Proceedings of the 2022 European School of High-Energy Physics

Jerusalem, Israel, 30 November–13 December 2022

Editors: Markus Elsing, Alexander Huss

CERN Yellow Reports: School Proceedings
Published by CERN, CH-1211 Geneva 23, Switzerland

ISBN 978-92-9083-675-9 (paperback)
ISBN 978-92-9083-676-6 (PDF)
ISSN 2519-8041 (Print)
ISSN 2519-805X (Online)
DOI https://doi.org/10.23730/CYRSP-2025-001

This volume should be cited as:

Proceedings of the 2022 European School of High-Energy Physics,
CERN Yellow Reports: School Proceedings, CERN-2025-001 (CERN, Geneva, 2024)
https://doi.org/10.23730/CYRSP-2025-001.

A contribution in this report should be cited as:

[Author name(s)], in: Proceedings of the 2022 European School of High-Energy Physics,
CERN-2025-001 (CERN, Geneva, 2024), pp. [first page]–[last page],
https://doi.org/10.23730/CYRSP-2025-001.[first page]

Corresponding editor: Markus.Elsing@cern.ch.

Accepted in March 2025, by the CERN Reports Editorial Board (contact Carlos.Lourenco@cern.ch).

Published by the CERN Scientific Information Service (contact Jens.Vigen@cern.ch).

Indexed in the CERN Document Server and in INSPIRE.

Published Open Access to permit its wide dissemination, as knowledge transfer is an integral part of the mission of CERN.

Proceedings of the 2022 European School of High-Energy Physics

Editors: Markus Elsing, Alexander Huss

Abstract

The 2022 European School of High-Energy Physics (ESHEP) took place in Israel. ESHEP is intended to give young physicists an introduction to recent theoretical and experimental advances in elementary particle physics. These proceedings contain lecture notes on field theory and the electro-weak Standard Model, on neutrino physics and on flavour physics and CP violation.

Keywords

Field theory, Standard Model, Neutrino physics, Flavour physics, Lecture notes

Preface

Martijn Mulders[a]
[a]CERN

The twenty-eighth event in the series of the European School of High-Energy Physics took place in Ma'ale Hachamisha (near Jerusalem), Israel, from 30 November to 13 December 2022. It was organized by CERN, with support from the Weizmann Institute, Tel Aviv University, Ben Gurion University, Technion, the Hebrew University of Jerusalem and the Azrieli Foundation. The local organization team was chaired by Gilad Perez (Weizmann Institute).

A total of 71 students of 21 different nationalities attended the school, mainly from institutes in member states of CERN, but also some from other regions. The participants were generally students in experimental High-Energy Physics in the final years of work towards their PhDs.

The School was hosted at the Yearim Hotel, in the Jerusalem hills. According to the tradition of the School, the students shared twin rooms mixing participants of different nationalities. A total of 31 lectures were complemented by daily discussion sessions led by five discussion leaders. The students displayed their own research work in the form of posters in an evening session in the first week, and the posters stayed on display until the end of the School. The full scientific programme was arranged in the on-site conference facilities.

The School also included an element of outreach training, complementing the main scientific programme. This consisted of a two-part course from the Inside Edge media training company. Additionally, students had the opportunity to act out radio interviews under realistic conditions based on a hypothetical scenario. The students from each discussion group subsequently carried out a collaborative project, preparing a talk on a physics-related topic at a level appropriate for a general audience. The talks were given by student representatives of each group in an evening session in the second week of the School. A jury, chaired by Dana Bernstein (Weizmann), judged the presentations; other members of the jury were Yossi Nir (Weizmann), and Guy Wilkinson (University of Oxford). We are very grateful to all of these people for their help.

Our thanks go to the local-organization team for all of their work and assistance in preparing the School, on both scientific and practical matters, and for their presence throughout the event. Our thanks also go to the efficient and friendly hotel management and staff who assisted the School organizers and the participants in many ways.

Very great thanks are due to the lecturers and discussion leaders for their active participation in the School and for making the scientific programme so stimulating. The students, who in turn manifested their good spirits during two intense weeks, appreciated listening to and discussing with the teaching staff of world renown.

In addition to the rich academic programme, the participants enjoyed leisure and cultural activities in Israel. There was a half-day excursion to Jerusalem, including a fascinating guided tour around the historical sites of the city. A full-day excursion to Masada, an impressive archaeological site in the Judean Desert, was followed by a visit to the Dead Sea. On the final Saturday afternoon, the students were able to make use of the hotel facilities during free time or visit the city of Jerusalem independently. The excursions provided an excellent environment for informal interactions between staff and students. We are very grateful to the School Administrator, Kate Ross (CERN) and Administrator for the LOC, Adi Zehavi (Weizmann), for their untiring efforts in the lengthy preparations for and the day-to-day operation of the School. Their continuous care of the participants and their needs during the School was highly appreciated.

The success of the School was to a large extent due to the students themselves. Their poster session was very well prepared and highly appreciated, their group projects were a big success, and throughout the School they participated actively during the lectures, in the discussion sessions and in the different activities and excursions.

Martijn Mulders
(On behalf of the Organizing Committee)


The 2022
European School
of High-Energy
Physics
Israel, 17 - 30 June 2020


THE LOWEST BAR IN THE WORLD
DEADSEA.com | -420 | KALIA BEACH
e.g. SUSY searches, or WIMPs
Look for unexpected effects in suppressed processes, or processes that are forbidden e.g. charged lepton flavour violation
BREAKING
NEWS
NEW PHYSICS? W BOSON HEAVIER THAN EXPECTED

Lecture summaries

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

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.

Neutrino physics

This is an update of the lectures previously published in arXiv:1708.01046. The topics discussed in this lecture include: general properties of neutrinos in the SM, the theory of neutrino masses and mixings (Dirac and Majorana), neutrino oscillations both in vacuum and in matter, as well as an overview of the experimental evidence for neutrino masses and of the prospects in neutrino oscillation physics. We also briefly comment on the relevance of neutrinos in leptogenesis and in beyond-the-Standard-Model physics.

Flavour physics

We explain the reasons for the interest in flavor physics. We describe flavor physics and the related CP violation within the Standard Model, with emphasis on the predictions of the model related to features such as flavor universality and flavor diagonality. We describe the flavor structure of flavor changing charged current interactions, and how they are used to extract the CKM parameters. We describe the structure of flavor changing neutral current interactions, and explain why they are highly suppressed in the Standard Model. We explain how the B-factories proved that the CKM (KM) mechanism dominates the flavor changing (CP violating) processes that have been observed in meson decays. We explain the implications of flavor physics for new physics, with emphasis on the "new physics flavor puzzle", and present the idea of minimal flavor violation as a possible solution. We explain the "Standard Model flavor puzzle", and present the Froggatt–Nielsen mechanism as a possible solution. We show that measurements of the Higgs boson decays may provide new opportunities for making progress on the various flavor puzzles. We briefly discuss two sets of measurements and some of their possible theoretical implications: $R(K^{(*)})$ and $R(D^{(*)})$.

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 (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 [energy] $= ML^2T^{-2}$ we see that both lengths and times have dimensions of (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, λ_4 is dimensionless while $[\lambda_6] = E^{-2}$.

Terms such as ϕ^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 Λ, 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 energies [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 λ_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 Λ, a cutoff, so fluctuations with wavelength smaller than Λ^{-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 Λ, 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 Λ, 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 (χPT), is a nonrenormalizable QFT whose action contains a plethora of higher-dimensional operators. Its cutoff, however, is not some arbitrary energy Λ 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

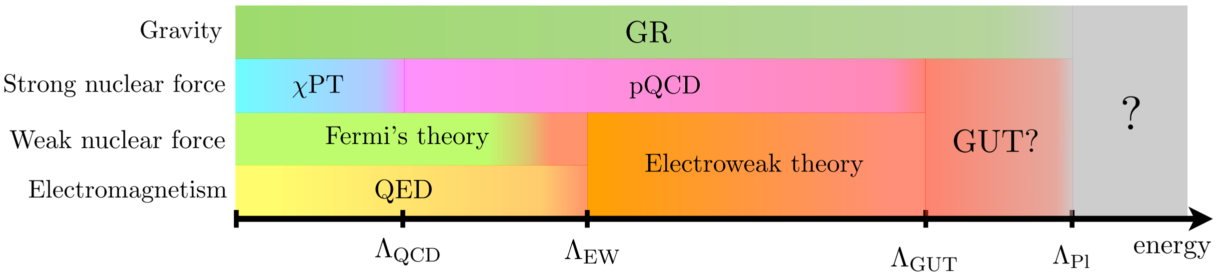


Fig. 1: Simplified cartoon showing the network of EFTs behind our understanding of subatomic physics.

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

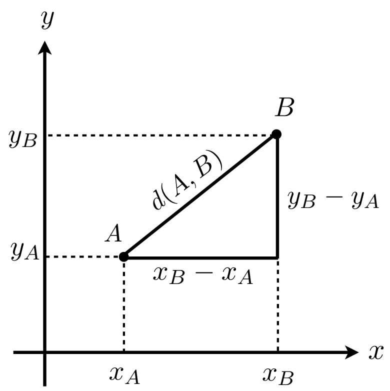


Fig. 2: Euclidean distance between two points on the plane.

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 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

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

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

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

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 Δt, 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^μ is an arbitrary constant vector. The linear coordinate change (2.9) generates the Poincaré group, ISO$(1,3)$, that includes all transformations Λ^μ_ν in the Lorentz group SO$(1,3)$ in addition to rigid translations. Notice that Λ^μ_ν is a 4×4 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

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

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

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

thus preserving the form $T'^{\mu_1\ldots\mu_k}_{\nu_1,\ldots,\nu_n}(x') = 0$ 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

$$\left(\begin{array}{c} t' \\ x' \end{array}\right) = \left(\begin{array}{cc} \Lambda^0{}_0 & \Lambda^0{}_1 \\ \Lambda^1{}_0 & \Lambda^1{}_1 \end{array}\right)\left(\begin{array}{c} t \\ x \end{array}\right). \tag{2.13}$$

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 α and β, with the third one requiring $\beta = \alpha$. The sought transformation is therefore parametrized as

$$\left(\begin{array}{c} t' \\ x' \end{array}\right) = \left(\begin{array}{cc} \cosh\alpha & -\sinh\alpha \\ -\sinh\alpha & \cosh\alpha \end{array}\right)\left(\begin{array}{c} t \\ x \end{array}\right), \tag{2.15}$$

where the parameter α 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 α 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^μ 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 Δs^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 Δs^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.

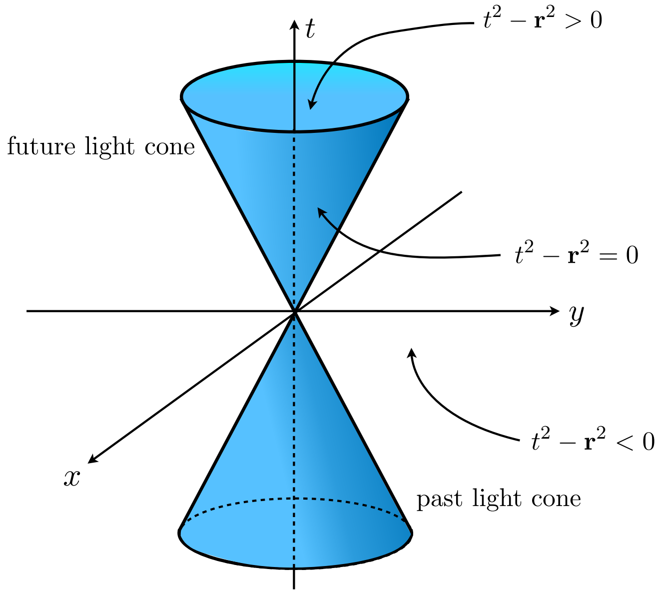
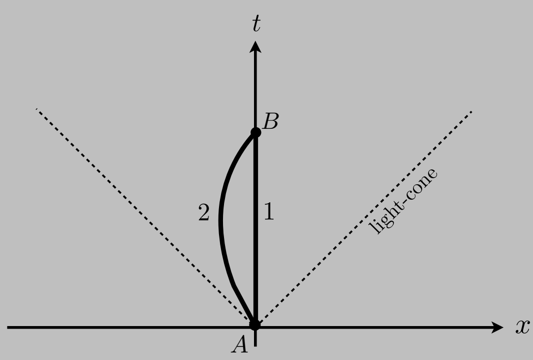


Fig. 3: Representation of the light cone at the origin in a $(2+1)$-dimensional spacetime.

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 γ_{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

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

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

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

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 $x_1^\mu(t') = \big(t', -\mathbf{R}(t')\big)$ and $x_2^\mu(t') =$

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

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

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

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

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^μ 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, $\eta_{\mu\nu}\dot{x}^\mu(s)\dot{x}^\nu(s) = 1$.

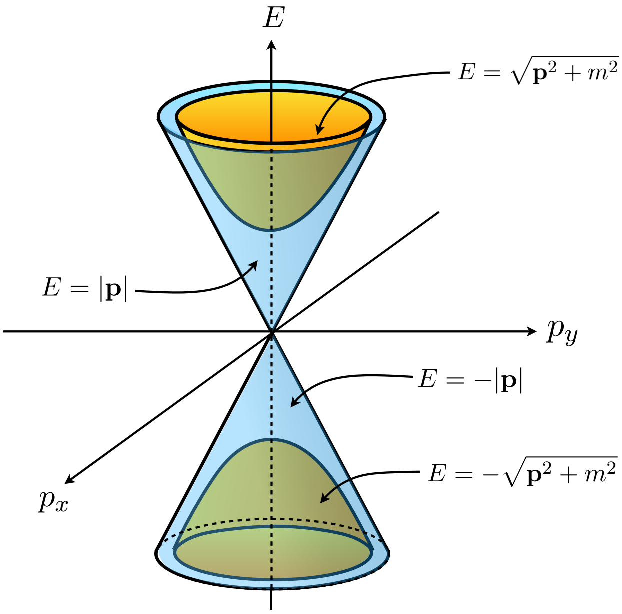


Fig. 4: Energy–momentum hyperboloid for a particle of mass $m \neq 0$ (orange). The energy-momentum vector of a massless particle lies on the blue cone.

In the four-dimensional energy–momentum space spanned by E and $\mathbf{p}$, the particle's four-momentum p^μ 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

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

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

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

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

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

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

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).

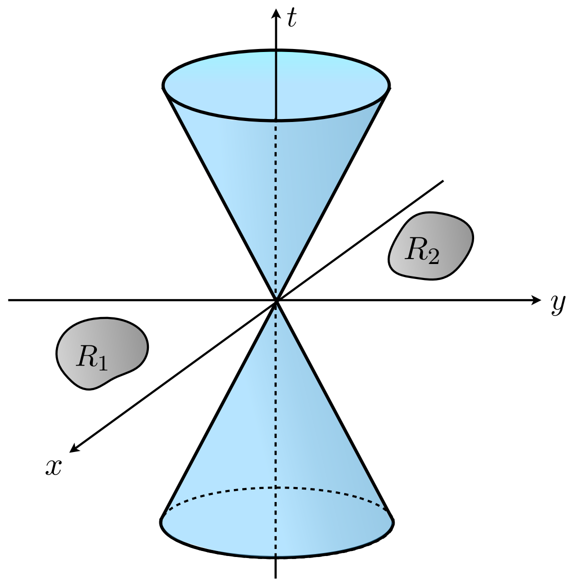


Fig. 5: The two spacelike-separated regions R_1 and R_2 cannot causally influence one another.

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 ρ_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

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

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

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

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

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

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 ϕ 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 ϕ 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 γ 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 γ 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 γ.

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

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

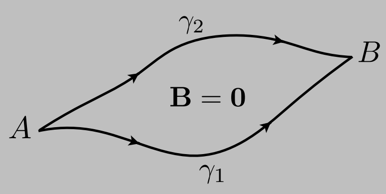
Box 4. Wilson loops and quantum interference

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 γ 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 γ across any region where the magnetic field vanishes. Let us consider two paths γ_1 and γ_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 γ_1 and back to A along γ_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

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

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 τ.

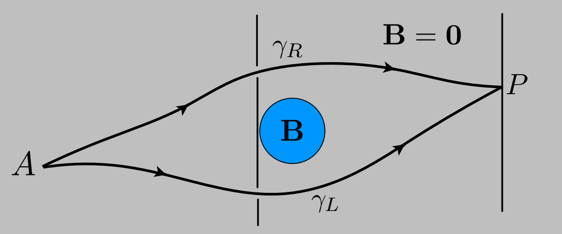
Recall that the magnetic field outside the solenoid is equal to zero and we can thus apply the topological property (3.14) to conclude that the second term in the exponential of (3.16) takes the same value for *all* trajectories γ_L passing through the left slit, and the same for *all* paths γ_R going through the right one. The total propagator can then be written as

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

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 Φ 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

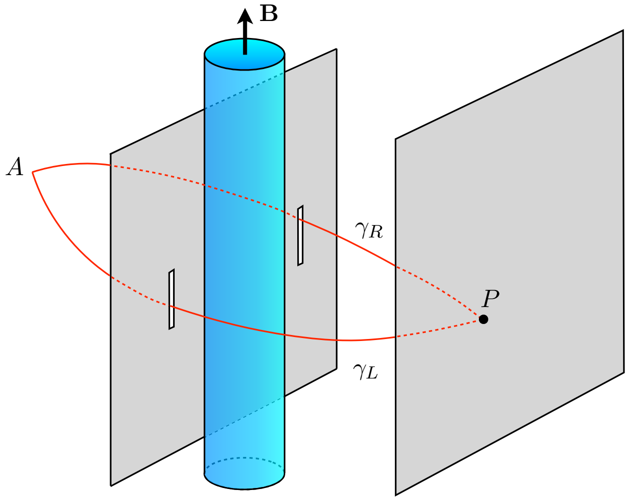


Fig. 6: Experimental setup to exhibit the Aharonov–Bohm effect explained in Box 4.

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 θ 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 ρ_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 ρ_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, φ, θ). 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

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

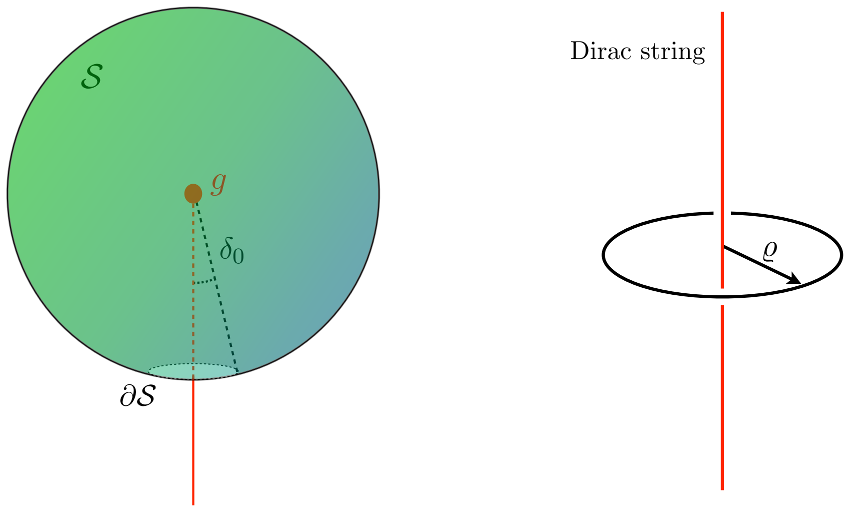


Fig. 7: Left: Section of a sphere around a Dirac magnetic monopole with charge g, resulting from cutting out a region around the south pole. Its boundary $\partial\mathcal{S}$ surrounds the singular Dirac string located along $\theta = \pi$ (in red). Right: Closed path surrounding the Dirac string.

