. We could identify it as the value of the scattering amplitude for four scalar particles when all $\mathbf{p}_i^2$ are equal

$\lambda_{\text{phys}} \equiv$ 

, (10.2)

where the blob stands for all diagrams contributing at a given order in perturbation theory and $\mu$ is the energy scale of the process. The dependence of the action parameters on $\Lambda$ is then chosen so this renormalization condition remains cutoff independent. Once this is done not just for the coupling constant but also for *all* physical quantities (e.g., masses), the theory is renormalized and everything can be computed in terms of experimentally defined physical couplings and masses.

In the case of the scalar theory defined by the action (10.1), as well as in other physically relevant theories like QED, QCD or the SM as a whole, it is possible to get rid of the cutoff dependence in any physical process by "hiding" it in a *finite* number of parameters. Those theories for which this can be accomplished are called renormalizable. Nonrenormalizable theories, on the other hand, require the introduction of an infinite number of parameters to absorb the cutoff dependence, that in turn means that we need to specify an infinite number of operationally-defined physical quantities. In this picture, nonrenormalizability seems quite a disaster, since it seems that to compute physical observables we need to specify an infinite number of physical renormalization conditions. This is the reason why, historically, nonrenormalizable theories were considered to be no good for physics.

Regularization and renormalization may have important consequences for classical symmetries, and we have seen examples of this in Section 7. One of the immediate consequences of regularization is the necessity of introducing a cutoff in the theory and therefore an energy scale. This has the result that, after renormalization, the physical couplings acquire a dependence on the energy scale where they are measured. This scale dependence is codified in the $\beta$ function, containing information on how the coupling constant $g$ depends on the scale where it is measured,

$$\beta(g) \equiv \mu \frac{dg}{d\mu}. \tag{10.3}$$

This function can be computed order by order in perturbation theory. In QCD $\beta(g) < 0$, which means that the coupling constant decreases as the energy grows, a property known as asymptotic freedom. Besides, the theory dynamically generates an energy scale $\Lambda_{\text{QCD}}$ below which it becomes strongly coupled, with quarks and gluons confined into mesons and baryons. Asymptotic freedom is the reason behind QCD's success as a description of strong interactions. It allows us to understand, for example, why in deep inelastic scattering experiments electrons seem to interact with quasifree partons inside the proton.

To summarize, we can say that generically classical scale invariance is anomalous, in the sense that it disappears as the result of renormalization[30]. The $\beta$-function is just one example of a set of functions describing how couplings and masses change with the energy scale. Together, they build the coefficients of a set of first-order differential equations satisfied by the theory's correlation functions and other quantities and known as the *renormalization group equations*.

The cartoon description of renormalization presented above might lead to thinking that it is just a smart trick, somehow justifying Feynman's dictum that renormalization is sweeping the infinities under the rug [160]. We have come, however, a long way from there. The current understanding of renormalization, dating back to the groundbreaking work of Kenneth Wilson [161–163], goes much deeper and beyond the mere mathematics of shifting the cutoff dependence from one place to another. It is also closely related to the idea of EFTs, so now we can revisit our discussion on pages 3-7 in more precise terms.

Everything boils down to making a physical interpretation of the cutoff. Instead of seeing it as an artificial scale introduced to render integrals finite, we can regard it as the upper energy scale at which our theory is defined. At energies above $\Lambda$, new physics may pop up, but we do not really care too much, since all we need to know are the values of the masses $m_i(\Lambda)$ and dimensionless couplings $g_i(\Lambda)$.

Now we ask ourselves how the theory looks at some lower energy scale $\mu < \Lambda$. To answer, we need to "integrate out" all physical processes taking place in the range $\mu \leq E \leq \Lambda$, which results in a new field theory now defined at scale $\mu$ and expressed in terms of some "renormalized" fields. Generically, the masses and couplings of this theory will differ from the original ones, so we have $m_i(\mu) \neq m_i(\Lambda)$ and $g_i(\mu) \neq g_i(\Lambda)$. But, in addition to this, the new theory might also contain additional couplings not present at the scale $\Lambda$, in principle an infinite number of them. Using the language of path integrals, we symbolically summarize all this by writing

$$\int\limits_{\mu \leq E \leq \Lambda} \mathscr{D}\Phi_0 \, e^{iS_0[\Phi_0]} = e^{iS[\Phi]}, \tag{10.4}$$

where $\Phi_0$ collectively denotes the fields of the original theory and $\Phi$ their renormalized counterparts, while $S[\Phi]$ is the action of the new theory defined at the energy scale $\mu$. On general grounds, it can be written as

$$S[\Phi] = S_0[\Phi] + \sum_n \frac{g'_n(\mu)}{\Lambda^{\dim \mathcal{O}_n - 4}} \int d^4x \, \mathcal{O}_n[\Phi]. \tag{10.5}$$

In this expression $S_0[\Phi]$ is the action of the original theory with all fields, masses, and couplings replaced by the corresponding renormalized quantities, and $\mathcal{O}_i[\Phi]$ are new operators with dimensions greater than or equal to four induced by the physics integrated out between the scales $\Lambda$ and $\mu$. Their couplings $g'_n(\mu)$ are dimensionless and we see that higher-dimensional operators are suppressed by inverse powers of the high energy scale $\Lambda$.