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 ϱ 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

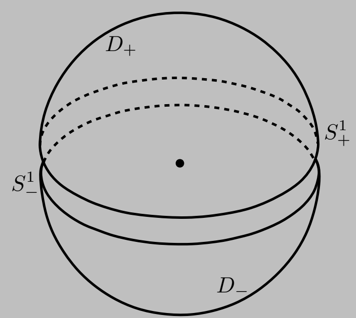
$$eg = 2\pi n \qquad \Longrightarrow \qquad e = \frac{2\pi}{g}n. \tag{3.34}$$

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)

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

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 2π.

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

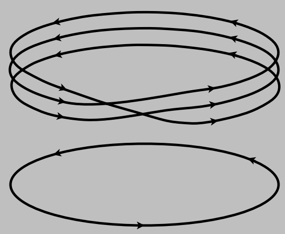
$$
e^{-ie\epsilon(0)} = e^{-ie\epsilon(2\pi)} \qquad \Longrightarrow \qquad e^{ieg} = 1, \tag{3.39}
$$

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 φ changes from 0 to 2π, 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

$$F_{\mu\nu} = \left(\begin{array}{cccc} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{array}\right), \tag{3.43}$$

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_μ, 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

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

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

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

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

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

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

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

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^μ is null

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

The parameter λ 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

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

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

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

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

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

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 λ. 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

$$\widehat{H} = \sum_{\lambda=\pm1} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2|\mathbf{k}|} |\mathbf{k}| \widehat{a}(\mathbf{k}, \lambda)^\dagger \widehat{a}(\mathbf{k}, \lambda) + \frac{1}{2} \sum_{\lambda=\pm1} \int d^3k\, |\mathbf{k}| \delta^{(3)}(\mathbf{0}). \tag{3.73}$$

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

$$\widehat{H}|0\rangle = \left(\frac{1}{2} \sum_{\lambda=\pm1} d^3k\, |\mathbf{k}| \delta^{(3)}(\mathbf{0}) \right) |0\rangle \tag{3.74}$$

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

$$\rho_{\text{vac}} \equiv \frac{E_{\text{vac}}}{V} = \frac{1}{2} \sum_{\lambda=\pm1} \int \frac{d^3k}{(2\pi)^3} |\mathbf{k}|. \tag{3.75}$$

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 ρ_{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 Λ_c contribution to the total vacuum energy of the universe as

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

where the sum is over all quantum fields in nature. Identifying the UV cutoff with the Planck energy, $\Lambda_{\text{UV}} \simeq \Lambda_{\text{Pl}}$, 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

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

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 λ 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

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

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

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

which admits plane wave solutions of the form

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

with p_μ 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 ϑ 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

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

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 λ 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 Δ_Φ is called the field's scaling dimension. Applying these transformations to the particular case of the action (3.103), we find

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

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

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

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 φ, 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,

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

[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.

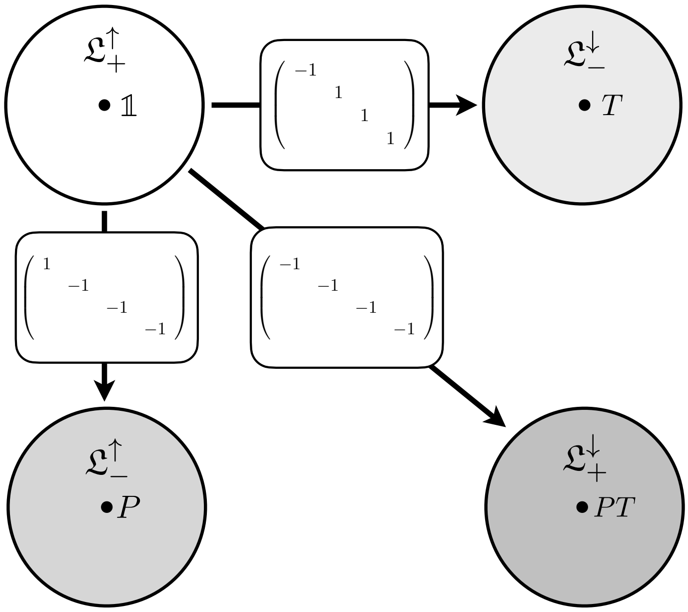


Fig. 8: The four connected components of the Lorentz group. The matrices indicate the transformations P, T, and PT mapping the connected component of the identity $\mathfrak{L}_+^\uparrow$ to the other three.

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 λ 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 χ_+ 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

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

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

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

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

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

Accordingly, we impose

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

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),

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

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

$$\psi \equiv \left(\begin{array}{c}\chi_+ \\ \chi_-\end{array}\right), \tag{4.43}$$

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, Λ_{SUSY}, rather than the much higher Λ_{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 ψ^* does not transform in the same representation as ψ, 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 4×4 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

$$\gamma_5 \equiv -i\gamma^0\gamma^1\gamma^2\gamma^3 = \left(\begin{array}{cc} \mathbb{1} & 0 \\ 0 & -\mathbb{1}\end{array}\right), \tag{4.53}$$

the previous identity is recast as

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

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

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

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)

$$j^\mu = \overline{\psi}\gamma^\mu\psi. \tag{4.57}$$

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

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

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

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

where the structure constants are given by

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

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 λ_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 ($i = 1, 2, 3$),

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

where S_{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_{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 π^0, $\pi^\pm$, K^0, $\overline{K}^0$, $K^\pm$, and η_8 mesons, while the singlet is the η_1 meson. In fact, the η_1 and η_8 mesons mix together into the η and the η' 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 Σ^0, $\Sigma^\pm$, Ξ^0, and Ξ^- 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 ϵ_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 ϵ_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^μ is linear in the ϵ_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 ϵ. 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 ϵ_A the current can be written as $j^\mu(\epsilon) = \epsilon_A j^\mu_A$, and (5.9) is satisfied for arbitrary values of ϵ_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_μ 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^μ (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^μ 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 ϵ_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^μ 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 Σ 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 ϵ_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 ϕ, θ, and ψ,

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

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 φ^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 ξ^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 ξ^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

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

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 φ^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 ξ. 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

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

where in the second line we split the quark fields into its right- and left-handed chiralities and in S_{int} we include all interaction terms. This theory is invariant under global 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 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 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 U(N_f) to U$(N_f)_R \times$ 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 α^a_R and α^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 U$(N_f)_R \times$ U$(N_f)_L$.

Taking into account that U(N_f) = U(1) × SU(N_f), the theory's global symmetry group can be written as

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

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_π 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 π^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 η-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_π. 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 $\pi^0 \to 2\gamma$

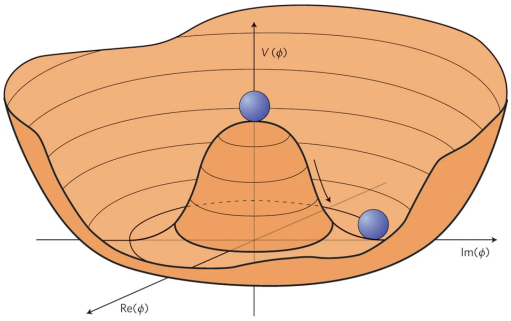


Fig. 9: Illustration from Ref. [88] depicting the celebrated Mexican hat potential shown in Eq. (5.61).

decay (see Section 7).

In our analysis of chiral symmetry breaking we encountered two energy scales: $\Lambda_{\chi\mathrm{SSB}}$ appearing in (5.51) as a consequence of the quark condensate having dimensions of (energy)^3, and f_π 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_π that determines the relative size of the infinite terms in the effective action (5.58). Operators weighted by f_π^{-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 electromagnetic 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

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

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 ϕ have been recast in terms of the field h and the NG boson ϑ. 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_μ, as befits a massive gauge field (see the footnote on page 32). We then arrive at the gauge-fixed action

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

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 ψ_R, ψ_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 ϕ 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 ϕ acquires the vev (5.62), and after gauging away the field ϑ, 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 $ev^2A^\mu\partial_\mu\vartheta$ 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^μ.

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)

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

Here $\phi(\mathbf{r})$ is the Cooper pair condensate, μ 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^μ 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,

$$S = \int d^4x\, \overline{\boldsymbol{\psi}}(i\partial\!\!\!/ - m)\boldsymbol{\psi}. \tag{6.2}$$

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 α^a.

To overcome this problem we have to find a covariant derivative D_μ, similarly to the one we introduced for Maxwell's theory, with the transformation

$$D_\mu\boldsymbol{\psi} \longrightarrow gD_\mu\boldsymbol{\psi}. \tag{6.4}$$

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 ∂_μ 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_μ 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 ∂_μ with D_μ defined in Eq. (6.5). In addition, we must include the dynamics of the new field A_μ 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_μ, 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,

$$
\begin{aligned}
S_{\text{YM}} &= \int d^4x \left[-\frac{1}{2g_{\text{YM}}^2} \text{tr}\left(F_{\mu\nu} F^{\mu\nu} \right) + \overline{\psi}(i\not{D} - m)\psi \right] \\
&= \int d^4x \left[-\frac{1}{4g_{\text{YM}}^2} F^a_{\mu\nu} F^{a\mu\nu} + \overline{\psi}(i\not{\partial} - m)\psi + A^a_\mu \overline{\psi}\gamma^\mu T^a_{\mathbf{R}}\psi \right],
\end{aligned}
\tag{6.13}
$$

where g_{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_{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

$$
H = \int d^3x \left[\frac{1}{2} \mathbf{E}^a \cdot \mathbf{E}^a + \frac{1}{2} \mathbf{B}^a \cdot \mathbf{B}^a + A^a_0 (\mathbf{D} \cdot \mathbf{E})^a \right].
\tag{6.16}
$$

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_{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_{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 ρ^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_∞, 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) $\times$ SU(2) $\times$ U(1), the color SU(3)$_\infty$ symmetry remains unbroken by the vacuum, whereas due to the BEH mechanism the electroweak factor [SU(2) $\times$ U(1)]$_\infty$ is partially realized à la Nambu–Goldstone, with a preserved 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) $\times$ 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 γ_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

$$j^\mu(x,\epsilon)_{\text{reg}} \equiv \overline{\psi}\left(x - \frac{1}{2}\epsilon\right)\gamma^\mu\psi\left(x + \frac{1}{2}\epsilon\right)\exp\left(ie\int_{x-\frac{1}{2}\epsilon}^{x+\frac{1}{2}\epsilon} dx^\mu A_\mu\right), \tag{7.4}$$

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

$$S_{\text{reg}} = \int d^4x \left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \overline{\psi}(i\not{D} - m)\psi + \sum_{i=1}^n c_k\overline{\Psi}_k(i\not{D} - M_k)\Psi_k\right], \tag{7.5}$$

with n and c_k chosen so that the limit

$$j^\mu(x)_{\text{reg}} \equiv \lim_{x'\to x}\left[\overline{\psi}(x')\gamma^\mu\psi(x) + \sum_{k=1}^n c_k\overline{\Psi}_k(x')\gamma^\mu\Psi_k(x)\right] \tag{7.6}$$

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 δt

$$G|\Psi(\delta t)\rangle = -i\delta t GH|\Psi(0)\rangle = -\delta t\partial_\mu j^\mu|\Psi(0)\rangle \neq 0, \tag{7.8}$$

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

$$[G(\mathbf{r}), G(\mathbf{r}')] = c\mathbf{B}(\mathbf{r}) \cdot \boldsymbol{\nabla}\delta^{(3)}(\mathbf{r} - \mathbf{r}'), \tag{7.9}$$

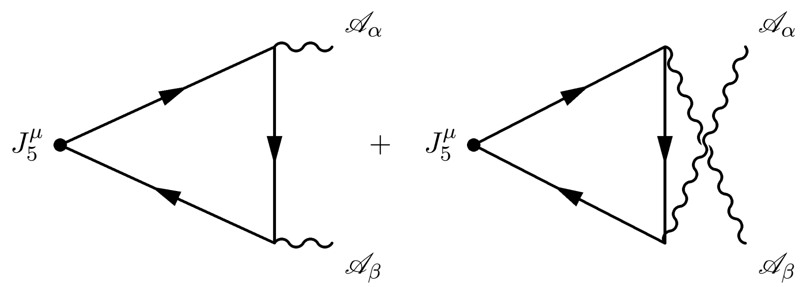
where $c \neq 0$ is a constant determined by the value of $\partial_\mu j^\mu$. This result implies that $G(\mathbf{r})|\text{phys}\rangle = 0$ cannot be consistently imposed, since this condition would imply $[G(\mathbf{r}), G(\mathbf{r}')]|\text{phys}\rangle = 0$ whereas the right-hand side of Eq. (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

$$\partial_\mu\langle J_5^\mu\rangle_{\mathscr{A},\mathcal{A}} = -\frac{N_c}{16\pi^2}\left(\sum_{f=1}^{N_f} q_f^2\right)\epsilon^{\mu\nu\alpha\beta}\mathscr{F}_{\mu\nu}\mathscr{F}_{\alpha\beta} - \frac{N_f}{16\pi^2}\epsilon^{\mu\nu\alpha\beta}\mathcal{F}_{\mu\nu}^a\mathcal{F}_{\alpha\beta}^a, \tag{7.13}$$

where $\mathcal{F}_{\mu\nu}^a$ 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 τ^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

$$\mathrm{SU}(N_f)_V : \boldsymbol{q} \to e^{i\alpha_V^I t_{\mathbf{f}}^I}\boldsymbol{q}, \qquad\qquad \mathrm{SU}(N_f)_A : \boldsymbol{q} \to e^{i\alpha_A^I t_{\mathbf{f}}^I\gamma_5}\boldsymbol{q}, \tag{7.15}$$

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

$$\partial_\mu \langle J_5^{I\mu}\rangle_{\mathscr{A},\mathcal{A}} = -\frac{N_c}{16\pi^2}\left[\sum_{f=1}^{N_f} q_f^2 (t_{\mathbf{f}}^I)_{ff}\right]\epsilon^{\mu\nu\alpha\beta}\mathscr{F}_{\mu\nu}\mathscr{F}_{\alpha\beta} - \frac{N_f}{16\pi^2}\big(\operatorname{tr} t_{\mathbf{f}}^I\big)\epsilon^{\mu\nu\alpha\beta}\mathcal{F}_{\mu\nu}^a\mathcal{F}_{\alpha\beta}^a. \tag{7.17}$$

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 σ^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_π is the pion mass and f_π 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].

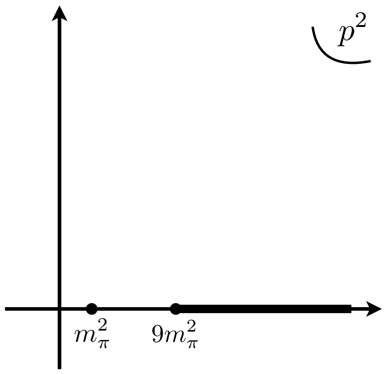


Fig. 10: Complex p^2-plane showing the structure of singularities of the function $f(p^2)$ in Eq. (7.24): a pole at $p^2 = m_\pi^2$ and a branch cut beginning at $p^2 = 9m_\pi^2$.

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

$$\Gamma(\pi^0 \to 2\gamma) = \frac{\alpha^2 N_c^2 m_\pi^3}{576\pi^3 f_\pi^2} = 7.73\,\mathrm{eV}, \tag{7.22}$$

which is perfectly consistent with experimental measurements [107]

$$\Gamma(\pi^0 \to 2\gamma)_{\mathrm{exp}} = 7.798 \pm 0.056\ (\mathrm{stat.})\ \pm 0.109\ (\mathrm{syst.})\ \mathrm{eV}. \tag{7.23}$$

Incidentally, the presence of $f_\pi = 93\,\mathrm{MeV}$ 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 π^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

$$\begin{aligned} S[\psi,\overline{\psi},\mathscr{A}_\mu] &= \int d^4x \overline{\psi}\gamma^\mu(i\partial_\mu + e\mathscr{A}_\mu)\psi \\ &= \int d^4x \Big[\overline{\psi}_R(i\partial_\mu + e\mathscr{A}_\mu)\psi_R + \overline{\psi}_L(i\partial_\mu + e\mathscr{A}_\mu)\psi_L\Big], \end{aligned} \tag{7.26}$$

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 α,

$$\mathscr{D}\overline{\psi}\mathscr{D}\psi = J[\mathscr{A}_\mu]\mathscr{D}\overline{\psi}'\mathscr{D}\psi'. \tag{7.30}$$

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,

$$S'[\psi',\overline{\psi}',\mathscr{A}_\mu] = \int d^4x\, \overline{\psi}'\gamma^\mu(i\partial_\mu + e\mathscr{A}_\mu)\psi' - i\log J[\mathscr{A}_\mu], \tag{7.32}$$

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:

$$\begin{aligned} S'[\psi',\overline{\psi}',\mathscr{A}_\mu] &= \int d^4x \Big[\overline{\psi}'_R\gamma^\mu(i\partial_\mu + e\mathscr{A}_\mu)\psi'_R + \overline{\psi}'_L\gamma^\mu(i\partial_\mu + e\mathscr{A}_\mu)\psi'_L \\ &\quad + me^{2i\alpha}(\overline{\psi}'_R\psi'_L + \overline{\psi}'_L\psi'_R)\Big] + \frac{e^2\alpha}{16\pi^2}\int d^4x\, \epsilon^{\mu\nu\alpha\beta}\mathscr{F}_{\mu\nu}\mathscr{F}_{\alpha\beta}. \end{aligned} \tag{7.34}$$