In this Wilsonian picture of renormalization the dependence of the coupling constants with the

[30] This happens, for example, in QCD with massless quarks. There are however a few examples of theories for which this does not happen, most notably $\mathcal{N} = 4$ supersymmetric Yang Mills theory in four dimensions. Due to its large symmetry, classical conformal invariance is preserved by quantization.

scale has a clear physical meaning: as we go to lower energies, their changing values incorporate the physics that we are integrating out at intermediate scales. But not only this, also the difference between renormalizable and nonrenormalizable theories gets blurred. All theories are defined at a given energy scale $\Lambda$. In order to describe the physics above this scale, the theory would have to be "completed" with additional degrees of freedom and/or interactions. What is special in renormalizable theories is that they are their own UV completion, in the sense that they can be extended to arbitrarily high energies without running into trouble, although technically this only makes sense for asymptotically free theories.

Nonrenormalizable theories need to be completed in the UV to make sense of them above $\Lambda$. Let us look at the example of Fermi's theory of weak interaction. It has a natural cutoff given by $\Lambda = m_W$, and if we try to go beyond this energy we run into trouble. For example, the theory violates unitarity at high energies. The theory, however, can be completed in the UV by the electroweak model studied in Section 9, which being renormalizable can in principle be extended to higher energies without inconsistencies.

Another case of nonrenormalizable theories encountered in section 5 is the chiral Lagrangian (see page 67). Again, the theory is endowed with a physical cutoff, in this case $\Lambda_{\rm QCD}$, above which the description in terms of pions is no longer valid. In fact, we can see the chiral Lagrangian as resulting from Wilsonian renormalization applied to QCD: by integrating out the physics of strongly coupled quarks and gluons we get a low energy action for the new fields (the pions) and their interactions. Since the resulting theory does not make sense above $\Lambda_{\rm QCD}$ there is no problem with the divergences appearing in loops. After all, before the momenta running in them can reach infinity the pion as such ceases to exist.

The final instance of a nonrenormalizable theory we discuss is gravity, which, as explained in section 1, has to be completed above the Planck scale (1.7). But here we have to remember that everything couples to gravity, including the SM. Thus, we are led to conclude that despite being renormalizable, the SM itself has to be regarded as an effective description to be supplemented at the Planck scale, if not earlier. In fact, phenomena like the nonzero neutrino masses strongly indicate new physics lurking somewhere between the electroweak scale and the Planck scale.

The bottomline of our discussion is that nonrenormalizability is just a sign that we are dealing with an EFT and that the ubiquitous presence of gravity in nature forces us to regard *all* QFTs as EFTs (have a look again at Fig. 1 in page 7). Nonrenormalizable theories are not anymore those sinister objects they were when renormalization was seen as nothing but infinites removal. They are perfectly reasonable theories, provided we are aware of what they are and of what they are good for (and they are indeed *very* good for quite many things!).

**Box 17. The Planck chimney**

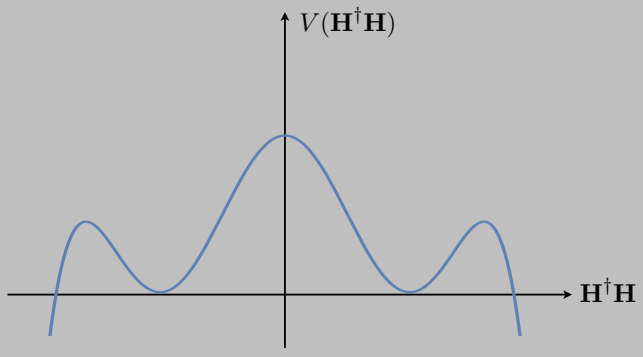
Let us go back to the Higgs action (9.31) and particularly to the potential

$$V(\mathbf{H},\mathbf{H}^\dagger) = \frac{\lambda}{4}\left(\mathbf{H}^\dagger\mathbf{H} - \frac{v^2}{2}\right)^2 . \tag{10.6}$$

We have seen that after symmetry breaking the parameter $\lambda$ directly relates to the Higgs mass (9.56) and determines its self couplings in the action (9.55). Since after quantization masses and couplings

get a dependence on the energy scale, we would like to know how $\lambda(\mu)$ or the Higgs mass $m_H(\mu)$ depend on the scale $\mu$. At this point we should recall that the strength of the coupling of the Higgs boson to fermions is proportional to the latter's masses [see Eq. (9.60)], so its interactions with the matter fields are dominated by the top quark. Thus the renormalization group equations determining the evolution of $\lambda(\mu)$ and $\mu_H(\mu)$ with the energy scale should also involve the top quark mass $m_t(\mu)$.

An important question is whether the evolution of these parameters with the scale changes in a significant way the shape of the Mexican hat potential and, most importantly, whether this jeopardizes the existence of a stable Higgs vacuum (see [164] and references therein). It might be that the sombrero's brim get flattened at higher energies, or even inverted like in the case shown here:




If this happens, the Higgs vacuum becomes metastable or outright unstable.