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 $A_0^a = 0$, 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 θ 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 θ 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 θ-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

$$\epsilon^{\mu\nu\alpha\beta} F_{\mu\nu}^a F_{\alpha\beta}^a = 4\frac{\partial}{\partial t}\left[\mathbf{A}^a \cdot \left(\boldsymbol{\nabla}\times\mathbf{A}^a\right) + \frac{1}{3}f^{abc}\mathbf{A}^a\cdot\left(\mathbf{A}^b\times\mathbf{A}^c\right)\right], \tag{8.16}$$

which, once integrated and with the proper normalization, gives the following expression of the θ-term

$$\begin{aligned} S_\theta = -\frac{\theta}{8\pi^2}\Bigg\{\int d^3r \left[\mathbf{A}^a \cdot \left(\boldsymbol{\nabla}\times\mathbf{A}^a\right) + \frac{1}{3}f^{abc}\mathbf{A}^a\cdot\left(\mathbf{A}^b\times\mathbf{A}^c\right)\right]\Bigg|_{t=\infty} \\ - \int d^3r \left[\mathbf{A}^a \cdot \left(\boldsymbol{\nabla}\times\mathbf{A}^a\right) + \frac{1}{3}f^{abc}\mathbf{A}^a\cdot\left(\mathbf{A}^b\times\mathbf{A}^c\right)\right]\Bigg|_{t=-\infty}\Bigg\}. \end{aligned} \tag{8.17}$$

To ensure finiteness, we take the gauge field $\mathbf{A} = \mathbf{A}^a T_{\mathbf{R}}^a$ 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 θ-term is nonzero (again, a derivation is outlined in Box 11),

$$\begin{aligned} S_\theta = \frac{\theta}{24\pi^2}\int d^3r\, \mathrm{tr}\left\{\left(g_+^{-1}\boldsymbol{\nabla} g_+\right)\cdot\left[\left(g_+^{-1}\boldsymbol{\nabla} g_+\right)\times\left(g_+^{-1}\boldsymbol{\nabla} g_+\right)\right]\right\} \\ - \frac{\theta}{24\pi^2}\int d^3r\, \mathrm{tr}\left\{\left(g_-^{-1}\boldsymbol{\nabla} g_-\right)\cdot\left[\left(g_-^{-1}\boldsymbol{\nabla} g_-\right)\times\left(g_-^{-1}\boldsymbol{\nabla} g_-\right)\right]\right\}. \end{aligned} \tag{8.18}$$

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 $A_\mu = A_\mu^a T_{\mathbf{R}}^a$ 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

$$F \equiv -\frac{i}{2}F_{\mu\nu}dx^\mu\wedge dx^\nu = dA + A\wedge A, \tag{8.22}$$

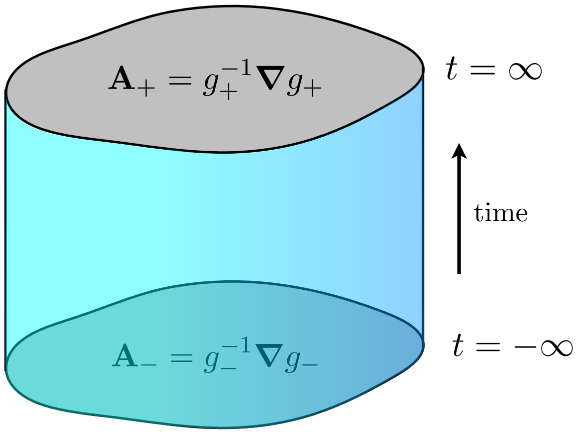


Fig. 11: Representation of the spacetime interpolating between two pure gauge configurations $\mathbf{A}_\pm = g_\pm \boldsymbol{\nabla} g_\pm$ at $t = \pm\infty$, in the $A_0 = 0$ gauge.

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

$$\Phi_p = -\frac{i}{p!}\Phi_{\mu_1\ldots\mu_p}dx^{\mu_1}\wedge\ldots\wedge dx^{\mu_p} \qquad \Longrightarrow \qquad \Phi_p \rightarrow \Phi'_p = g^{-1}\Phi_p g, \tag{8.25}$$

a covariant exterior derivative is defined acting as

$$D\Phi_p \equiv d\Phi_p + A\wedge\Phi_p - (-1)^p\Phi_p\wedge A \qquad \Longrightarrow \qquad (D\Phi_p)' = g^{-1}(D\Phi_p)g, \tag{8.26}$$

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

$$S_{\text{YM}} = \frac{1}{2g_{\text{YM}}^2}\int_{\mathcal{M}_4}\mathrm{tr}\,(F\wedge\star F) + \frac{\theta}{8\pi^2}\int_{\mathcal{M}_4}\mathrm{tr}\,(F\wedge F), \tag{8.31}$$

where $\mathcal{M}_4$ represents the four-dimensional spacetime. The two terms correspond respectively to the kinetic and θ 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 θ-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 θ-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 θ-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

$$J = \star\text{tr}\left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A\right). \tag{8.34}$$

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

$$\begin{aligned} \omega_3(A) \equiv \text{tr}\left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A\right) \\ \longrightarrow \quad \omega_3(A) - \frac{1}{3}\text{tr}\Big[(g^{-1}dg) \wedge (g^{-1}dg) \wedge (g^{-1}dg)\Big]. \end{aligned} \tag{8.35}$$

The Chern–Simons form is a very interesting object for many reasons. One is that it gives rise to the action

$$S_{\text{CS}} = -\frac{k}{4\pi}\int_{\mathcal{M}_3} \text{tr}\left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A\right), \tag{8.36}$$

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

$$S_{\text{CS}} \longrightarrow S_{\text{CS}} + \frac{k}{12\pi}\int_{\mathcal{M}_3} \text{tr}\Big[(g^{-1}dg) \wedge (g^{-1}dg) \wedge (g^{-1}dg)\Big], \tag{8.37}$$

the extra term equals $2\pi nk$, 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 $\exp(iS_{\text{CS}})$, 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]

$$\int_{\mathcal{M}_4} d^4x\, \partial_\mu \langle J_A^\mu(x)\rangle = -2i\big(N_+ - N_-\big), \tag{8.38}$$

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

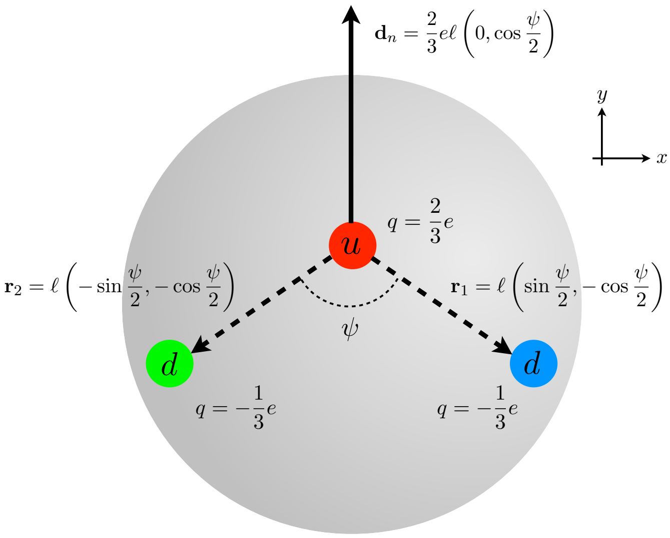


Fig. 12: Classical depiction of the neutron and its electric dipolar moment $\mathbf{d}_n$. The components of the d quarks position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are written using the coordinate axes shown in the picture, with origin on the position of the u quark.

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 θ-term (8.12) is that it violates both parity and CP, the combination of parity and charge conjugation,

$$\text{CP}: \left\{\begin{array}{lll} \mathbf{E}^a(t,\mathbf{r}) & \longrightarrow & \mathbf{E}^a(t,-\mathbf{r}) \\ \mathbf{B}^a(t,\mathbf{r}) & \longrightarrow & -\mathbf{B}^a(t,-\mathbf{r}) \end{array}\right. \qquad \Longrightarrow \qquad \text{CP}: S_\theta \longrightarrow -S_\theta. \tag{8.40}$$

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 ℓ of the u quark and their position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ span an angle ψ 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 ℓ 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 π, 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 θ 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 γ_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

$$S = \int d^4x \left(i\overline{\boldsymbol{q}}_R \not{D} \boldsymbol{q}_R + i\overline{\boldsymbol{q}}_L \not{D} \boldsymbol{q}_L + \overline{\boldsymbol{q}}_L M \boldsymbol{q}_R + \overline{\boldsymbol{q}}_R M^T \boldsymbol{q}_L - \frac{\theta}{32\pi^2}\epsilon^{\mu\nu\alpha\beta} F^a_{\mu\nu} F^a_{\alpha\beta} + \ldots \right), \tag{8.48}$$

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 θ-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 SU(2)$_R \times$ 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

$$S_{\rm NG} = \int d^4x \left[\frac{f_\pi^2}{4} \mathrm{tr}\left(D_\mu \boldsymbol{\Sigma}^\dagger D^\mu \boldsymbol{\Sigma} \right) + f_\pi^3 B_0 \mathrm{tr}\left(M^\dagger \boldsymbol{\Sigma} + \boldsymbol{\Sigma}^\dagger M \right) \right]. \tag{8.52}$$

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

$$\Delta S_{\rm NG} = -f_\pi B_0 (m_u + m_d) \int d^4x \Big[(\pi^0)^2 + 2\pi^+\pi^- \Big], \tag{8.53}$$

[24] The pion effective action $S_{\rm NG}$ 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 $\big(\mathrm{tr}\,\log\boldsymbol{\Sigma} - \mathrm{tr}\,\log\boldsymbol{\Sigma}^\dagger\big) F_{\mu\nu}\widetilde{F}^{\mu\nu}$, 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

$$N = \left(\begin{array}{c} p \\ n \end{array}\right), \tag{8.55}$$

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

$$\Delta S_{\pi N} = \int d^4x\, \overline{N}\Big[i\partial\!\!\!/ - f(\boldsymbol{\Sigma})\Big]N, \tag{8.57}$$

with $f(\boldsymbol{\Sigma})$ a matrix-valued function depending on the NG boson matrix and such that $D\!\!\!\!/ \equiv \partial\!\!\!/ + if(\boldsymbol{\Sigma})$ defines a covariant derivative with respect to the local transformation (8.56), $D\!\!\!\!/ \to KD\!\!\!\!/K^\dagger$. 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 $i\partial\!\!\!/N = m_N N$ 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 θ-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 θ-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

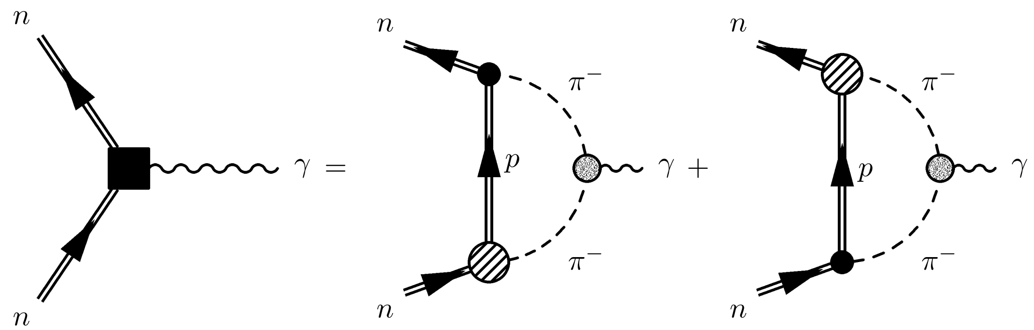
$$S = \int d^4x \left(i\overline{\boldsymbol{q}}_R \not{D} \boldsymbol{q}_R + i\overline{\boldsymbol{q}}_L \not{D} \boldsymbol{q}_L + \overline{\boldsymbol{q}}_L M \boldsymbol{q}_R + \overline{\boldsymbol{q}}_R M^\dagger \boldsymbol{q}_L - \frac{\theta - 2\alpha}{32\pi^2} \epsilon^{\mu\nu\alpha\beta} F^a_{\mu\nu} F^a_{\alpha\beta} + \ldots \right), \tag{8.61}$$

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 θ-term cancels and all dependence on θ 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 θ-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 $M = \mathrm{diag}(e^{i\theta} m_u, m_d)$ could be removed by an appropriate chiral transformation of the NG field $\mathbf{\Sigma}(x)$. In doing so, however, we introduce a θ-dependence in $f(\mathbf{\Sigma}, \theta)$ 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 θ parameter needed to explain experiments is not very far off the one obtained for the angle θ 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 θ

We would like to understand how the energy of the ground state of QCD depends on the parameter θ. 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 θ-term is a topological object and any physical quantity depending on it like $V(\theta)$ should be periodic in θ with period equal to 2π,

$$V(\theta+2\pi) = V(\theta). \tag{8.67}$$

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),

$$\mathcal{V}(\mathbf{\Sigma}) = -\frac{m_\pi^2 f_\pi^2}{2(m_u+m_d)}\text{tr}\left(M^\dagger\mathbf{\Sigma}+M\mathbf{\Sigma}^\dagger\right), \tag{8.69}$$

where M is given by

$$M = \left(\begin{array}{cc} e^{i\theta}m_u & 0 \\ 0 & m_d \end{array}\right). \tag{8.70}$$

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 $\mathbf{\Sigma}=\text{diag}(e^{i\varphi_1},e^{i\varphi_2})$. Furthermore, the dependence on θ in the mass matrix can be shifted to the NG boson matrix by the field redefinition

$$\mathbf{\Sigma} \longrightarrow \widetilde{\mathbf{\Sigma}} \equiv \left(\begin{array}{cc} e^{-\frac{i\theta}{2}} & 0 \\ 0 & 1 \end{array}\right)\mathbf{\Sigma}\left(\begin{array}{cc} e^{-\frac{i\theta}{2}} & 0 \\ 0 & 1 \end{array}\right) = \left(\begin{array}{cc} e^{i(\varphi_1-\theta)} & 0 \\ 0 & e^{i\varphi_2} \end{array}\right). \tag{8.71}$$

Imposing the condition $\det\widetilde{\mathbf{\Sigma}}=1$, we have $\varphi_1+\varphi_2=\theta \mod 2\pi$.

Substituting the redefined NG matrix field $\widetilde{\Sigma}$ into (8.69) with $M=\text{diag}(m_u,m_d)$, 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 θ

$$V(\theta) = -\frac{m_\pi^2 f_\pi^2}{m_u+m_d}\sqrt{m_u^2+m_d^2+2m_um_d\cos\theta}. \tag{8.75}$$

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 θ 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 θ 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.

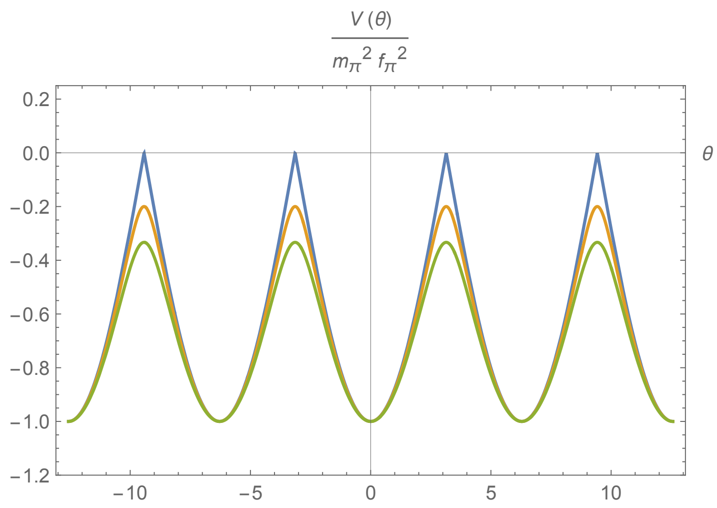


Fig. 13: Plot of $V(\theta)$ in Eq. (8.75) for three different values of the $\frac{m_u}{m_d}$ ratio: 1 (blue), 0.3 (orange), and 0.5 (green).

8.3 Enters the axion

We would like to understand the smallness of θ 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 θ-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 θ-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

$$S = \int d^4x \left(i\overline{\boldsymbol{q}}_R \not{D}\boldsymbol{q}_R + i\overline{\boldsymbol{q}}_L \not{D}\boldsymbol{q}_L + \overline{\boldsymbol{q}}_L M\boldsymbol{q}_R + \overline{\boldsymbol{q}}_R M^\dagger \boldsymbol{q}_L - \frac{1}{32\pi^2 f_a} aF^a_{\mu\nu}\widetilde{F}^{a\mu\nu} \right), \tag{8.78}$$

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 $f_\pi^3 B_0 \text{tr}\left(M^\dagger\boldsymbol{\Sigma} + \boldsymbol{\Sigma}^\dagger M\right)$, however, induces a potential that can be read off Eq. (8.75) with θ re-

[26]The fact that in the CO_2 molecule the angle θ is zero is a consequence of the dynamics of the atomic orbitals and is therefore "natural".

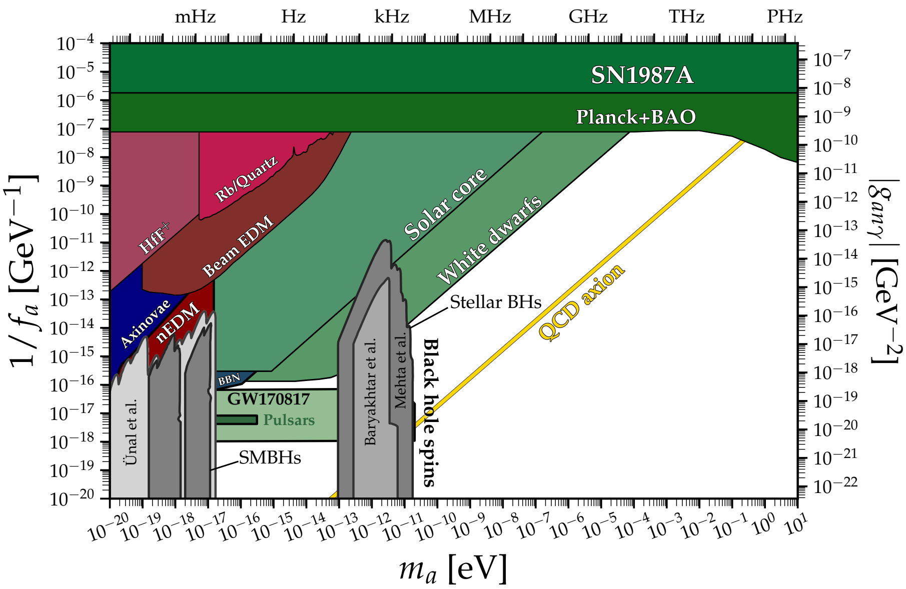


Fig. 14: Exclusion plot from Ref. [134] for the axion parameters f_a (resp. $g_{an\gamma}$) and m_a. The yellow line represents the relation given in Eq. (8.81).

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 SU(2)$\times$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 SU(2) $\times$ 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 β-decay and other processes mediated by the weak force [135–138]. Unlike gluons, which couple to quarks through a vector current $J^{\mu}_{\rm QCD} = \overline{q}\gamma^{\mu}q$, the carriers of the weak force interact with matter via the V $-$ A current $J^{\mu}_{\rm weak} = \overline{\psi}\gamma^{\mu}(\mathbb{1}-\gamma_5)\psi$, with ψ either a lepton or a quark field [139, 140].