Since the renormalization group equations are first order, we need to specify some "initial conditions". In this case they are the values of the Higgs and top masses measured at the LHC. Assuming that the SM correctly describes the physics all the way to $\Lambda_{\rm Pl}$, the bounds to be satisfied by the masses in order to preserve the stability of the Higgs vacuum are [165–167]

$$\begin{aligned} m_H &> (129.1 \pm 1.5)\ \text{GeV}, \\ m_t &< (171.53 \pm 0.42)\ \text{GeV}. \end{aligned} \tag{10.7}$$

Comparing with the experimental values $m_H = (125.25 \pm 0.17)\,\text{GeV}$ and $m_t = (172.69 \pm 0.30)\,\text{GeV}$ [117], we see that the SM lies slightly outside the stability zone. In fact, the SM seems to be metastable, with the Higgs boson trapped in a false vacuum. The energy scale where the instability appears turns out to be of the order of the geometric mean of the $W$ mass and the Planck scale $\Lambda_{\rm inst} \sim \sqrt{m_W \Lambda_{\rm Pl}}$. This is quite a discovery made at the LHC!

The instability of the Higgs vacuum is indeed no good news. Of course, living in a metastable universe is no major problem if its tunneling probability is so low that its decay time turns out to be much larger than the age of the universe, around 13.6 Gyr. But we have to remember that the bounds (10.7) are obtained with the proviso that there are no new degrees of freedom between the electroweak and the Planck scales. This is yet another reason to expect some physics beyond the SM making the universe stable.

The apparent metastability of the Higgs vacuum highlights a very important feature of the renormalization group. We can run it from high to low energies with total confidence. Knowing the degrees of freedom and interactions at a certain scale $\Lambda$, everything is determined at energies $\mu < \Lambda$. The worst thing that may happen is that the degrees of freedom get "rearranged", as it happens in QCD where mesons and baryons replace quarks and gluons at low energies. But if the aim is getting information about what is going on at $\mu > \Lambda$, additional assumptions are required: either that no new degrees of freedom emerge above $\Lambda$, or that there is some UV completion whose details are necessarily an educated guess. After all, this is why particle physics is hard. Whatever happens above the energies we explore is blurred in the parameters of the theory we test. The best we can do is to play the model building game to reproduce this blurriness, and hopefully predict distinct signals that could be detected in some future facility.

## 11 Closing remarks

The SM is a vast and complex subject, providing the best description of particle physics and its applications at energies below a few TeV. It explains a large amount of phenomena in microphysics and in cosmology. However, its precise formulation delineates some of its limitation, as illustrated by the following list:

- The SM does not predict the values for the masses and mixing angles of quarks and leptons (including neutrino masses).
- The SM does not provide adequate candidates to explain dark matter.
- The only real progress in the study of dark energy has been to change its name from the previous one: the cosmological constant.
- We know that CP needs to be violated in the universe in order to generate a matter–antimatter asymmetry. Thus, three families are the minimum needed to generate a CP violating angle, apart from the QCD vacuum angle. Unfortunately, CP violation from the CKM matrix is not enough to generate the observed asymmetry. The equivalent angle in the neutrino sector has not yet been measured. It would be ironical if the ultimate origin of "humans" was related to properties of the ghostly neutrinos. Theories beyond the Standard Model provide many scenarios with larger amounts of CP violation.
- The currently preferred paradigm in cosmology is inflation. We still do not have a convincing candidate for what the inflaton is, or how the big bang was triggered, if that question makes any sense at all. There are still many open questions in cosmology, including what is the correct paradigm.

This is just a sample of the most pressing issues for which the SM cannot provide a satisfactory answer. For decades now the scientific community has been trying to address these problems through extensions of the SM, from minimal ones inspired by supersymmetry to radical proposals rethinking the very structure of the elementary constituents, like string theory.

So far the experiments have not given any positive indication as to where the answers to the open questions might lie. Despite transient anomalies or data bumps, the more we probe the Higgs particle

the more it looks like its "vanilla version". It is truly fascinating that, in order to give masses to the SM particles, nature has chosen the simplest solution we came up with, the Higgs field. The SM's definite triumph, the discovery of the Higgs particle in 2012, was also a disappointment, because it apparently closed the door to more exciting possibilities with a clear bearing on new physics.

One of the reasons for the impasse might be that we are at the end of a cycle and the current conceptual framework based on symmetry and locality has been exhausted, or maybe the idea of naturalness, a basic guiding principle in our understanding of particle physics, is after all a red herring. We still need to bring gravity into the SM and this opens a plethora of problems and questions, some of them touching notions like landscapes or multiverses loaded with philosophical or just metascientific ideas.

Cosmology and astroparticle physics might offer some hope. In recent years, we have witnessed important discoveries, from the first direct detection of gravitational waves in 2015 [168] to the "photo" of the black hole at the center of the M87 galaxy [169] in 2019. The rapidly developing field of gravitational wave astronomy opens up new windows to phenomena up to now out of observational reach, and it may allow unprecedented glimpses into the physics of compact astrophysical objects or the very early universe.

We should not give up hope. Maybe we are on the verge of a golden era of discoveries that will leave us gasping with awe and laughing with joy in amazement of a new vision of the universe. One never knows, and dreaming is for free.

## Acknowledgments

These lecture notes contain an extended version of courses taught by the authors at the 2022 European School for High Energy Physics (L.A.-G.), the TAE 2017 and 2019 schools, and graduate courses at Madrid Autónoma University (M.A.V.-M.). L.A.-G. would like to thank Markus Elsing, Martijn Mulders, Gilad Perez, and Kate Ross for their invitation to present the lectures at the 2022 ESHEP Jerusalem school, and fun moments together. We would also like to thank Het Joshi, student assistant at the Simons Center for Geometry and Physics, for her excellent work editing the first draft of these lecture notes. M.A.V.-M. acknowledges financial support from the Spanish Science Ministry through research grant PID2021-123703NB-C22 (MCIN/AEI/FEDER, EU), as well as from Basque Government grant IT1628-22.

## References

[1] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields*, McGraw-Hill 1965.

[2] C. Itzykson and J. B. Zuber, *Quantum Field Theory*, McGraw-Hill 1980.

[3] P. Ramond, *Field Theory: A Modern Primer*, Addison-Wesley 1990.

[4] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Addison-Wesley 1995.

[5] S. Weinberg, *The Quantum Theory of Fields*, vol. 1, vol. 2, and vol. 3, Cambridge 1995, 1996, and 2000.

[6] B. DeWitt, *The Global Approach to Quantum Field Theory*, Oxford 2003.

[7] M. Maggiore, *A Modern Introduction to Quantum Field Theory*, Oxford 2005.

[8] V. P. Nair, *Quantum Field Theory: A Modern Perspective*, Springer 2005.

[9] C. Burgess and G. Moore, *The Standard Model: A Primer*, Cambridge 2006.

[10] E. A. Paschos, *Electroweak Theory*, Cambridge 2007.

[11] W. N. Cottingham and D. A. Greenwood, *An Introduction to the Standard Model of Particle Physics (2nd edition)*, Cambridge 2007.

[12] T. Banks, *Modern Quantum Field Theory: A Concise Introduction*, Cambridge 2008.

[13] A. Zee, *Quantum Field Theory in a Nutshell (2nd edition)*, Princeton 2010.

[14] L. Álvarez-Gaumé and M. Á. Vázquez-Mozo, *An Invitation to Quantum Field Theory*, Springer 2012.

[15] M. D. Schwartz, *Quantum Field Theory and the Standard Model*, Cambridge 2013.

[16] G. Kane, *Modern Elementary Particle Physics: Explaining and Extending the Standard Model (2nd edition)*, Cambridge 2017.

[17] D. Goldberg, *The Standard Model in a Nutshell*, Princeton 2017.

[18] S. Raby, *Introduction to the Standard Model and Beyond: Quantum Field Theory, Symmetries and Phenomenology*, Cambridge 2021.

[19] G. Aad *et al.* [ATLAS], *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1 [arXiv:1207.7214 [hep-ex]].

[20] S. Chatrchyan *et al.* [CMS], *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30 [arXiv:1207.7235 [hep-ex]].

[21] A. A. Abrikosov, L. P. Gorkov and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics*, Prentice-Hall 1963, reprint edition Dover, 1975.

[22] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems*, McGraw-Hill 1971, reprint edition by Dover, 2003.

[23] H. Bruss and K. Flensberg, *Many-Body Quantum Theory in Condensed Matter Physics: An Introduction*, Oxford, 2004.

[24] C. P. Burgess, *Introduction to Effective Field Theories and Inflation*, in: "Effective Field Theory in Particle Physics and Cosmology", eds. S. Davidson *et al.*, pp. 220–306, Oxford 2020. [arXiv:1711.10592 [hep-th]]

[25] T. Baldauf, *Effective Field Theory of Large-Scale Structure*, in: "Effective Field Theory in Particle Physics and Cosmology", eds. S. Davidson *et al.*, pp. 415–478, Oxford 2020.

[26] G. Cabass *et al.*, *Snowmass white paper: Effective field theories in cosmology,* Phys. Dark Univ. **40** (2023) 101193 [arXiv:2203.08232 [astro-ph.CO]].

[27] A. Pich, *Effective Field Theory*, in: "Probing the Standard Model of Particle Interactions", eds. R. Gupta, A. Morel, E. de Rafael and F. David, pp. 949–1049, North Holland 1999 [arXiv:hep-ph/9806303 [hep-ph]]

[28] D. B. Kaplan, *Five lectures on effective field theory*, [arXiv:nucl-th/0510023 [nucl-th]]

[29] R. Feynman, *Feynman Lectures on Gravitation*, Addison-Wesley 1995, ebook CRC Press 2019.

[30] E. Álvarez, *Quantum gravity: an introduction to some recent results*, Rev. Mod. Phys. **61** (1989) 561.

[31] H. W. Hamber, *Quantum Gravitation: The Feynman Path Integral Approach*, Springer 2008.

[32] L. E. Ibáñez and Á. M. Uranga, *String Theory and Particle Physics*, Cambridge 2012.

[33] E. Kiritsis, *String Theory in a Nutshell (2nd edition)*, Princeton 2021.