9.1 Implementing SU(2) $\times$ U(1)$_Y$

To be more precise, β-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

$$\mathbf{L} = \left(\begin{array}{c} \nu_e \\ e^- \end{array}\right)_L, \qquad \mathbf{Q} = \left(\begin{array}{c} u \\ d \end{array}\right)_L, \tag{9.1}$$

and assume they transform in the fundamental representation $\mathbf{2}$ of the SU(2) algebra. At the same time, since right-handed electrons and quarks do not undergo β-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 U(1)$_Y$ of the electroweak gauge group is not the U(1) of Maxwell's theory. The

Leptons

	$i=1$	$i=2$	$i=3$	$t^3_{\mathbf{R}}$	$Y_{\mathbf{R}}$
$\mathbf{L}^i$	$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L$	$\frac{1}{2}\sigma^3$	$-\frac{1}{2}\mathbb{1}$
ℓ^i_R	e^-_R	μ^-_R	τ^-_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) $\times$ 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) $\times$ 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 (μ^-)

and its neutrino (ν_μ), together with the charm (c) and strange (s) quarks. The third family, on the other hand, contains the τ^- lepton, its neutrino (ν_τ), 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 $\{\ell^i_R, U^i_R, D^i_R\}$ 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_μ is to be identified with the electromagnetic field, whose gauge group will be denoted by U(1)$_{\rm em}$ to distinguish it from the one associated with the gauge field $\mathbf{B}_\mu$. The parameter θ_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_μ 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 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 θ_w. With this, the SM matter action reads

$$S_{\text{matter}} = \sum_{k=1}^{3}\int d^4x\,\Big(i\overline{\mathbf{L}}^k \not{D}\mathbf{L}^k + i\overline{\ell}_R^k\not{D}\ell_R^k + i\overline{\mathbf{Q}}^k\not{D}\mathbf{Q}^k + i\overline{U}_R^k\not{D}U_R^k + i\overline{D}_R^k\not{D}D_R^k\Big). \tag{9.12}$$

Next we look at the gauge action

$$S_{\text{gauge}} = -\frac{1}{2}\int d^4x\,\left[\operatorname{tr}\left(\mathbf{W}_{\mu\nu}\mathbf{W}^{\mu\nu}\right) + \operatorname{tr}\left(\mathbf{B}_{\mu\nu}\mathbf{B}^{\mu\nu}\right)\right], \tag{9.13}$$

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_μ gauge fields defined in Eq. (9.8), we have

$$\begin{aligned} S_{\text{gauge}} = -\int d^4x\,\bigg\{&\frac{1}{4}W_{\mu\nu}^+W^{-\mu\nu} + \frac{1}{4}Z_{\mu\nu}Z^{\mu\nu} + \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{ie}{2}\cot\theta_w W_\mu^+W_\nu^-Z^{\mu\nu} \\ &-\frac{ie}{2}W_\mu^+W_\nu^-F^{\mu\nu} + \frac{e^2}{2\sin\theta_w}\Big[(W_\mu^+W^{+\mu})(W_\mu^-W^{-\mu}) - (W_\mu^+W^{-\mu})^2\Big]\bigg\}, \end{aligned} \tag{9.14}$$

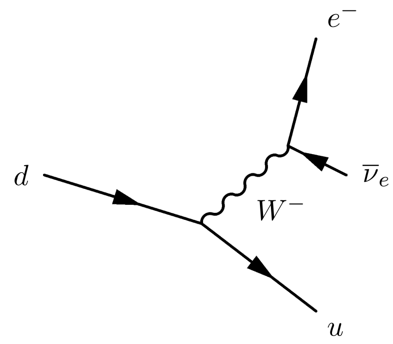
where $Z_{\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu$, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and we have defined

$$W_{\mu\nu}^\pm = \partial_\mu W_\nu^\pm - \partial_\nu W_\mu^\pm \mp e\big(W_\mu^\pm A_\nu - W_\nu^\pm A_\mu\big) \mp ie\cot\theta_w\big(W^\pm Z_\nu - W_\nu^\pm Z_\mu\big). \tag{9.15}$$

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_μ 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 SU(2) $\times$ 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 β-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^-

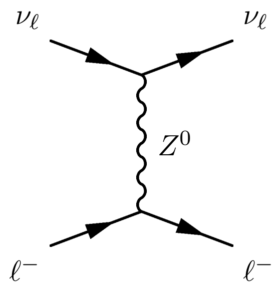
then decays into an electron and an electronic antineutrino.

$$n[udd] \longrightarrow p^+[uud] + e^- + \overline{\nu}_e \qquad \Longrightarrow \tag{9.16}$$



As a second example, we also have lepton–neutrino scattering mediated by the interchange of a Z^0

$$\ell^- + \nu_\ell \longrightarrow \ell^- + \nu_\ell \qquad \Longrightarrow \tag{9.17}$$



where ℓ stands for e, μ or τ. 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

$$(T^a_{\mathbf{R}})_{ij}(T^b_{\mathbf{R}})_{jk}(T^c_{\mathbf{R}})_{ki} + (T^a_{\mathbf{R}})_{ij}(T^c_{\mathbf{R}})_{jk}(T^b_{\mathbf{R}})_{ki} = \operatorname{tr}\left(T^a_{\mathbf{R}}\{T^b_{\mathbf{R}}, T^c_{\mathbf{R}}\}\right). \tag{9.18}$$

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

$$\sum_{L} \operatorname{tr}\left(T^a_{\mathbf{R}}\{T^b_{\mathbf{R}}, T^c_{\mathbf{R}}\}\right)_L - \sum_{R} \operatorname{tr}\left(T^a_{\mathbf{R}}\{T^b_{\mathbf{R}}, T^c_{\mathbf{R}}\}\right)_R = 0, \tag{9.19}$$

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.

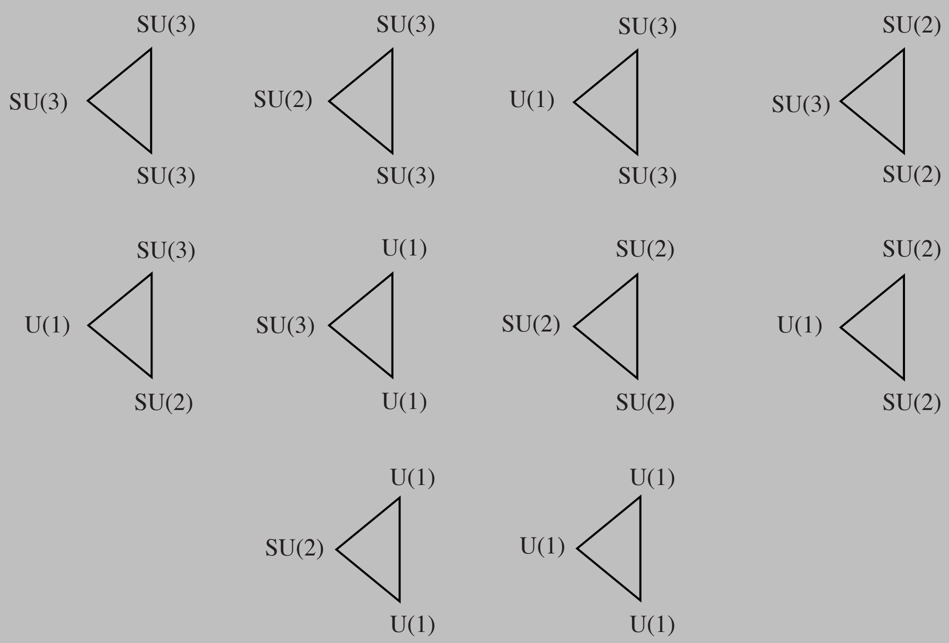
There is a simple way to summarize the group-theoretical information contained in Table 2 by just indicating the representations of the different fermion species with respect to SU(3) $\times$ SU(2) $\times$ U(1)$_Y$, including also now the gauge group factor associated with the strong force. Using the 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: $\mathrm{SU}(2)^3$, $\mathrm{SU}(2)^2\mathrm{U}(1)$, $\mathrm{SU}(3)^2\mathrm{U}(1)$, and $\mathrm{U}(1)^3$. 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 $\mathrm{SU}(2)^2\mathrm{U}(1)$, 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 $\mathrm{SU}(3)^2\mathrm{U}(1)$ triangle the color factor rules out leptons, so we have

$$\sum_{\mathrm{quarks},L} Y_L - \sum_{\mathrm{quarks},R} Y_R = 0. \tag{9.24}$$

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

$$\sum_L \operatorname{tr}(T^a_{\mathbf{R}})_L - \sum_R \operatorname{tr}(T^a_{\mathbf{R}})_R = 0. \tag{9.26}$$

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 $S_{\text{matter}} + S_{\text{gauge}}$ 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

$$\mathbf{H} = \left(\begin{array}{c} H^+ \\ H^0 \end{array}\right), \tag{9.30}$$

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

$$S_{\text{Higgs}} = \int d^4x \left[(D_\mu \mathbf{H})^\dagger D^\mu \mathbf{H} - \frac{\lambda}{4}\left(\mathbf{H}^\dagger\mathbf{H} - \frac{v^2}{2}\right)^2 \right], \tag{9.31}$$

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

$$\langle \mathbf{H} \rangle = \frac{1}{\sqrt{2}} \left(\begin{array}{c} 0 \\ v \end{array}\right). \tag{9.32}$$

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

$$\mathrm{SU}(2) \times \mathrm{U}(1)_Y \longrightarrow \mathrm{U}(1)_{\text{em}}. \tag{9.33}$$

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 β-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

$$\langle\widetilde{\mathbf{H}}\rangle = \frac{1}{\sqrt{2}}\left(\begin{array}{c} v \\ 0 \end{array}\right). \tag{9.40}$$

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

$$\begin{aligned} S_{\rm mass} = -\int d^4x \Bigg[&(\overline{e}_L,\overline{\mu}_L,\overline{\tau}_L)M^{(\ell)}\left(\begin{array}{c} e_R \\ \mu_R \\ \tau_R \end{array}\right) + (\overline{d}_L,\overline{s}_L,\overline{b}_L)M^{(q)}\left(\begin{array}{c} d_R \\ s_R \\ b_R \end{array}\right) \\ &+ (\overline{u}_L,\overline{c}_L,\overline{t}_L)\widetilde{M}^{(q)}\left(\begin{array}{c} u_R \\ c_R \\ t_R \end{array}\right) + \text{H.c.}\Bigg], \end{aligned} \tag{9.41}$$

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

$$\begin{aligned} U_L^{(\ell)\dagger}M^{(\ell)}U_R^{(\ell)} &= \text{diag}(m_e,m_\mu,m_\tau), \\ V_L^{(q)\dagger}M^{(q)}V_R^{(q)} &= \text{diag}(m_d,m_s,m_b), \\ \widetilde{V}_L^{(q)\dagger}\widetilde{M}^{(q)}\widetilde{V}_R^{(q)} &= \text{diag}(m_u,m_c,m_t), \end{aligned} \tag{9.43}$$

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

$$\left(\begin{array}{c} u'_{L,R} \\ c'_{L,R} \\ t'_{L,R} \end{array}\right) = \widetilde{V}^{(q)\dagger}_{L,R}\left(\begin{array}{c} u_{L,R} \\ c_{L,R} \\ t_{L,R} \end{array}\right), \qquad \left(\begin{array}{c} d'_{L,R} \\ s'_{L,R} \\ b'_{L,R} \end{array}\right) = V^{(q)\dagger}_{L,R}\left(\begin{array}{c} d_{L,R} \\ s_{L,R} \\ b_{L,R} \end{array}\right), \tag{9.44}$$

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

$$S \supset \int d^4x\,(\overline{u}_L,\overline{c}_L,\overline{t}_L)\gamma^\mu \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} W^+_\mu = \int d^4x\,(\overline{u}'_L,\overline{c}'_L,\overline{t}'_L)\widetilde{V}^{(q)\dagger}_L V^{(q)}_L \gamma^\mu \begin{pmatrix} d'_L \\ s'_L \\ b'_L \end{pmatrix} W^+_\mu, \quad (9.45)$$

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

$$\widetilde{V}^{(q)\dagger}_L V^{(q)}_R \equiv V_{\text{CKM}} \quad (9.46)$$

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 Λ^0 baryon (uds) is that the strange quark emits a W^- and transforms into a u-quark, thus converting the Λ^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 $V^{(q)}_{L,R}$ and $\widetilde{V}^{(q)}_{L,R}$ 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:

$$\begin{aligned} S \supset \int d^4x \Bigg[& \left(\frac{1}{2} - \frac{2}{3}\sin^2\theta_w\right)(\overline{u}_L,\overline{c}_L,\overline{t}_L)\gamma^\mu \begin{pmatrix} u_L \\ c_L \\ t_L \end{pmatrix} - \left(\frac{1}{2} - \frac{1}{3}\sin^2\theta_w\right)(\overline{d}_L,\overline{s}_L,\overline{b}_L)\gamma^\mu \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} \\ & + \frac{2}{3}\sin^2\theta_w(\overline{u}_R,\overline{c}_R,\overline{t}_R)\gamma^\mu \begin{pmatrix} u_R \\ c_R \\ t_R \end{pmatrix} - \frac{1}{3}\sin^2\theta_w(\overline{d}_R,\overline{s}_R,\overline{b}_R)\gamma^\mu \begin{pmatrix} d_R \\ s_R \\ b_R \end{pmatrix}\Bigg], \end{aligned} \quad (9.47)$$

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 $V^{(q)\dagger}_{L,R}V^{(q)}_{L,R} = \mathbb{1} = \widetilde{V}^{(q)\dagger}_{L,R}\widetilde{V}^{(q)}_{L,R}$ 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 ${}^{62}_{28}\text{Ni}$. 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 $m_p[uud] \simeq 938.3\,\text{MeV}$ and $m_d[udd] = 939.6\,\text{MeV}$, 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

$$\mathbf{H}(x) = \frac{1}{\sqrt{2}} e^{i a^I(x) t^I_{\mathbf{2}}} \left(\begin{array}{c} 0 \\ v + h(x) \end{array} \right), \tag{9.54}$$

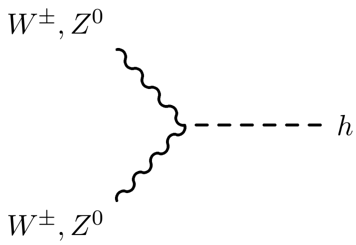
where $a^I(x)$ and $h(x)$ are the four real degrees of freedom encoding the two complex components in (9.30). In fact, as in the Abelian case of Section 5.4, we can use the gauge invariance of $S_{\rm Higgs} + S_{\rm Yukawa}$ to eliminate the global SU(2) global factor, after which we are left with a single real degree of freedom representing the Higgs boson [36]. Substituting into (9.31) and expanding, we get

$$\begin{aligned} S_{\text{Higgs}} = \int d^4x \, \Bigg[& \frac{1}{2}\partial_\mu h \partial^\mu h - \frac{\lambda v^2}{4} h^2 - \frac{\lambda v}{4} h^3 - \frac{\lambda}{16} h^4 + \frac{2m_W^2}{v} W_\mu^- W^{+\mu} h \\ & + \frac{m_W^2}{v^2} W_\mu^- W^{+\mu} h^2 + \frac{m_Z^2}{v} Z_\mu Z^\mu h + \frac{m_Z^2}{2v^2} Z_\mu Z^\mu h^2 + m_W^2 W_\mu^+ W^{-\mu} + \frac{m_Z^2}{2} Z_\mu Z^\mu \Bigg], \end{aligned} \tag{9.55}$$

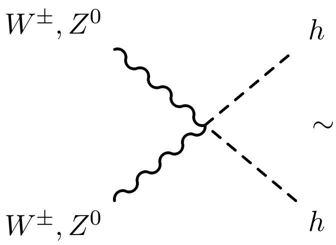
where in the last two terms we recognize the masses for the $W^\pm$ and Z^0 gauge bosons. The first thing to be noticed is that the mass of the Higgs boson is determined by the vev v and the strength λ of the Higgs quartic self-couplings,

$$m_H = v\sqrt{\frac{\lambda}{2}} = (125.25 \pm 0.17)\ \text{GeV}, \tag{9.56}$$

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



$\sim \dfrac{m_{W,Z}^2}{v}$,



$\sim \dfrac{m_{W,Z}^2}{v^2}$. (9.57)

In both cases, the strength of the coupling is proportional to the mass squared of the corresponding intermediate bosons.