[34] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.

[35] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132.

[36] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[37] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

[38] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[39] A. Salam, *Weak and Electromagnetic Interactions*, in: "Proceedings of the 8th Nobel Symposium", pp. 367–377, Almqvist & Wiksell 1968, reprinted in "Selected Papers of Abdus Salam", pp. 244–254, World Scientific 1994.

[40] G. 't Hooft, *Renormalizable Lagrangians for Massive Yang-Mills Fields*, Nucl. Phys. B **35** (1971) 167.

[41] G. 't Hooft and M. J. G. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl. Phys. B **44** (1972) 189.

[42] F. Klein, *Vergleichende Betrachtungen über neuere geometrische Forschungen*, Math. Ann. **43** (1893) 63. (For an English translation, see `arXiv:0807.3161`)

[43] M. Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford 1990.

[44] P. Curie, *Sur la symétrie dans les phénomenes physiques, symétrie d'un champ électrique et d'un champ magnétique*, J. de Phys. **3** (1894) 26.

[45] E. Noether, *Invariante Variationsprobleme*, Nachr. v. d. Kgl. Ges. d. Wiss. zu Göttingen, Math.-phys. Kl. (1918) 235. (For an English translation, see `arXiv:physics/0503066`)

[46] A. Pais, *"Subtle is the Lord...": The Science and Life of Albert Einstein*, Oxford 1982.

[47] B. R. Holstein, *Klein's paradox*, Am. J. Phys. 66 (1999) 507.

[48] E. C. G. Stueckelberg, *La Mecanique du point materiel en theorie de relativite et en theorie des quanta,* Helv. Phys. Acta **15** (1942) 23.

[49] R. P. Feynman, *A Relativistic Cutoff for Classical Electrodynamics,* Phys. Rev. **74** (1948) 939.

[50] R. P. Feynman, *The reason for antiparticles*, in: R. P. Feynman and S. Weinberg, *Elementary Particles and the Laws of Physics. The 1986 Dirac Memorial Lectures*, pp. 1–60, Cambridge 1987.

[51] J. D. Jackson, *Classical Electrodynamics (3rd edition)*, Wiley 1999.

[52] Y. Aharonov and D. Bohm, *Significance of electromagnetic potentials in the quantum theory,* Phys. Rev. **115** (1959) 485.

[53] P. A. M. Dirac, *Quantised singularities in the electromagnetic field,* Proc. Roy. Soc. Lond. A **133** (1931) 60.

[54] B. Cabrera, *First Results from a Superconductive Detector for Moving Magnetic Monopoles,* Phys. Rev. Lett. **48** (1982) 1378.

[55] P. B. Price, E. K. Shirk, W. Z. Osborne and L. S. Pinsky, *Evidence for Detection of a Moving Magnetic Monopole,* Phys. Rev. Lett. **35** (1975) 487.

[56] T. T. Wu and C. N. Yang, *Dirac's Monopole Without Strings: Classical Lagrangian Theory,* Phys. Rev. D **14** (1976), 437.

[57] J. A. Azcárraga and J. M. Izquierdo, *Lie groups, Lie algebras, cohomology and some applications in physics*, Cambridge 1995.

[58] M. Nakahara, *Geometry, Topology and Physics (2nd edition)*, CRC Press 2003.

[59] C. Nash and S. Sen, *Topology and Geometry for Physicists*, Dover 2011.

[60] T. Frankel, *The Geometry of Physics (3rd edition)*, Cambridge 2011.

[61] S. Weinberg, *The Cosmological Constant Problem,* Rev. Mod. Phys. **61** (1989) 1.

[62] T. Padmanabhan, *Cosmological Constant: The Weight of the Vacuum,* Phys. Rept. **380** (2003) 235. [arXiv:hep-th/0212290 [hep-th]].

[63] R. Bousso, *TASI Lectures on the Cosmological Constant,* Gen. Rel. Grav. **40** (2008) 607 [arXiv:0708.4231 [hep-th]].

[64] R. Haag, *On Quantum Field Theories*, Danske Vid. Selsk. Mat.-Fys. Medd. **29** (1955) 669.

[65] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That*, Princeton 1989.

[66] R. Haag, *Local Quantum Physics (2nd edition)*, Springer 1996.

[67] F. Strocchi, *An Introduction to Non-Perturabative Foundations of Quantum Field Theory*, Oxford 2013.

[68] H. Georgi, *Lie Algebras in Particle Physics: From Isospin to Unified Field Theories (2nd edition)*, Perseus Books 1999.

[69] P. Ramond, *Group Theory: A Physicist's Survey*, Cambridge 2010.

[70] M. Gell-Mann, *The Eightfold Way: A Theory of strong interaction symmetry,* Caltech report CTSL-20/TID-12608 (1961).

[71] Y. Ne'eman, *Derivation of strong interactions from a gauge invariance,* Nucl. Phys. **26** (1961) 222.

[72] M. Gell-Mann, *A Schematic Model of Baryons and Mesons,* Phys. Lett. **8** (1964) 214.

[73] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking (versions 1 & 2),* CERN reports CERN-TH-401 and CERN-TH-412 (1964).

[74] E. Wigner, *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren*, Vieweg+Teubner Verlag 1931. English translation: *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra*, Academic Press 1959.

[75] F. J. Belinfante, *On the Current and the Density of the Electric Charge, the Energy, the Linear Momentum and the Angular Momentum of Arbitrary Fields*, Physica **7** (1940) 29.

[76] L. Rosenfeld, *Sur le tenseur d'impulsion-energie*, Mem. Acad. Belgique cl. sc. **18** (1940) 1, reprinted in "Selected Papers of Léon Rosenfeld", Boston Stud. Phil. Sci. **21** (1979) 711—735, for an English translation, see http://neo-classical-physics.info.

[77] E. Merzbacher, *Quantum Mechanics (3rd edition)*, Wiley 1998.

[78] J. Goldstone, *Field Theories with 'Superconductor' Solutions,* Nuovo Cim. **19** (1961) 154.

[79] J. Goldstone, A. Salam and S. Weinberg, *Broken Symmetries*, Phys. Rev. **127** (1962) 965.

[80] Y. Nambu, *Dynamical theory of elementary particles suggested by superconductivity,* in: "Proceedings of the 10th International Conference on High-Energy Physics", pp. 858–864, 1960.

[81] Y. Nambu and G. Jona-Lasinio, *Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. 1.,* Phys. Rev. **122** (1961) 345.

[82] N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group*, Addison-Wesley 1992.

[83] J. F. Annett, *Superconductivity, Superfluids and Condensates*, Oxford 2004.

[84] A. M. J. Schakel, *Boulevar of Broken Symmetries: Effective Field Theories of Condensed Matter*, World Scientific 2008.

[85] A. Pich, *Chiral perturbation theory*, Rept. Prog. Phys. **58** (1995) 563.

[86] S. Scherer and M. R. Schinler, *A Primer for Chiral Perturbation Theory*, Springer 2012.

[87] P. W. Anderson, *Plasmons, Gauge Invariance, and Mass,* Phys. Rev. **130** (1963) 439.

[88] L. Álvarez-Gaumé and J. Ellis, *Eyes on a prize particle*, Nature Phys. **7** (2011) 2.

[89] H. Yukawa, *On the Interaction of Elementary Particles I,* Proc. Phys. Math. Soc. Jap. **17** (1935) 48.

[90] P. B. Littlewood and C. M. Varma, *Amplitude collective modes in superconductors and their coupling to charge-density waves*, Phys. Rev. B **26** (1982) 4883.

[91] R. Shimano and N. Tsuji, *Higgs Mode in Superconductors*, Ann. Rev. Condensed Matter Phys. **11** (2020) 103 [arXiv:1906.09401 [cond-mat.supr-con]].

[92] C. N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance,* Phys. Rev. **96** (1954) 191.

[93] L. Álvarez-Gaumé, *An Introduction to Anomalies*, in *"Fundamental Problems of Gauge Field Theory"* edited by G. Velo and A. S. Wightman, pp. 93—206, Plenum Press, 1985.

[94] R. A. Bertlmann, *Anomalies in Quantum Field Theory*, Oxford 1996.

[95] K. Fujikawa and H. Suzuki, *Path Integrals and Quantum Anomalies*, Oxford 2004.

[96] L. Alvarez-Gaumé and M. Á. Vázquez-Mozo, *Anomalies and the Green-Schwarz Mechanism,* in: "Handbook of Quantum Gravity", eds. C. Bambi, L. Modesto and I. L. Shapiro, Springer 2024, [arXiv:2211.06467 [hep-th]].

[97] L. D. Faddeev, *Operator Anomaly for the Gauss Law,* Phys. Lett. B **145** (1984) 81.

[98] P. Nelson and L. Álvarez-Gaumé, *Hamiltonian Interpretation of Anomalies,* Commun. Math. Phys. **99** (1985) 103.

[99] M. Kobayashi, K. Seo and A. Sugamoto, *Commutator Anomaly for the Gauss Law Constraint Operator,* Nucl. Phys. B **273** (1986) 607.

[100] R. Jackiw, *Field Theoretic Investigations in Current Algebra*, in "Current Algebras and Anomalies", eds. S. B. Treiman, R. Jackiw, B. Zumino and E. Witten, pp. 81–210, World Scientific 1985.

[101] J. Steinberger, *On the use of subtraction fields and the lifetimes of some types of meson decay,* Phys. Rev. **76** (1949) 1180.

[102] J. Schwinger, *On gauge invariance and vacuum polarization,* Phys. Rev. **82** (1951) 664.

[103] H. Fukuda and Y. Miyamoto, *On the $\gamma$-Decay of Neutral Mesons*, Prog. Theor. Phys. **4** (1949) 347.

[104] S. Ozaki, S. Oneda and S. Sasaki, *On the Decay of Heavy Mesons I*, Prog. Theor. Phys. **4** (1949) 524.

[105] S. L. Adler, *Axial vector vertex in spinor electrodynamics,* Phys. Rev. **177** (1969) 2426.

[106] J. S. Bell and R. Jackiw, *A PCAC puzzle: $\pi^0 \to \gamma\gamma$ in the $\sigma$ model,* Nuovo Cim. A **60** (1969) 47.