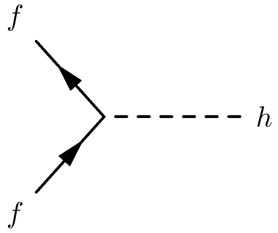
As to the coupling of the Higgs boson to fermions, this is obtained by replacing (9.54) into the Yukawa action (9.38),

$$\begin{aligned} S_{\text{Yukawa}} = -\int d^4x \, \Bigg[& (\overline{e}_L, \overline{\mu}_L, \overline{\tau}_L) \left(\frac{1}{v} M^{(\ell)}\right) \begin{pmatrix} e_R \\ \mu_R \\ \tau_R \end{pmatrix} h \\ & + (\overline{d}_L, \overline{s}_L, \overline{b}_L) \left(\frac{1}{v} M^{(q)}\right) \begin{pmatrix} d_R \\ s_R \\ b_R \end{pmatrix} h + (\overline{u}_L, \overline{c}_L, \overline{t}_L) \left(\frac{1}{v} \widetilde{M}^{(q)}\right) \begin{pmatrix} u_R \\ c_R \\ t_R \end{pmatrix} h + \text{H.c.} \Bigg]. \end{aligned} \tag{9.58}$$

This, upon switching to mass eigenstates, takes the general form

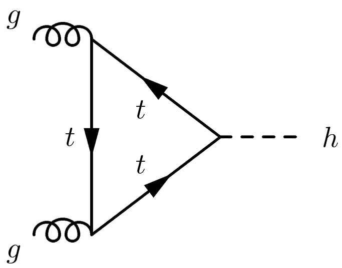
$$S_{\text{Yukawa}} = -\sum_f \frac{m_f}{v} \int d^4x \, \overline{f} f h, \tag{9.59}$$

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



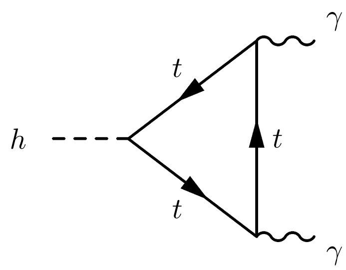
$$\sim \frac{m_f}{v}. \tag{9.60}$$

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



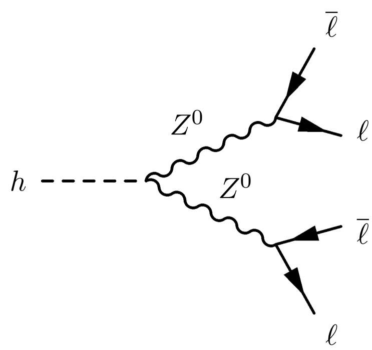
(9.61)

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 ν_R^i transforming as $(\mathbf{1},\mathbf{1})_0$ under SU(3) $\times$ SU(2) $\times$ U(1)$_Y$ (see the notation introduced on page 105), adding then the following terms to the Yukawa action

$$\Delta S_{\text{Yukawa}} = -\sum_{i=1}^{3}\int d^4x\,\Big(\widetilde{C}^{(\nu)}\overline{\mathbf{L}}^i\widetilde{\mathbf{H}}\nu_R^i + \widetilde{C}_{ji}^{(\nu)*}\overline{\nu}_R^i\widetilde{\mathbf{H}}\mathbf{L}^j\Big). \tag{9.64}$$

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 η_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 η_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 χ_-. 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

$$\psi = \frac{1}{\sqrt{2}}\begin{pmatrix} \chi_+ \\ 0 \end{pmatrix} + \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ -i\eta_C\sigma^2\chi_+ \end{pmatrix} \equiv \frac{1}{\sqrt{2}}(\psi_+ + \psi_+^c). \tag{9.73}$$

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 electromagnetic 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

$$j^{\mu}=\overline{\psi}\gamma^{\mu}\psi=\frac{1}{2}\Big(\chi_{+}^{\dagger}\sigma_{+}^{\mu}\chi_{+}+\chi_{+}^{T}\sigma_{+}^{\mu T}\chi_{+}^{*}\Big)=0. \tag{9.75}$$

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 $\psi\rightarrow e^{i\vartheta}\psi$. 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

$$\Delta S=-\sum_{i,j=1}^{3}\int d^{4}x\left(\frac{1}{2}M_{ij}\overline{\nu^{ic}}_{L}\nu_{L}^{j}+\text{H.c.}\right), \tag{9.77}$$

where because of Fermi statistics $\overline{\nu^{ic}}_{L}\nu_{L}^{j}=\overline{\nu^{jc}}_{L}\nu_{L}^{i}$ and the mass matrix $M_{ij}^{(\nu)}$ 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,

$$\Delta S=-\frac{1}{M}\sum_{i,j=1}^{3}\int d^{4}x\left[C_{ij}^{(\nu)}\left(\overline{\mathbf{L}^{ic}}\,\widetilde{\mathbf{H}}^{*}\right)\left(\widetilde{\mathbf{H}}^{\dagger}\mathbf{L}^{j}\right)+\text{H.c.}\right] \tag{9.78}$$

is invariant under $\mathrm{SU(2)}\times\mathrm{U(1)}_{Y}$. 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,

$$\Delta S = -\frac{1}{2}\sum_{i,j=1}^{3}\int d^4x\left(M_{ij}^{(\nu)}\overline{\nu^{ic}}_L\nu_L^j + \text{H.c.}\right), \tag{9.79}$$

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 ν_L^j and $\overline{\nu^{ic}}_L$ transform with the same phase [cf. (9.76)].

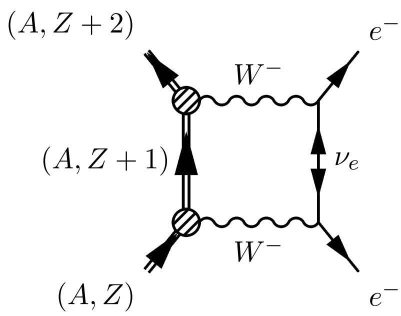
Neutrinos are regarded as one of the most promising windows to physics beyond the SM, being the main reason why neutrino physics has remained for decades one of the most exciting fields in (astro)particle physics and cosmology [147–149]. As to the question of whether the neutrino is a Dirac or a Majorana particle, however, the jury is still out. Some processes can only take place if the neutrino is its own antiparticle, most notably neutrinoless double β decay [150, 151]. A nucleus with mass and atomic numbers (A, Z) can undergo double β-decay and transmute into the nucleus $(A, Z+2)$ with emission of two electrons and two antineutrinos:

$$\begin{array}{ccl} (A,Z) & \longrightarrow & (A,Z+1) \;+\; e^- + \;\overline{\nu}_e \\ & & \qquad\quad \hookrightarrow \quad (A,Z+2) \;+\; e^- \;+\; \overline{\nu}_e \end{array}. \tag{9.81}$$

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 β-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 θ dependence in the QCD action from the θ-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 θ-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 θ_{12}, θ_{13}, and θ_{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

$$S \supset \int d^4x \left[(\overline{e}'_L, \overline{\mu}'_L, \overline{\tau}'_L) U_L^{(\ell)\dagger} \gamma^\mu \begin{pmatrix} \nu_{eL} \\ \nu_{\mu L} \\ \nu_{\tau L} \end{pmatrix} W^+_\mu + \text{H.c.} \right]. \tag{9.90}$$

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 $U_L^{(\ell)\dagger}$ 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 δ_{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 θ_{12}, θ_{13}, θ_{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 δ_{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 β decay [152]. Fits of neutrino data (including the Super-Kamiokande atmospheric neutrino data) give the following 3σ 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, λ for the second diagonal, and λ^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 Λ, 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 Λ. 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

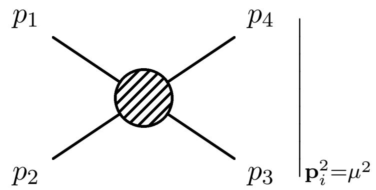
[29] Here we are going to be concerned with UV divergences associated with the high energy regime of the theory. IR divergences, which appear in the limit of low momenta, cancel once the physical question is properly posed and all contributions to the given process are taken into account.

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 λ_{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 μ is the energy scale of the process. The dependence of the action parameters on Λ 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 β 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 Λ_{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 β-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 Λ, 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 μ 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 Λ, in principle an infinite number of them. Using the language of path integrals, we symbolically summarize all this by writing

$$\int_{\mu\leq E\leq\Lambda} \mathscr{D}\Phi_0\, e^{iS_0[\Phi_0]} = e^{iS[\Phi]}, \tag{10.4}$$

where Φ_0 collectively denotes the fields of the original theory and Φ their renormalized counterparts, while $S[\Phi]$ is the action of the new theory defined at the energy scale μ. 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 Λ and μ. Their couplings $g'_n(\mu)$ are dimensionless and we see that higher-dimensional operators are suppressed by inverse powers of the high energy scale Λ.

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 Λ. 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 Λ. 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!).

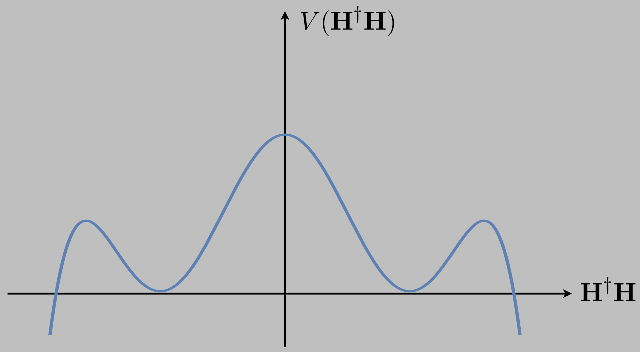
Box 17. The Planck chimney

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 λ 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 μ. 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 Λ_{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 $m_H = (125.25 \pm 0.17)$ GeV and $m_t = (172.69 \pm 0.30)$ 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 $\Lambda_{\text{inst}} \sim \sqrt{m_W \Lambda_{\mathrm{Pl}}}$. 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 Λ, 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 Λ, 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 γ-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 σ 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 β 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]].

Neutrino physics

Pilar Hernández
IFIC, Universidad de València and CSIC, E-46071 Valencia, Spain.
Email: Pilar.Hernandez@ific.uv.es

This is an update of the lectures previously published in arXiv:1708.01046. The topics discussed in this lecture include: general properties of neutrinos in the SM, the theory of neutrino masses and mixings (Dirac and Majorana), neutrino oscillations both in vacuum and in matter, as well as an overview of the experimental evidence for neutrino masses and of the prospects in neutrino oscillation physics. We also briefly comment on the relevance of neutrinos in leptogenesis and in beyond-the-Standard-Model physics.

This article should be cited as: Neutrino physics, Pilar Hernández, DOI: 10.23730/CYRSP-2025-001.139, 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. 139.

1 Introduction

Neutrinos made their first *invisible* appearance at the beginning of the 20th century as *dark* particles in radioactive β-decay. In this process a nucleus undergoes a transition

$$^A_ZX \rightarrow ^A_{Z+1} X' + e^-, \tag{1.1}$$

emitting an electron, which, by energy–momentum conservation, should have an energy approximately equal to the difference of the parent and daughter nuclear masses, Q, see Fig. 1.

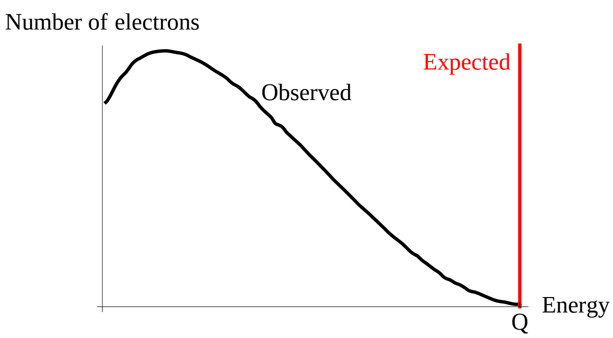


Fig. 1: Electron spectrum of β-decay.

The spectrum of the electrons was measured to be instead continuous with an end-point at Q. It took almost 20 years to come up with an explanation to this apparent violation of energy–momentum conservation. W. Pauli called for a *desperate remedy*, suggesting that in the decay, a neutral and light particle was being emitted together with the electron and escaped undetected. In that case the spectrum of the electron would indeed be continuous since only the sum of the energy of the electron and the phantom particle should equal Q. The dark particle got an Italian name: *neutrino* in honour of E. Fermi, who was among the first to take seriously Pauli's hypothesis, from which he constructed the famous theory of β-decay [1]. In this theory, the interaction responsible for β-decay is shown in Fig. 2, a four-fermion interaction with strength given by G_F, the Fermi constant.

Such interaction implies that neutrinos should also scatter off matter through the inverse beta

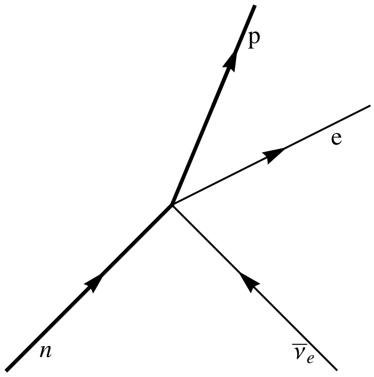


Fig. 2: Fermi four-fermion coupling responsible for β-decay.

process, $\bar{\nu}\, p \to ne^+$. Bethe and Pearls [2] estimated the cross section for such process to be

$$\sigma_{\bar{\nu}} \leq 10^{-44} \text{ cm}^2, \quad E_{\bar{\nu}} \simeq 2 \text{ MeV}, \tag{1.2}$$

and concluded that *"it is absolutely impossible to observe processes of this kin"*. Indeed this tiny cross section implies that a neutrino has a mean free path of thousands of light-years in water.

Pontecorvo [3] however was among the first to realise that it was not so hopeless. One could get a few events per day in a ton-mass scale detector with a neutrino flux of $10^{11}\nu$/cm^2/s. Such is the neutrino flux from a typical nuclear reactor at a few tens of meters distance from its core. Reines and Cowen (RC) succeeded in detecting reactor neutrinos [4, 5]. They were able to detect neutrinos via inverse beta decay in a very massive detector thanks to the extremely clean signal which combines the detection of the positron and the neutron in delayed coincidence, see Fig. 3. This experiment not only led to the discovery of anti-neutrinos, but introduced a detection technique that is still being used today in state-of-the-art reactor neutrino experiments and continues to make fundamental discoveries in neutrino physics.

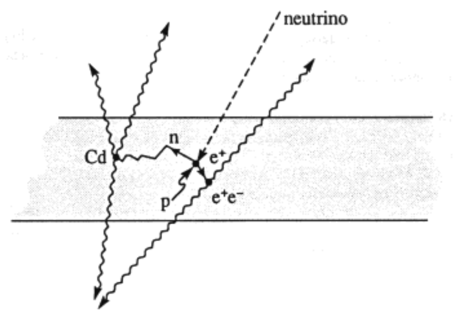


Fig. 3: Detection technique in the Reines–Cowan experiment.

Shortly after anti-neutrinos were discovered, it was realised that they come in flavours or families. The muon had been discovered in cosmic rays much earlier, and pion decay to muons is an analogous process to β-decay:

$$\pi^- \to \mu^- \bar{\nu}_\mu. \tag{1.3}$$

It was understood that also in this case a (anti-)neutrino is emitted but, accompanying a μ instead of an electron, it had a different identity to that in β-decay. Since the energy transfer in this process is higher than in β-decay, and the neutrino cross-sections grow fast with energy in the Fermi theory, it would actually be easier to detect this new type of neutrino.

In 1962 Lederman, Schwartz and Steinberger (LSS) detected for the first time neutrinos from pion decay by creating the first accelerator neutrino beam [6]. The accelerated proton beam is made to hit a fixed target producing pions and other hadrons that decay into neutrinos and other particles, mimicking what happens in cosmic rays. If a thick shield intercepts the secondary particles, all particles except the neutrinos are stopped, see Fig. 4. Finally a neutrino detector is located behind the shield. A neutrino event will induce the appearance of a muon in the detector. Again this was such a great idea that we are still making discoveries with the modern versions of the LSS experiment, in the so-called conventional accelerator neutrino beams.

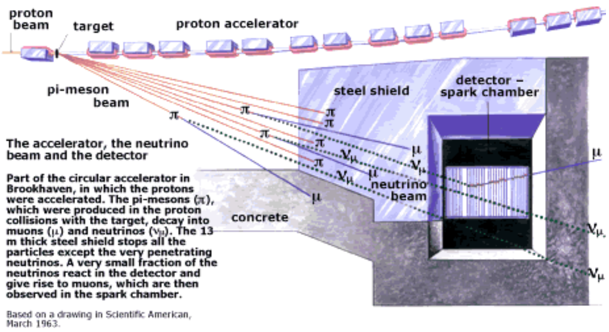


Fig. 4: Lederman, Schwartz, Steinberger experiment.

Kinematical effects of neutrino masses were searched for by measuring very precisely the end-point of the lepton energy spectrum in weak decays, that gets modified if neutrinos are massive. In particular the most stringent limit is obtained from tritium β-decay for the "electron" neutrino:

$$^3H \to ^3\text{He} + e^- + \bar{\nu}_e. \tag{1.4}$$

Figure 5 shows the effect of a neutrino mass in the end-point electron energy spectrum in this decay.

The best limit has been recently improved by the Katrin experiment [7]:

$$m_{\nu_e} < 0.8\,\text{eV}(90\%\text{CL})\;, \tag{1.5}$$

which aims at reaching a sensitivity of 0.2 eV. The direct limits from processes involving μ, τ leptons are much weaker. The best limit on the ν_μ mass (m_{ν_μ} < 170 keV [8]) was obtained from the end-point spectrum of the decay $\pi^+ \to \mu^+\nu_\mu$, while that on the ν_τ mass was obtained at LEP (m_{ν_τ} <

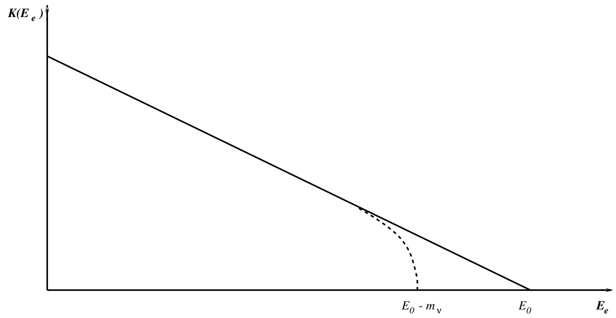


Fig. 5: Effect of a neutrino mass in the end-point of the lepton energy spectrum in β decay.

Table 1: Irreducible fermionic representations in the Standard Model: $(d_{SU(3)}, d_{SU(2)})_Y$.

$(\mathbf{1},\mathbf{2})_{-\frac{1}{2}}$	$(\mathbf{3},\mathbf{2})_{\frac{1}{6}}$	$(\mathbf{1},\mathbf{1})_{-1}$	$(\mathbf{3},\mathbf{1})_{\frac{2}{3}}$	$(\mathbf{3},\mathbf{1})_{-\frac{1}{3}}$
$\begin{pmatrix}\nu_e\\ e\end{pmatrix}_L$	$\begin{pmatrix}u^i\\ d^i\end{pmatrix}_L$	e_R	u^i_R	d^i_R
$\begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}_L$	$\begin{pmatrix}c^i\\ s^i\end{pmatrix}_L$	μ_R	c^i_R	s^i_R
$\begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}_L$	$\begin{pmatrix}t^i\\ b^i\end{pmatrix}_L$	τ_R	t^i_R	b^i_R

18.2 MeV [9]) from the decay $\tau \to 5\pi\nu_\tau$. Neutrinos in the Standard Model were therefore conjectured to be massless.

2 Neutrinos in the Standard Model

The Standard Model (SM) is a gauge theory based on the gauge group $SU(3) \times SU(2) \times U_Y(1)$. All elementary particles arrange in irreducible representations of this gauge group. The quantum numbers of the fermions $(d_{SU(3)}, d_{SU(2)})_Y$ are listed in Table 1.

Under gauge transformations neutrinos transform as doublets of $SU(2)$, they are singlets under $SU(3)$ and their hypercharge is $-1/2$. The electric charge, given by $Q = T_3 + Y$, vanishes. They are therefore the only particles in the SM that carry no conserved charge.

The two most intriguing features of Table 1 are its left–right or chiral asymmetry, and the three-fold repetition of family structures. Neutrinos have been essential in establishing both features.

2.1 Chiral structure of the weak interactions

The left and right entries in Table 1 have well defined chirality, negative and positive respectively. They are two-component spinors or Weyl fermions, the smallest irreducible representation of the Lorentz group representing spin $1/2$ particles. Only fields with negative chirality carry the $SU(2)$ charge. For free fermions moving at the speed of light (i.e., massless), the chiral states have a well defined helicity, i.e they are eigenstates of the helicity operator, $\mathbf{\Sigma} = \frac{\mathbf{s}\cdot\mathbf{p}}{|\mathbf{p}|}$, that measures the component of the spin in the direction of the momentum. This is not inconsistent with Lorentz invariance, since for a fermion travelling at the speed of light, the helicity is the same in any reference frame. In other words, the helicity operator commutes with the Hamiltonian for a massless fermion and is thus a good quantum number.

The discrete symmetry under CPT (charge conjugation, parity, and time reversal), which is a basic building block of any Lorentz invariant and unitary quantum field theory (QFT), requires that for any left-handed particle, there exists a right-handed antiparticle, with opposite charge, but the right-handed particle state may not exist. A Weyl fermion field represents therefore a particle of negative helicity and an antiparticle with positive one.

Parity however transforms left and right fields into each other, thus the left-handedness of the weak interactions implies that parity is maximally broken in the SM. The breaking is nowhere more obvious than for neutrinos since the parity partner of the neutrino does not exist. All the remaining fermions in the SM come in parity pairs, albeit with different $SU(2) \times U(1)$ charges. Since this gauge symmetry is spontaneously broken, the left and right fields combine into massive Dirac fermions, that is a four component representation of the Lorentz group and parity, which represents a particle and an antiparticle with either helicity. The chirality components are recovered from the four-component Dirac spinor by the chiral projectors

$$\psi_L = P_L\psi = \frac{1-\gamma_5}{2}\psi, \quad \psi_R = P_R\psi = \frac{1+\gamma_5}{2}\psi. \tag{2.1}$$

The SM resolved the Fermi interaction as being the result of the exchange of the $SU(2)$ massive W boson as in Fig. 6.

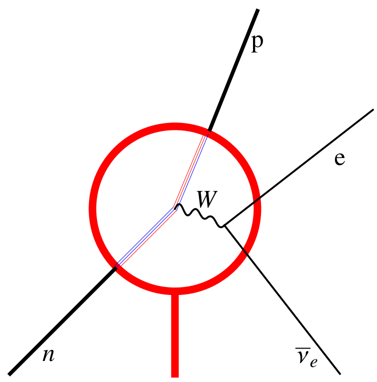


Fig. 6: β-decay process in the SM.

Neutrinos interact in the SM via charged and neutral currents:

$$\mathcal{L}_{SM} \supset -\frac{g}{\sqrt{2}}\sum_\alpha \bar{\nu}_\alpha\gamma_\mu P_L l_\alpha W^+_\mu - \frac{g}{2\cos\theta_W}\sum_\alpha \bar{\nu}_\alpha\gamma_\mu P_L \nu_\alpha Z^+_\mu + h.c. \tag{2.2}$$

The weak current is therefore $V{-}A$ since it only couples to the left fields: $\gamma_\mu P_L \propto \gamma_\mu{-}\gamma_\mu\gamma_5$. This structure is clearly seen in the kinematics of weak decays involving neutrinos, such as the classic

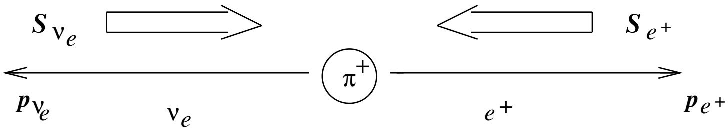


Fig. 7: Kinematics of pion decay: two recoiling particles must have same helicity to ensure angular momentum conservation.

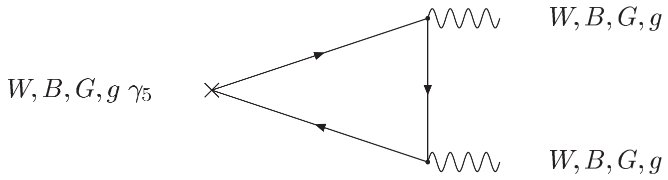


Fig. 8: Triangle diagrams that can give rise to anomalies. W, B, G are the gauge bosons associated to the $SU(2), U_Y(1), SU(3)$ gauge groups, respectively, and g is the graviton

example of pion decay to $e\bar{\nu}_e$ or $\mu\bar{\nu}_\mu$. In the limit of vanishing electron or muon mass, this decay is forbidden, because the spin of the initial state is zero and thus it is impossible to conserve simultaneously momentum and angular momentum if the two recoiling particles must have opposite helicities, as shown in Fig. 7. The decay amplitude is therefore proportional to the lepton mass and the ratio of the decay rates to electrons and muons, in spite of the larger phase space in the former, is strongly suppressed by the factor $\left(\frac{m_e}{m_\mu}\right)^2 \sim 2 \times 10^{-5}$.

Another profound consequence of the chiral nature of the weak interaction is anomaly cancellation. The chiral coupling of fermions to gauge fields leads generically to inconsistent gauge theories due to chiral anomalies: if any of the diagrams depicted in Fig. 8 is non-vanishing, the weak current which is conserved at tree level is not at one loop, implying a catastrophic breaking of gauge invariance. Anomaly cancellation is the requirement that all these triangle diagrams vanish, which imposes strong constraints on the hypercharge assignments of the fermions in the SM, which are *miraculously* satisfied:

$$\overbrace{\sum_{i=\text{quarks}} Y_i^L - Y_i^R}^{GGB} = \overbrace{\sum_{i=\text{doublets}} Y_i^L}^{WWB} = \overbrace{\sum_i Y_i^L - Y_i^R}^{Bgg} = \overbrace{\sum_i (Y_i^L)^3 - (Y_i^R)^3}^{B^3} = 0, \tag{2.3}$$

where $Y_i^{L/R}$ are the hypercharges of the left/right components of the fermionic field i, and the triangle diagram corresponding to each of the sums is indicated above the bracket.

2.2 Family structure

Concerning the family structure, we know, thanks to neutrinos, that there are exactly three families in the SM. An extra SM family with quarks and charged leptons so heavy that cannot be produced at the energies explored so far in colliders, would also have massless neutrinos that would contribute to the

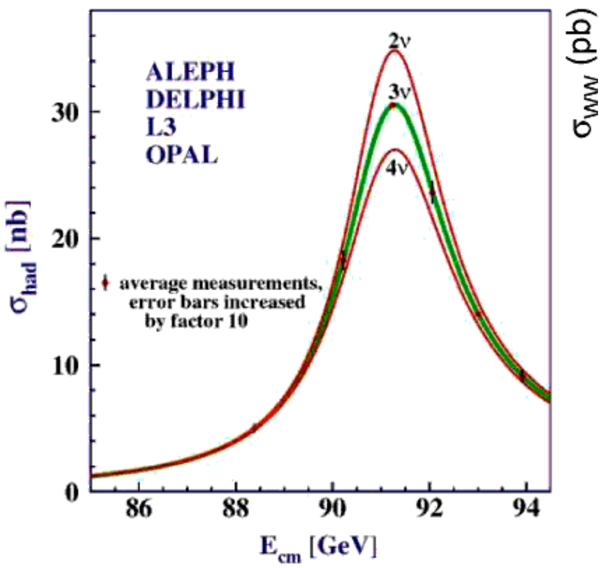


Fig. 9: Z^0 resonance from the LEP experiments. Data are compared to the case of $N_\nu = 2, 3$ and 4

invisible Z^0 decay:

$$Z^0 \to \bar{\nu}_\alpha \nu_\alpha. \tag{2.4}$$

The invisible width of the Z^0 has been measured at LEP with an impressive precision, as shown in Fig. 9 [10]. This measurement has been recently revised [11, 12] with a reduced systematic error and excludes any number of standard families different from three:

$$N_\nu = \frac{\Gamma_{\rm inv}}{\Gamma_{\bar{\nu}\nu}} = 2.9963 \pm 0.00074. \tag{2.5}$$

3 Massive neutrinos

Neutrinos are ubiquitous in our surroundings. If we open our hand, it will be crossed each second by about $\mathcal{O}(10^{12})$ neutrinos from the sun, about $\mathcal{O}(10)$ from the atmosphere, about $\mathcal{O}(10^9)$ from natural radioactivity in the earth and even $\mathcal{O}(10^{12})$ relic neutrinos from the Big Bang. In 1987, the Kamiokande detector in Japan observed the neutrino burst from a SuperNova that exploded in the Large Magellanic Cloud, at a distance of 160 thousand light years from earth. For a few seconds, the supernova neutrino flux was of the same order of magnitude as the flux of solar neutrinos!

Using many of these sources as well as others from reactors and accelerators, a decade of revolutionary neutrino experiments have demonstrated that, for the time being, neutrinos are the less standard of the SM particles. They have tiny masses and this necessarily requires new degrees of freedom with respect to those in Table 1.

A massive fermion necessarily has two states of helicity, since it is always possible to reverse the helicity of a state that moves at a slower speed than light by looking at it from a boosted reference frame. What is the right-handed state of the neutrino? It turns out there are two ways to proceed.

Let us consider the case of free fermions. A four-component Dirac fermion can be made massive adding the following mass term to the Lagrangian:

$$-\mathcal{L}_m^{\rm Dirac} = m\bar{\psi}\psi = m(\overline{\psi_L + \psi_R})(\psi_L + \psi_R) = m(\overline{\psi_L}\psi_R + \overline{\psi_R}\psi_L). \tag{3.1}$$

A Dirac mass term couples the left-handed and right-handed chiral components of the fermion field, and therefore this coupling vanishes identically in the case of a Weyl fermion.

Can one give a mass to a two-component Weyl fermion? As first noticed by Majorana, this indeed can be done with the following mass term:

$$-\mathcal{L}_m^{\rm Majorana} = \frac{m}{2}\overline{\psi^c}\psi + \frac{m}{2}\overline{\psi}\psi^c = \frac{m}{2}\psi^T C\psi + \frac{m}{2}\bar{\psi}C\bar{\psi}^T, \tag{3.2}$$

where

$$\psi^c \equiv C\bar{\psi}^T = C\gamma_0\psi^*. \tag{3.3}$$

It is easy to check that the Majorana mass term satisfies the required properties:

1) It can be constructed with a two-component spinor or Weyl fermion: if $\psi = P_L\psi$

$$\psi^T C\psi = \psi_L^T i\sigma_2\psi_L, \tag{3.4}$$

which does not vanish in the absence of the right chiral component.

2) It is Lorentz invariant. It is easy to show, using the properties of the gamma matrices that under a Lorentz transformation ψ and ψ^c transform in the same way,

$$\psi \to e^{-\frac{i}{4}\omega_{\mu\nu}\sigma^{\mu\nu}}\psi \equiv S(\Lambda)\psi, \quad \psi^c \to S(\Lambda)\psi^c, \tag{3.5}$$

with $\sigma_{\mu\nu} \equiv = \frac{i}{4}[\gamma_\mu, \gamma_\nu]$, and therefore the bilinear $\overline{\psi^c}\psi$ is Lorentz invariant.

3) The equation of motion derived from Eq. (3.2) for a free majorana fermion has plane wave solutions satisfying the relativistic relation for a massive fermion:

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

In the SM none of the mass terms of Eqs. (3.1) and (3.2) are gauge invariant. Spontaneous symmetry breaking allows to generate the Dirac mass term from Yukawa couplings for all fermions in the SM, while the Majorana mass term can only be generated for neutrinos. Let us see how this works.

3.1 Massive Dirac neutrinos

We can enlarge the SM by adding a set of three right-handed neutrino, ν_R states, with quantum numbers $(1,1)_0$, i.e. singlets under all the gauge groups. A new Yukawa (Fig. 10) coupling of these new states with the lepton doublet is exactly gauge invariant and therefore can be added to the SM:

$$-\mathcal{L}_m^{\rm Dirac} = \overline{L}\,\lambda\tilde{\Phi}\,\nu_R + \text{ h.c.} \tag{3.6}$$

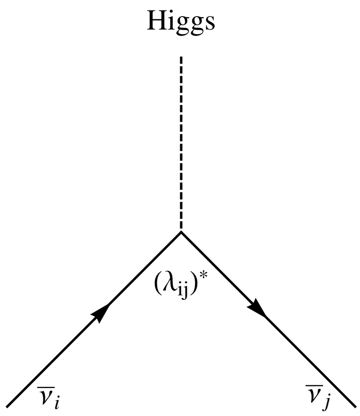


Fig. 10: Neutrino Yukawa coupling.

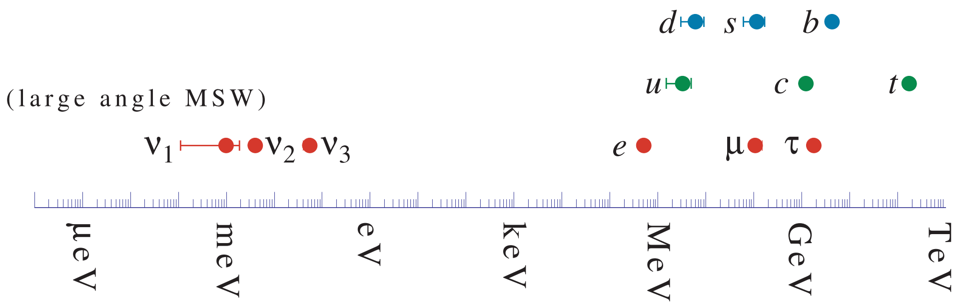


Fig. 11: Fermion spectrum in the Standard Model.

where $L = (\nu\ l)$ is the lepton doublet, $\tilde{\Phi} \equiv i\sigma_2\phi^*$ and ϕ is the Higgs field, with quantum numbers $(\mathbf{1},\mathbf{2})_{-\frac{1}{2}}$. Upon spontaneous symmetry breaking the scalar doublet gets a vacuum expectation value $\langle\tilde{\Phi}\rangle = (\frac{v}{\sqrt{2}}\ 0)$, and therefore a neutrino Dirac mass term is generated

$$-\mathcal{L}_m^{\text{Dirac}} \to -\ \overline{\nu_L}\ \lambda\frac{v}{\sqrt{2}}\nu_R +\ \text{h.c.} \tag{3.7}$$

The neutrino mass matrix is proportional to the Higgs vacuum expectation value, in complete analogy to the remaining fermions:

$$m_\nu = \lambda\frac{v}{\sqrt{2}}. \tag{3.8}$$

There are two important consequences of Dirac neutrinos. First, there is a new hierarchy problem in the SM to be explained: why are neutrinos so much lighter than the remaining leptons, even those in the same family (see Fig. 11), if they get the mass in the same way? This requires a large hierarchy in the Yukawa couplings that should differ in many orders of magnitude. Secondly, an accidental global symmetry, lepton number L, that counts the number of leptons minus that of antilepton, remains exactly conserved at the classical level,[1] just as baryon number, B, is.

[1]As usual $\mathrm{B}+\mathrm{L}$ is broken by the anomaly and only $\mathrm{B}-\mathrm{L}$ remains exact at all orders.

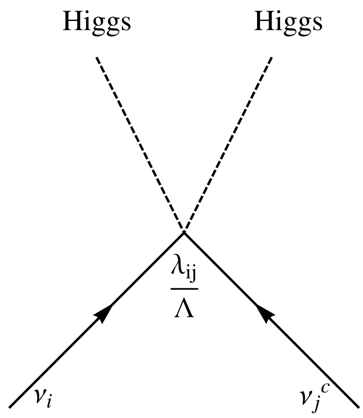


Fig. 12: Weinberg operator.

3.2 Massive Majorana neutrinos

Since the combination $\bar{L}\tilde{\phi}$ is a singlet under all gauge groups, the Majorana-type contraction (see Fig. 12):

$$-\mathcal{L}_m^{\rm Majorana} = \bar{L}\tilde{\phi}\,\alpha C\tilde{\phi}^T\bar{L}^T + h.c., \tag{3.9}$$

is gauge invariant. This term, first writen down by Weinberg [13], gives rise to a Majorana mass term for neutrinos upon spontaneous symmetry breaking:

$$-\mathcal{L}_m^{\rm Majorana} \to \bar{\nu}_L\alpha\frac{v^2}{2}C\bar{\nu}_L^T + h.c., \tag{3.10}$$

The neutrino mass matrix in this case is given by:

$$m_\nu = \alpha v^2. \tag{3.11}$$

The Weinberg operator has dimension 5, and therefore the coupling $[\alpha] = -1$. We can write it in terms of a dimensionless coupling as

$$\alpha = \frac{\lambda}{\Lambda}, \tag{3.12}$$

where Λ is a new physics scale, in principle unrelated to the electroweak scale.

The consequences of the SM neutrinos being massive Majorana particles are profound. If the scale Λ is much higher than the electroweak scale v, a strong hierarchy between the neutrino and the charged lepton masses arises naturally. If all dimensionless couplings λ are of the same order, neutrino masses are suppressed by a factor v/Λ with respect to the charged fermions. On the other hand, Weinberg's operator violates lepton number L and provides a new seed for generating the matter/antimatter asymmetry in the Universe as we will see.

Even though the Majorana mechanism to generate neutrino masses does not involve any extra degree of freedom with respect to those in the SM, the existence of the Weinberg coupling implies that cross sections involving for example the scattering of neutrinos and the Higgs will grow with energy, ultimately violating unitarity. The situation is analogous to that of the Fermi interaction of Fig. 2. The SM resolved this interaction at higher energies as being the result of the interchange of a heavy vector boson, Fig. 6. The Majorana coupling, if it exists, should also represent the effect at low energies of the exchange of one or more unknown massive states. What those states are remains one of the most interesting open questions in neutrino physics.

Finally, it is interesting to note that the anomaly cancellation conditions fix all the hypercharges in this case (i.e., there is only one possible choice for the hypercharges that satisfies Eq. (2.3)), which implies that electromagnetic charge quantization is the only possibility in a field theory with the same matter content as the SM.

3.3 Neutrino masses and physics beyond the Standard Model

Any new physics beyond the standard model (BSM) characterized by a high scale, Λ, will induce effects at low energies $E \ll \Lambda$ that can be described by an effective field theory [14, 15] of the form:

$$\mathcal{L}_{\rm eff} = \mathcal{L}_{\rm SM} + \sum_i \frac{\alpha_i}{\Lambda} O_i^{d=5} + \sum_i \frac{\beta_i}{\Lambda^2} O_i^{d=6} + ... \tag{3.13}$$

It is the most general Lagrangian which includes the SM and an infinite tower of operators constructed out of the SM fields respecting Lorentz and gauge symmetries. In principle such a theory depends on infinite new couplings, one per new independent operator, and it is therefore not predictive. However, if we are interested in describing processes at energies $E \ll \Lambda$, we can truncate the sum of operators up to a given dimension d in such a way that our predictions are correct up to order $\left(\frac{E}{\Lambda}\right)^{d-4}$.

The operators of lowest dimension are the most relevant at low energies. It turns out that there is only one such operator of the lowest possible dimension, $d = 5$, which is precisely the Weinberg operator of Eq. (3.9). In this perspective, it is natural to expect that the first indication of BSM physics is precisely Majorana neutrino masses. While many types of BSM theories can give rise to neutrino masses, generically they will induce other new physics effects represented by the operators of $d = 6$ and higher.