[107] I. Larin *et al.* [PrimEx-II Collaboration], *Precision measurement of the neutral pion lifetime,* Science **368** (2020) 506.

[108] D. G. Sutherland, *Current algebra and some nonstrong mesonic decays,* Nucl. Phys. B **2** (1967) 433.

[109] M. J. G. Veltman, *Theoretical aspects of high-energy neutrino interactions*, Proc. R. Soc. A **301** (1967) 107.

[110] K. Fujikawa, *Path Integral Measure for Gauge Invariant Fermion Theories,* Phys. Rev. Lett. **42** (1979) 1195.

[111] K. Fujikawa, *Path Integral for Gauge Theories with Fermions,* Phys. Rev. D **21** (1980) 2848.

[112] A. Achúcarro and P. K. Townsend, *A Chern-Simons Action for Three-Dimensional anti-De Sitter Supergravity Theories,* Phys. Lett. B **180** (1986) 89.

[113] E. Witten, *(2+1)-Dimensional Gravity as an Exactly Soluble System,* Nucl. Phys. B **311** (1988) 46.

[114] Z. F. Ezawa, *Quantum Hall Effects (3rd edition)*, World Scientific 2013.

[115] G. Villadoro, *Axions*, Lectures at the Galileo Galilei Institute, Arcetri 2015.

[116] A. Hook, *TASI Lectures on the Strong CP Problem and Axions*, PoS **TASI2018** (2019) 004 [arXiv:1812.02669 [hep-ph]].

[117] R. L. Workman *et al.* [Particle Data Group], *Review of Particle Physics,* PTEP **2022** (2022) 083C01.

[118] R. Alarcon *et al.*, *Electric dipole moments and the search for new physics,* [arXiv:2203.08103 [hep-ph]].

[119] V. Bernard, N. Kaiser and U. G. Meißner, *Chiral dynamics in nucleons and nuclei,* Int. J. Mod. Phys. E **4** (1995) 193.

[120] H. Georgi, *Weak Interactions and Modern Particle Physics*, Benjamin-Cummings 1984, Revised and updated edition published by Dover in 2009.

[121] S. Coleman, J. Wess and B. Zumino, *Structure of phenomenological Lagrangians. 1.*, Phys. Rev. **177** (1969) 2239.

[122] C. G. Callan, Jr., S. Coleman, J. Wess and B. Zumino, *Structure of phenomenological Lagrangians. 2.*, Phys. Rev. **177** (1969) 2247.

[123] J. Gasser, M. E. Sainio and A. Svarc, *Nucleons with Chiral Loops,* Nucl. Phys. B **307** (1988) 779.

[124] M. L. Goldberger and S. B. Treiman, *Decay of the pi meson,* Phys. Rev. **110** (1958) 1178.

[125] R. J. Crewther, P. Di Vecchia, G. Veneziano and E. Witten, *Chiral Estimate of the Electric Dipole Moment of the Neutron in Quantum Chromodynamics,* Phys. Lett. B **88** (1979) 123 [Erratum: Phys. Lett. B **91** (1980) 487].

[126] C. Vafa and E. Witten, *Parity Conservation in QCD*, Phys. Rev. Lett. **53** (1984) 535.

[127] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*, Phys. Rev. Lett. **38** (1977) 1440.

[128] R. D. Peccei and H. R. Quinn, *Constraints Imposed by CP Conservation in the Presence of Instantons,* Phys. Rev. D **16** (1977) 1791.

[129] S. Weinberg, *A New Light Boson?*, Phys. Rev. Lett. **40** (1978) 223.

[130] F. Wilczek, *Problem of Strong $P$ and $T$ Invariance in the Presence of Instantons*, Phys. Rev. Lett. **40** (1978) 279.

[131] J. Redondo and A. Ringwald, *Light shining through walls*, Contemp. Phys. **52** (2011) 211 [arXiv:1011.3741 [hep-ph]]

[132] G. Sigl, *Astroparticle Physics: Theory and Phenomenology*, Springer 2017.

[133] D. J. E. Marsh, *Axion Cosmology*, Phys. Rept. **643** (2016) 1 [arXiv:1510.07633 [astro-ph.CO]].

[134] C. O'Hare, *cajohare/Axionlimits: AxionLimits*, Zenodo 2020.

[135] T. D. Lee and C. N. Yang, *Question of Parity Conservation in Weak Interactions*, Phys. Rev. **104** (1956) 254.

[136] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, *Experimental Test of Parity Conservation in $\beta$ Decay,* Phys. Rev. **105** (1957) 1413.

[137] R. L. Garwin, L. M. Lederman and M. Weinrich, *Observations of the Failure of Conservation of Parity and Charge Conjugation in Meson Decays: The Magnetic Moment of the Free Muon*, Phys. Rev. **105** (1957) 1415.

[138] J. I. Friedman and V. L. Telegdi, *Nuclear Emulsion Evidence for Parity Nonconservation in the Decay Chain $\pi^+ \to \mu^+ \to e^+$,* Phys. Rev. **106** (1957) 1290.

[139] E. C. G. Sudarshan and R. E. Marshak, *Chirality Invariance and the Universal Fermi Interaction*, Phys. Rev. **109** (1958) 1860. (Originally published in "Proceedings of the Padua-Venice Conference on Mesons and Recently Discovered Particles", 1957.)

[140] R. P. Feynman and M. Gell-Mann, *Theory of the Fermi Interaction*, Phys. Rev. **109** (1958) 193.

[141] F. J. Hasert *et al.* [Gargamelle Neutrino], *Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment*, Phys. Lett. B **46** (1973) 138.

[142] D. Haidt, *The Discovery of Weak Neutral Currents*, in "60 Years of CERN Experiments and Discoveries", eds. H. Schopper and L. Di Lella, World Scientific 2015.

[143] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

[144] S. L. Glashow, J. Iliopoulos and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2** (1970) 1285.

[145] Z. Fodor and C. Hoelbling, *Light Hadron Masses from Lattice QCD*, Rev. Mod. Phys. **84** (2012) 449 [arXiv:1203.4789 [hep-lat]].

[146] T. Hatsuda, *Lattice Quantum Chromodynamics*, in: "An Advanced Course in Computational Nuclear Physics", eds. M. Hjorth-Jensen, M. P. Lombardo and U. van Kolck, Springer 2017.

[147] J. Lesgourges, G. Mangano, G. Miele and S. Pastor, *Neutrino Cosmology*, Cambridge 2013.

[148] C. Giunti and C. W. Kim, *Fundamental of Neutrino Physics and Astrophysics*, Oxford 2007.

[149] S. Bilenky, *Introduction to the Physics of Massive and Mixed Neutrinos (2nd edition)*, Springer 2018.

[150] S. M. Bilenky and C. Giunti, *Neutrinoless Double-Beta Decay: a Probe of Physics Beyond the Standard Model*, Int. J. Mod. Phys. A **30** (2015) 1530001 [arXiv:1411.4791 [hep-ph]].

[151] B. J. P. Jones, *The Physics of Neutrinoless Double Beta Decay: A Beginners Guide*, PoS **TASI2020** (2021) 007 [arXiv:2108.09364 [nucl-ex]].

[152] M. C. González-García and M. Yokoyama, *Neutrino Masses, Mixing, and Oscillations*, in [117].

[153] M. S. Sozzi, *Discrete Symmetries and CP Violation: From Experiment to Theory*, Oxford 2008.

[154] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.

[155] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531.

[156] J. M. Cline, *TASI Lectures on Early Universe Cosmology: Inflation, Baryogenesis and Dark Matter*, PoS **TASI2018** (2019) 001 [arXiv:1807.08749 [hep-ph]].

[157] B. Pontecorvo, *Neutrino Experiments and the Problem of Conservation of Leptonic Charge*, Zh. Eksp. Teor. Fiz. **53** (1967) 1717.

[158] Z. Maki, M. Nakagawa and S. Sakata, *Remarks on the unified model of elementary particles*, Prog. Theor. Phys. **28** (1962) 870.

[159] I. Esteban, M. C. González-García, M. Maltoni, T. Schwetz and A. Zhou, *The fate of hints: updated global analysis of three-flavor neutrino oscillations*, JHEP **09** (2020) 178 [arXiv:2007.14792 [hep-ph]].

[160] *Dr. Richard Feynman Nobel Laureate!*, California Tech (October 22, 1965).

[161] K. G. Wilson, *The Renormalization Group: Critical Phenomena and the Kondo Problem*, Rev. Mod. Phys. **47** (1975) 773.

[162] K. G. Wilson, *The renormalization group and critical phenomena*, Rev. Mod. Phys. **55** (1983) 583.

[163] J. Zinn-Justin, *Phase Transitions and Renormalization Group*, Oxford 2013.

[164] M. Sher, *Electroweak Higgs Potentials and Vacuum Stability*, Phys. Rept. **179** (1989) 273.

[165] G. Degrassi, S. Di Vita, J. Elias-Miró, J. R. Espinosa, G. F. Giudice, G. Isidori and A. Strumia, *Higgs mass and vacuum stability in the Standard Model at NNLO*, JHEP **08** (2012) 098 [arXiv:1205.6497 [hep-ph]].

[166] S. Alekhin, A. Djouadi and S. Moch, *The top quark and Higgs boson masses and the stability of the electroweak vacuum*, Phys. Lett. B **716** (2012) 214 [arXiv:1207.0980 [hep-ph]].

[167] D. Buttazzo, G. Degrassi, P. P. Giardino, G. F. Giudice, F. Sala, A. Salvio and A. Strumia, *Investigating the near-criticality of the Higgs boson*, JHEP **12** (2013) 089 [arXiv:1307.3536 [hep-ph]].

[168] B. P. Abbott *et al.* [LIGO Scientific and Virgo], *Observation of Gravitational Waves from a Binary Black Hole Merger*, Phys. Rev. Lett. **116** (2016) 061102 [arXiv:1602.03837 [gr-qc]].

[169] K. Akiyama *et al.* [Event Horizon Telescope], *First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole*, Astrophys. J. Lett. **875** (2019) L1 [arXiv:1906.11238 [astro-ph.GA]].