4 Neutrino masses and lepton mixing

Neutrino masses, whether Dirac or Majorana, imply lepton mixing [16, 17]. The Yukawa coupling in Eq. (3.6) is a generic complex matrix in flavour space, while that in Eq. (3.9) is a generic complex symmetric matrix, and the same holds for the corresponding leptonic mass matrices:

$$-\mathcal{L}_m^{\rm Dirac} = \overline{\nu_L^i}\,(M_\nu)_{ij}\,\nu_R^j + \overline{l_L^i}\,(M_l)_{ij}\,l_R^j + \text{ h.c.} \tag{4.1}$$

$$-\mathcal{L}_m^{\rm Majorana} = \frac{1}{2}\overline{\nu_L^i}\,(M_\nu)_{ij}\,\nu_L^{cj} + \overline{l_L^i}\,(M_l)_{ij}\,l_R^j + \text{ h.c.} \tag{4.2}$$

In the Dirac case, the two mass matrices can be diagonalized by a bi-unitary rotation:

$$M_\nu = U_\nu^\dagger {\rm Diag}(m_1, m_2, m_3) V_\nu, \quad M_l = U_l^\dagger {\rm Diag}(m_e, m_\mu, m_\tau) V_l, \tag{4.3}$$

while in the Majorana case, the neutrino mass matrix, being symmetric, can be taken to a diagonal form by

$$M_\nu = U_\nu^\dagger \mathrm{Diag}(m_1, m_2, m_3) U_\nu^*. \tag{4.4}$$

We can go to the mass basis by rotating the fields as:

$$\nu'_R = V_\nu \nu_R,\ \nu'_L = U_\nu \nu_L,\ \ l'_R = V_l l_R,\ l'_L = U_l l_L. \tag{4.5}$$

In this basis the charged-current interactions are no longer diagonal, in complete analogy with the quark sector (see Fig. 13):

$$\mathcal{L}_{CC}^{\text{lepton}} = -\frac{g}{\sqrt{2}} \bar{l}'_i \gamma_\mu P_L W_\mu^+ \underbrace{(U_l^\dagger U_\nu)_{ij}}_{U_{\text{PMNS}}} \nu'_j + \text{h.c.} \tag{4.6}$$

The mixing matrix in the lepton sector is referred to as the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix, analogous to the CKM one in the quark sector.

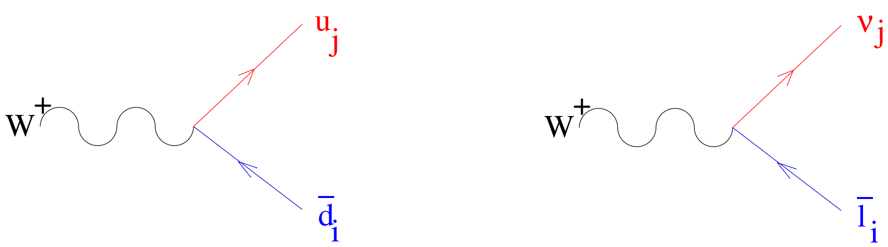


Fig. 13: Quark and lepton mixing.

The number of physical parameters in the lepton mixing matrix, U_{PMNS}, can easily be computed by counting the number of independent real and imaginary elements of the Yukawa matrices and eliminating those that can be absorbed in field redefinitions. The allowed field redefinitions are the unitary rotations of the fields that leave the rest of the Lagrangian invariant (only those that are not symmetries of the full Lagrangian when lepton masses are included are efficient in absorbing flavour parameters).

In the Dirac case, it is possible to rotate independently the left-handed lepton doublet, together with the right-handed charged leptons and neutrinos, that is $U(n)^3$, for a generic number of families n. However, this includes total lepton number which remains a symmetry of the massive theory and thus cannot be used to reduce the number of physical parameters in the mass matrix. The parameters that can be absorbed in field redefinitions are thus the parameters of the group $U(n)^3/U(1)$ (that is $\frac{3(n^2-n)}{2}$ real, $\frac{3(n^2+n)-1}{2}$ imaginary).

In the case of Majorana neutrinos, there is no independent right-handed neutrino field, nor is lepton number a good symmetry. Therefore the number of field redefinitions is the number of parameters of the elements in $U(n)^2$ (that is $n^2 - n$ real and $n^2 + n$ imaginary).

The resulting real physical parameters are the mass eigenstates and the mixing angles, while the resulting imaginary parameters are CP-violating phases. All this is summarized in Table 2. Dirac and Majorana neutrinos differ only in the number of observables phases. For three families ($n = 3$), there is

Table 2: Number of real and imaginary parameters in the Yukawa matrices, of those that can be absorbed in field redefinitions. The difference between the two is the number of observable parameters: the lepton masses (m), mixing angles (θ), and imaginary phases (ϕ).

	Yukawas	Field redefinitions	$No.\ m$	$No.\ \theta$	$No.\ \phi$
Dirac	$\lambda_l,\ \lambda_\nu$	$U(n)^3/U(1)$			
Real, Im	$2n^2, 2n^2$	$\frac{3(n^2-n)}{2}, \frac{3(n^2+n)-1}{2}$	$2n$	$\frac{n^2-n}{2}$	$\frac{(n-2)(n-1)}{2}$
Majorana	$\lambda_l,\ \alpha_\nu^T = \alpha_\nu$	$U(n)^2$			
Real,Im	$n^2 + \frac{n(n+1)}{2}, n^2 + \frac{n(n+1)}{2}$	$n^2 - n,\ n^2 + n$	$2n$	$\frac{n^2-n}{2}$	$\frac{n^2-n}{2}$

just one Dirac phase and three in the Majorana case.

A standard parametrization of the mixing matrices for Dirac, U_{PMNS}, and Majorana, $\tilde{U}_{\text{PMNS}}$, is given by

$$\begin{aligned} U_{\text{PMNS}} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ \tilde{U}_{\text{PMNS}} &= U_{\text{PMNS}}(\theta_{12}, \theta_{13}, \theta_{23}, \delta) \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\alpha_1} & 0 \\ 0 & 0 & e^{i\alpha_2} \end{pmatrix}, \end{aligned} \tag{4.7}$$

where in all generality $\theta_{ij} \in [0, \pi/2]$ and $\delta, \alpha_1, \alpha_2 \in [0, 2\pi]$.

5 Majorana versus Dirac

It is clear that establishing the Majorana nature of neutrinos is of great importance, since it would imply the existence of a new physics scale. In principle there are very clear signatures, such as the one depicted in Fig. 14, where a ν_μ beam from π^+ decay is intercepted by a detector, D. In the Dirac case, the interaction of neutrinos on the detector via a charged current interaction will produce only a μ^- in the final state. If neutrinos are Majorana, a wrong-sign muon in the final state is also possible. Unfortunately the rate for μ^+ production is suppressed by m_ν/E in amplitude with respect to the μ^-. For example, for $E_\nu = \mathcal{O}(1)$ GeV and $m_\nu \sim \mathcal{O}(1)$ eV the cross section for this process will be roughly 10^{-18} times the usual CC neutrino cross section.

The best hope of observing a rare process of this type seems to be the search for neutrinoless double-beta decay ($2\beta0\nu$), the right diagram of Fig. 15. The background to this process is the standard double-beta decay depicted on the left of Fig. 15, which has been observed to take place for various isotopes with a lifetime of $T_{2\beta2\nu} > 10^{19}$–10^{21} years.

If the source of this process is just the Majorana ν mass, the inverse lifetime for this process is

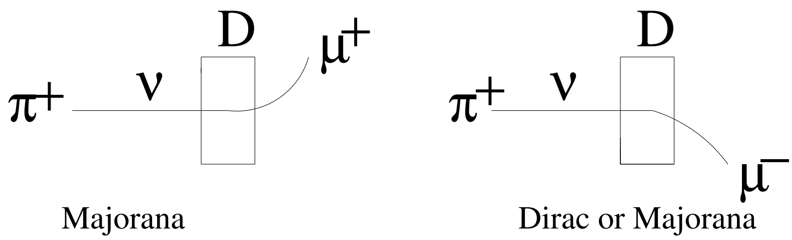


Fig. 14: A neutrino beam from π^+ decay (ν_μ) could interact in the magnetized detector producing a μ^+ only if neutrinos are Majorana.

given by

$$T_{2\beta 0\nu}^{-1} \simeq \underbrace{G^{0\nu}}_{\text{Phase}} \underbrace{\left|M^{0\nu}\right|^2}_{\text{NuclearM.E.}} \underbrace{\left|\sum_i \left(\tilde{U}_{\text{PMNS}}^{ei}\right)^2 m_i\right|^2}_{|m_{ee}|^2} . \tag{5.1}$$

In spite of the suppression in the neutrino mass (over the energy of this process), the neutrinoless mode has a phase factor orders of magnitude larger than the 2ν mode, and as a result present experiments searching for this rare process have already set bounds on neutrino masses in the eV range as shown in Table 3.

Fig. 15: 2β decay: normal (left) and neutrinoless (right).

Table 3: Present bounds at 90%CL from some recent neutrinoless double-beta-decay experiments [18].

Experiment	Nucleus	$\lvert m_{ee}\rvert$
EXO-200	^{136}Xe	< 0.093–0.286 eV
AMoRE	^{100}Mo	< 1.2–2.1 eV
GERDA	^{76}Ge	< 0.079–0.18 eV
KamLAND-Zen	^{136}Xe	< 0.061–0.165 eV
CUORE	^{130}Te	< 0.11–0.52 eV

6 Neutrino oscillations

The most spectacular implication of neutrino masses and mixings is the macroscopic quantum phenomenon of neutrino oscillations, first introduced by B. Pontecorvo [19]. The Nobel Prize of 2015 was awarded to T. Kajita (from the SuperKakiokande collaboration) and A. B. McDonald (from the SNO collaboration) for the *discovery of neutrino oscillations, which shows that neutrinos have a mass*.

We have seen that if neutrinos are massive the neutrino flavour fields $(\nu_e, \nu_\mu, \nu_\tau)$, that couple via CC to the leptons (e, μ, τ) , are unitary combinations of the mass eigenstates fields (ν_1, ν_2, ν_3):

$$\left(\begin{array}{c} \nu_e \\ \nu_\mu \\ \nu_\tau \end{array}\right) = U_{\rm PMNS}(\theta_{12}, \theta_{13}, \theta_{23}, {\rm phases}) \left(\begin{array}{c} \nu_1 \\ \nu_2 \\ \nu_3 \end{array}\right). \tag{6.1}$$

In a neutrino oscillation experiment, neutrinos are produced by a source (e.g. pion or μ decays, nuclear reactions, etc) and are detected some macroscopic distance, L, away from the production point. They are produced and detected via weak processes in combination with a given lepton flavour, that is in flavour states or a combination of mass eigenstates. As these states propagate undisturbed in space-time from the production to the detection regions, the different mass eigenstates, having slighly different phase velocities, pick up different phases, resulting in a non-zero probability that the state that arrives at the detector is in a different flavour combination to the one originally produced, see Fig. 16. The probability for this flavour transition oscillates with the distance travelled.

Two ingredients are mandatory for this phenomenon to take place:

– neutrinos must keep quantum coherence in propagation over macroscopic distances, which is only possible because they are so weakly interacting
– there is sufficient uncertainty in momentum at production and detection so that a coherent flavour state can be produced[2].

The master formula for the oscillation probability of ν_α turning into a ν_β is

$$P(\nu_\alpha \to \nu_\beta) = \sum_{i,j} U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta j} e^{-i\frac{\Delta m^2_{ji} L}{2|\mathbf{p}|}}, \tag{6.2}$$

where $\Delta m^2_{ji} \equiv m^2_i - m^2_j$, $U_{\alpha i}$ are the elements of the PMNS matrix, L is the baseline and $\mathbf{p}$ is the neutrino momentum.

There are many ways to derive this formula. The simplest way that appears in most textbooks uses simple quantum mechanics, where neutrinos are treated as plane waves. A slightly more rigorous method treats neutrinos as wave packets. Finally, it is also possible to derive it from QFT, where neutrinos are treated as intermediate virtual states. The different methods make more or less explicit the basic necessary conditions of neutrino oscillations mentioned above, and therefore are more or less prone to quantum paradoxes.

[2]If the momentum uncertainty is sufficiently small one could kinematically distinguish the mass eigenstate being produced/detected.

Fig. 16: Neutrino oscillations.

6.1 Plane wave derivation

Let us suppose that a neutrino of flavor α is produced at t_0. It is therefore a superposition of the mass eigenstates that we assume to be plane waves with spatial momentum $\mathbf{p}$:

$$|\nu_\alpha(t_0)\rangle = \sum_i U^*_{\alpha i}|\nu_i(\mathbf{p})\rangle. \tag{6.3}$$

The mass eigenstates are eigenstates of the free Hamiltonian:

$$\hat{H}|\nu_i(\mathbf{p})\rangle. = E_i(\mathbf{p})|\nu_i(\mathbf{p})\rangle, \quad E_i(\mathbf{p})^2 = \mathbf{p}^2 + m_i^2. \tag{6.4}$$

The time evolution operator from $t_0 \to t$ is given by $e^{-i\hat{H}(t-t_0)}$ and therefore the state at time t is given by

$$|\nu_\alpha(t)\rangle = e^{-i\hat{H}(t-t_0)}|\nu_\alpha(t_0)\rangle = \sum_i U^*_{\alpha i} e^{-iE_i(\mathbf{p})(t-t_0)}|\nu_i(\mathbf{p})\rangle. \tag{6.5}$$

The probability that at time t the state is in flavour β is

$$P(\nu_\alpha \to \nu_\beta)(t) = |\langle \nu_\beta|\nu_\alpha(t)\rangle|^2 = \left|\sum_i U_{\beta i} U^*_{\alpha i} e^{-iE_i(\mathbf{p})(t-t_0)}\right|^2, \tag{6.6}$$

where we have used the orthogonality relation $\langle \nu_i(\mathbf{p})|\nu_j(\mathbf{p})\rangle = \delta_{ij}$.

Since the neutrinos are ultrarelativistic, we can approximate

$$E_i(\mathbf{p}) - E_j(\mathbf{p}) \simeq \frac{1}{2}\frac{m_i^2 - m_j^2}{|\mathbf{p}|} + \mathcal{O}(m^4), \tag{6.7}$$

and $L \simeq (t - t_0)$, so that the master formula in Eq. (6.2) is recovered.

The well-founded criticism to this derivation can be summarized in the following questions: 1) why are all mass eigenstates of equal spatial momentum, $\mathbf{p}$? 2) is the plane wave treatment justified when the production and detection regions are localized? 3) why is it necessary to do the $t - t_0 \to L$ conversion?

A number of quantum paradoxes can be formulated from these questions, that can be resolved only when the two basic conditions for neutrino oscillations above are made explicit. This can be achieved in a wave packet treatment.

6.2 Wave packet derivation

Many authors have derived the master formula treating neutrinos involved as wave packets. For examples, see Refs. [20, 21].

A neutrino of flavour α is produced at time and position $(t_0, \mathbf{x}_0) = (0, \mathbf{0})$ as a superposition of *source* wave packets, $f_i^S(\mathbf{p})$, one for each mass eigenstate. The state at time and position $(t, \mathbf{x})$ is therefore

$$|\nu_\alpha(t,\mathbf{x})\rangle = \sum_i U^*_{\alpha i} \int_{\mathbf{p}} f_i^S(\mathbf{p}) e^{-iE_i(\mathbf{p})t} e^{i\mathbf{p}\mathbf{x}} |\nu_i\rangle. \tag{6.8}$$

For simplicity we will assume Gaussian wave packets, with an average momentum $\mathbf{Q}_i$ and width σ_S:

$$f_i^S(\mathbf{p}) \propto e^{-(\mathbf{p}-\mathbf{Q}_i)^2/2\sigma_S^2}. \tag{6.9}$$

Note that we have lifted the assumption that all mass eigenstates have the same spatial momentum.

A neutrino of flavour β is detected at time and position $(T, \mathbf{L})$ as a superposition of *detector* wave packets, $f_i^D(\mathbf{p})$, created at this space-time position. The state detected is therefore

$$|\nu_\beta(t,\mathbf{x})\rangle = \sum_j U^*_{\beta j} \int_{\mathbf{p}} f_j^D(\mathbf{p}) e^{-iE_j(\mathbf{p})(t-T)} e^{i\mathbf{p}(\mathbf{x}-\mathbf{L})} |\nu_j\rangle, \tag{6.10}$$

where we also assume Gaussian wave packets at detection, with average momentum $\mathbf{Q}'_j$ and width σ_D:

$$f_j^D(\mathbf{p}) \propto e^{-(\mathbf{p}-\mathbf{Q}'_j)^2/2\sigma_D^2}. \tag{6.11}$$

The probability amplitude for the first state to turn into the second is therefore

$$\mathcal{A}(\nu_\alpha \to \nu_\beta) \propto \int d\mathbf{x} \langle \nu_\beta(t,\mathbf{x}) | \nu_\alpha(t,\mathbf{x})\rangle = \sum_i U^*_{\alpha i} U_{\beta i} \int_{\mathbf{p}} e^{-iE_i(\mathbf{p})T} e^{i\mathbf{p}\mathbf{L}} f_i^S(\mathbf{p}) f_i^{D*}(\mathbf{p}) \tag{6.12}$$

For Gaussian wave packets we can rewrite the product of the S and D wave packets as a Gaussian wave packet:

$$f_i^{D*}(\mathbf{p}) f_i^S(\mathbf{p}) \propto f_i^{ov}(\mathbf{p}) e^{-(\mathbf{Q}_i-\mathbf{Q}'_i)^2/4(\sigma_S^2+\sigma_D^2)}, \tag{6.13}$$

where the overlap wave packet

$$f_i^{ov}(\mathbf{p}) \equiv e^{-(\mathbf{p}-\bar{\mathbf{Q}}_i)^2/2\sigma_{\rm ov}^2}, \quad \bar{\mathbf{Q}}_i \equiv \left(\frac{\mathbf{Q}_i}{\sigma_S^2} + \frac{\mathbf{Q}'_i}{\sigma_D^2}\right)\sigma_{\rm ov}^2, \quad \sigma_{\rm ov}^2 \equiv \frac{1}{1/\sigma_S^2 + 1/\sigma_D^2}. \tag{6.14}$$

The momentum integral in Eq. (6.12) can be done analytically if we approximate

$$E_i(\mathbf{p}) \simeq E_i(\bar{\mathbf{Q}}_i) + \sum_k \left.\frac{\partial E_i}{\partial p_k}\right|_{\bar{\mathbf{Q}}_i} (p_k - (\bar{Q}_i)_k) + \ldots = E_i(\bar{\mathbf{Q}}_i) + \mathbf{v}_i(\mathbf{p} - \bar{\mathbf{Q}}_i) + \ldots, \tag{6.15}$$

where $\mathbf{v}_i$ is the overlap wave packet group velocity.

The amplitude obtained is

$$\mathcal{A}(\nu_\alpha \to \nu_\beta) \propto \sum_i U^*_{\alpha i} U_{\beta i} e^{-iE_i(\bar{\mathbf{Q}}_i)T} e^{i\bar{\mathbf{Q}}_i \mathbf{L}} e^{-(\mathbf{Q}_i - \mathbf{Q}'_i)^2/4(\sigma_S^2+\sigma_D^2)} e^{-(\mathbf{L}-\mathbf{v}_i T)^2 \sigma_{\rm ov}^2/2}. \tag{6.16}$$

Note that the two last exponential factors impose momentum conservation (the average momentum of the source and detector wave packets should be equal up to the momentum uncertainty) and the classical relation $\mathbf{L} = \mathbf{v}_i T$ within the spatial uncertainty, $\sigma_{\rm ov}^{-1}$.

Since we usually do not measure the detection time T in a neutrino oscillation experiment, we should integrate the probability over this variable. For simplicity we assume $\mathbf{Q}_i \simeq \mathbf{Q}'_i$ and parallel to $\mathbf{L}$. In this case, the integral gives:

$$\begin{aligned} P(\nu_\alpha \to \nu_\beta) &\propto \int_{-\infty}^{\infty} dT |\mathcal{A}(\nu_\alpha \to \nu_\beta)|^2 \\ &\propto \sum_{i,j} U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta j} e^{-i\frac{\Delta m^2_{ji} L}{2|\mathbf{p}|}} \underbrace{e^{-\left(\frac{L}{L_{\rm coh}(i,j)}\right)^2}}_{\text{coherence}} \underbrace{e^{-\left(\frac{E_i(\bar{\mathbf{Q}}_i)-E_j(\bar{\mathbf{Q}}_j)}{2\sigma_{\rm ov}}\right)^2}}_{\text{momentum uncertainty}} \end{aligned} \tag{6.17}$$

where the coherence length

$$L_{\rm coh}(i,j) \simeq \sigma_{\rm ov} \frac{|\mathbf{v}_i - \mathbf{v}_j|}{\sqrt{\mathbf{v}_i^2 + \mathbf{v}_j^2}}, \tag{6.18}$$

represents the distance travelled by the two wave packets, moving at slightly different group velocities $\mathbf{v}_i$ and $\mathbf{v}_j$, such that the center of the two wave packets have separated spacially a distance of the order of the spatial uncertainty $\sigma_{\rm ov}^{-1}$. For $L \geq L_{\rm coh}(i,j)$ the coherence between the wave packets i, j is lost and the corresponding terms in the oscillation probability exponentially suppressed. The last exponential factor in Eq. (6.17) leads to a suppression of the oscillation probability when the difference in average energies of the two wave packets i, j is larger than the momentum uncertainty of the overlap wave packet, $\sigma_{\rm ov}$. Note that $\sigma_{\rm ov}$ is dominated by the smallest of the production and detection uncertainties, and therefore both should be large enough to ensure that the wave packets of the different mass eigenstates remain coherent. To the extent that $L \ll L_{\rm coh}$ and $|E_i - E_j| \ll {\rm Min}(\sigma_S, \sigma_D)$, the probability reduces to the master formula, with one caveat: we have lost the normalization along the way. This is usually unavoidable in the wave packet derivation. The right normalization can be imposed only a posteriori, for example, from unitarity, $\sum_\beta P(\nu_\alpha \to \nu_\beta) = 1$.

In summary, the wave packet derivation is clearly more physical, as it makes explicit the two necessary conditions for neutrino oscillations to take place: coherence and sufficient momentum uncertainty.

6.3 QFT derivation

Since we are dealing with relativistic quantum mechanics, QFT should be the appropriate framework to derive the oscillation probability.

In QFT we consider scattering processes where some asymptotic *in-states* that we can prepare

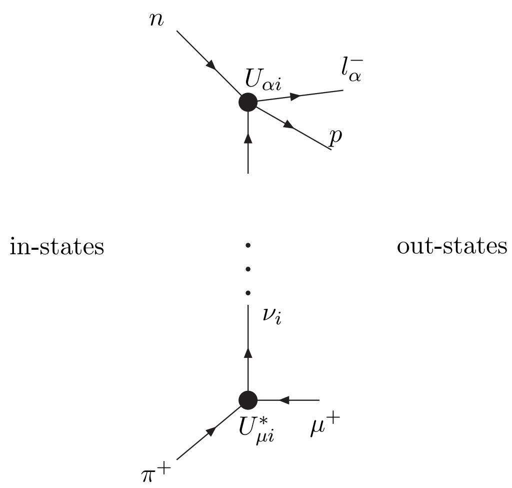


Fig. 17: Neutrino oscillations in QFT.

in the infinite past come close together at some finite time in an interaction region and scatter off into other asymptotic *out-states* at time $t \to \infty$. The probability amplitude for this process is just the scalar product of the in and out states. In computing this amplitude we usually idealise the asymptotic states as plane waves, which is a good approximation provided the interaction region is small compared to the Compton wavelength of the scattering states. In reality however the proper normalization of the scattering probability as a probability per unit time and volume requires that the initial states are normalized wave packets.

In a neutrino oscillation experiment, the asymptotic states are not the neutrinos, we cannot really prepare the neutrino states, but the particles that produce the neutrino at the source and those that interact with the neutrino in the detector. The neutrino is just a virtual particle being exchanged between the source and detector, see Fig. 17, and in this perspective the interaction region is as large as the baseline and therefore macroscopic, in particular much larger than the Compton wavelength of the asymptotic states involved. It is mandatory therefore to consider the in-states as wave packets to ensure the localization of the source and detector.

Consider for example a neutrino beam produced from pions at rest and a detector some distance apart, where neutrinos interact with nucleons that are also at rest, via a quasi-elastic event:

$$\pi n \to p \mu l_\beta. \tag{6.19}$$

The in-states therefore will be the two wave packets representing a static pion that decays and is localized at time and position $(0, \mathbf{0})$ within the uncertainty better defined than the decay tunnel, and a nucleon that is static and localized within the detector, at time and position $(T, \mathbf{L})$, when the interaction takes place. The out-states are the muon produced in pion decay and the lepton and hadron produced in the quasi-elastic event. The probability amplitude for the whole process includes the pion decay amplitude, the neutrino propagation and the scattering amplitude at the detector. Therefore in order to extract from the full amplitude an oscillation probability, it must be the case that there is factorization of the whole probability into three factors that can be identified with the flux of neutrino from pion decay, an oscillation probability and a neutrino cross section.

By explicit calculation [22], it is possible to show that such factorization does indeed take place as long as kinematical effects of neutrino masses can be neglected. The oscillation probability defined as the ratio of the probability for the whole process and the product of the neutrino flux from pion decay and the neutrino scattering cross-section is properly normalized.

6.4 Neutrino oscillations in vacuum

Let us analyse more closely the master formula Eq. (6.2). The probability is a superposition of oscillatory functions of the baseline with wavelengths that depend on the neutrino mass differences $\Delta m^2_{ij} = m^2_j - m^2_i$, and amplitudes that depend on different combinations of the mixing matrix elements. Defining $W^{ij}_{\alpha\beta} \equiv [U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j}]$ and using the unitarity of the mixing matrix, we can rewrite the probability in the more familiar form:

$$\begin{aligned} P(\nu_\alpha \;\; \rightarrow \;\; \nu_\beta) &= \delta_{\alpha\beta} - 4 \sum_{j>i} \mathrm{Re}[\mathrm{W}^{\mathrm{ij}}_{\alpha\beta}] \sin^2\left(\frac{\Delta \mathrm{m}^2_{\mathrm{ij}}\,\mathrm{L}}{4\mathrm{E}_\nu}\right) \\ &\mp \;\; 2 \sum_{j>i} \mathrm{Im}[\mathrm{W}^{\mathrm{ij}}_{\alpha\beta}] \sin\left(\frac{\Delta \mathrm{m}^2_{\mathrm{ij}}\,\mathrm{L}}{2\mathrm{E}_\nu}\right) , \end{aligned} \tag{6.20}$$

where the $\mp$ refers to neutrinos/antineutrinos and $|\mathbf{p}| \simeq E_\nu$.

We refer to an *appearance* or *disappearance* oscillation probability when the initial and final flavours are different ($\alpha \neq \beta$) or the same ($\alpha = \beta$), respectively. Note that oscillation probabilities show the expected GIM suppression of any flavour changing process: they vanish if the neutrinos are degenerate.

In the simplest case of two-family mixing, the mixing matrix depends on just one mixing angle:

$$U_{\mathrm{PMNS}} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} , \tag{6.21}$$

and there is only one mass square difference Δm^2. The oscillation probability of Eq. (6.20) simplifies to the well-known expression where we have introduced convenient physical units:

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta) &= \sin^2 2\theta \, \sin^2\left(1.27 \frac{\Delta m^2(\mathrm{eV}^2)\, L(\mathrm{km})}{E_\nu(\mathrm{GeV})}\right), \quad \alpha \neq \beta \, . \\ P(\nu_\alpha \rightarrow \nu_\alpha) &= 1 - P(\nu_\alpha \rightarrow \nu_\beta). \end{aligned} \tag{6.22}$$

The probability is the same for neutrinos and antineutrinos, because there cannot be CP violation when there are only two families. Indeed CPT implies that the disappearance probabilities are the same for neutrinos and antineutrinos, and therefore according to Eq. (6.22) the same must hold for the appearance probability. The latter is a sinusoidal function of the distance between source and detector, with a period determined by the oscillation length:

$$L_{\mathrm{osc}}\,(\mathrm{km}) = \pi \frac{E_\nu(\mathrm{GeV})}{1.27 \Delta m^2(\mathrm{eV}^2)} , \tag{6.23}$$

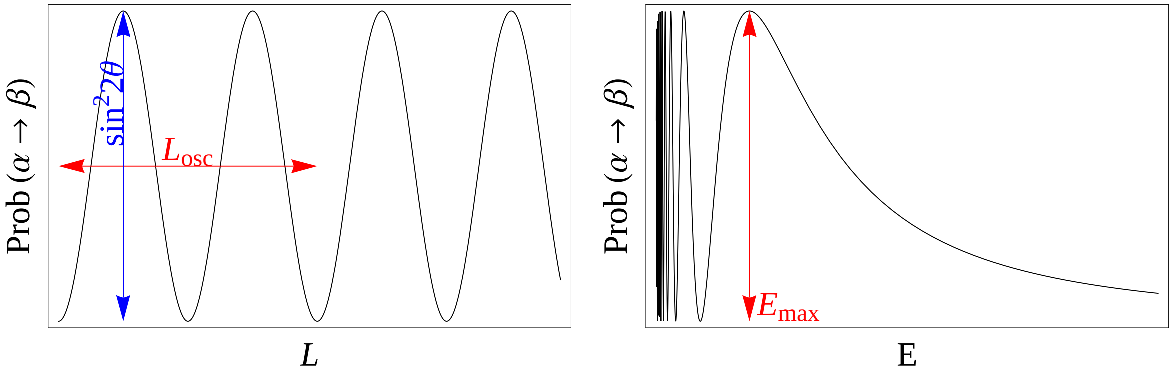


Fig. 18: Left: two-family appearance oscillation probability as a function of the baseline of L at fixed neutrino energy. Right: same probability shown as a function of the neutrino energy for fixed baseline.

which is proportional to the neutrino energy and inversely proportional to the neutrino mass square difference. The amplitude of the oscillation is determined by the mixing angle. It is maximal for $\sin^2 2\theta = 1$ or $\theta = \pi/4$. The oscillation probability as a function of the baseline is shown on the left plot of Fig. 18.

In many neutrino oscillation experiments the baseline is not varied but the oscillation probability can be measured as a function of the neutrino energy. This is shown on the right plot of Fig. 18. In this case, the position of the first maximum contains information on the mass splitting:

$$E_{\rm max}({\rm GeV}) = 1.27\frac{\Delta m^2({\rm eV}^2)L({\rm km})}{\pi/2}. \tag{6.24}$$

An optimal neutrino oscillation experiment in vacuum is such that the ratio of the neutrino energy and baseline are tuned to be of the same order as the mass splitting, $E/L \sim \Delta m^2$. If $E/L \gg \Delta m^2$, the oscillation phase is small and the oscillation probability is approximately $P(\nu_\alpha \to \nu_\beta) \propto \sin^2 2\theta (\Delta m^2)^2$, so the mixing angle and mass splitting cannot be disentangled. The opposite limit $E/L \ll \Delta m^2$ is the fast oscillation regime, where one can only measure an energy or baseline-smeared oscillation probability

$$\langle P(\nu_\alpha \to \nu_\beta)\rangle \simeq \frac{1}{2}\sin^2 2\theta, \tag{6.25}$$

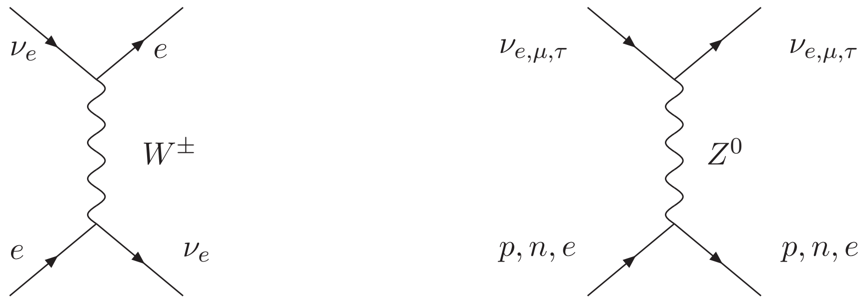
sensitivity to the mass splitting is lost in this limit. It is interesting, and reassuring, to note that this averaged oscillation regime gives the same result as the flavour transition probability in the case of incoherent propagation ($L \gg L_{\rm coh}$):

$$P\left(\nu_\alpha \to \nu_\beta\right) = \sum_i |U_{\alpha i}U_{\beta i}|^2 = 2\cos^2\theta\sin^2\theta = \frac{1}{2}\sin^2 2\theta. \tag{6.26}$$

Flavour transitions via incoherent propagation are sensitive to mixing but not to the neutrino mass splitting. The smoking gun for neutrino oscillations is not the flavour transition, which can occur in the presence of neutrino mixing without oscillations, but the peculiar L/E_ν dependence. An optimal experiment that intends to measure both the mixing and the mass splitting requires running $E/L \sim \Delta m^2$.

6.5 Neutrino propagation in matter

When neutrinos propagate in matter (earth, sun, etc.), their propagation is modified owing to coherent forward scattering on electrons and nucleons [23]:



The effective Hamiltonian density resulting from the charged current interaction is

$$\mathcal{H}_{\rm CC} = 2\sqrt{2}G_F\,[\bar{e}\gamma_\mu P_L \nu_e][\bar{\nu}_e\gamma^\mu P_L e] = 2\sqrt{2}G_F\,[\bar{e}\gamma_\mu P_L e][\bar{\nu}_e\gamma^\mu P_L \nu_e]. \tag{6.27}$$

Since the medium is not polarized, the expectation value of the electron current is simply the number density of electrons:

$$\langle \bar{e}\gamma_\mu P_L e\rangle_{\rm unpol.medium} \;=\; \delta_{\mu 0}\frac{N_e}{2}. \tag{6.28}$$

Including also the neutral current interactions in the same way, the effective Hamiltonian for neutrinos in the presence of matter is

$$\langle \mathcal{H}_{CC} + \mathcal{H}_{NC}\rangle_{\rm medium} = \bar{\nu} V_m \gamma^0 (1-\gamma_5)\nu \tag{6.29}$$

$$V_m = \begin{pmatrix} \frac{G_F}{\sqrt{2}}\left(N_e - \frac{N_n}{2}\right) & 0 & 0 \\ 0 & \frac{G_F}{\sqrt{2}}\left(-\frac{N_n}{2}\right) & 0 \\ 0 & 0 & \frac{G_F}{\sqrt{2}}\left(-\frac{N_n}{2}\right) \end{pmatrix}, \tag{6.30}$$

where N_n is the number density of neutrons. Due to the neutrality of matter, the proton and electron contributions to the neutral current potential cancel.

The plane wave solutions to the modified Dirac equation satisfy a different dispersion relation

$$E^2 = |\mathbf{p}|^2 + M_\nu^2 \pm 4EV_m, \tag{6.31}$$

where $\pm$ is for neutrinos/antineutrinos. The phases of neutrino oscillation phenomena change.

The effect of matter can be simply accommodated in an effective mass matrix:

$$\tilde{M}_\nu^2 = M_\nu^2 \pm 4EV_m. \tag{6.32}$$

The effective mixing matrix $\tilde{V}_{\rm MNS}$ is the one that takes us from the original flavour basis to that which diagonalizes this effective mass matrix:

$$\begin{pmatrix} \tilde{m}_1^2 & 0 & 0 \\ 0 & \tilde{m}_2^2 & 0 \\ 0 & 0 & \tilde{m}_3^2 \end{pmatrix} = \tilde{V}_{\rm MNS}^\dagger \left(M_\nu^2 \pm 4E \begin{pmatrix} V_e & 0 & 0 \\ 0 & V_\mu & 0 \\ 0 & 0 & V_\tau \end{pmatrix} \right) \tilde{V}_{\rm MNS}. \tag{6.33}$$

The effective mixing angles and masses depend on the energy.

The matter potential in the center of the sun is $V_m \sim 10^{-12}$ eV and in the earth $V_m \sim 10^{-13}$ eV. In spite of these tiny values, these effects are non-negligible in neutrino oscillations.

6.6 Neutrino oscillations in constant matter

In the case of two flavours, the effective mass and mixing angle have relatively simple expressions:

$$\Delta\tilde{m}^2 = \sqrt{\left(\Delta m^2 \cos 2\theta \mp 2\sqrt{2} E\, G_F\, N_e\right)^2 + (\Delta m^2 \sin 2\theta)^2}, \tag{6.34}$$

$$\sin^2 2\tilde{\theta} = \frac{\left(\Delta m^2 \sin 2\theta\right)^2}{(\Delta\tilde{m}^2)^2}, \tag{6.35}$$

where the sign $\mp$ corresponds to neutrinos/antineutrinos. The corresponding oscillation amplitude has a resonance [24], when the neutrino energy satisfies

$$\sqrt{2}\, G_F\, N_e \mp \frac{\Delta m^2}{2E} \cos 2\theta = 0 \quad \Rightarrow \quad \sin^2 2\tilde{\theta} = 1, \quad \Delta\tilde{m}^2 = \Delta m^2 \sin 2\theta. \tag{6.36}$$

The oscillation amplitude is therefore maximal, independently of the value of the vacuum mixing angle.

We also note that

- oscillations vanish at $\theta = 0$, because the oscillation length becomes infinite for $\theta = 0$;
- the resonance is only there for ν or $\bar{\nu}$ but not both;
- the resonance condition depends on the sign$(\Delta m^2 \cos 2\theta)$:

resonance observed in $\nu \to$ sign$(\Delta m^2 \cos 2\theta) > 0$,
resonance observed in $\bar{\nu} \to$ sign$(\Delta m^2 \cos 2\theta) < 0$.

The origin of this resonance is a would-be level crossing in the case of vanishing mixing. In the case of two families, for $\theta = 0$, the mass eigenstates as a function of the electron number density, at fixed neutrino energy, are depicted in Fig. 19 for $\Delta m^2 > 0$. As soon as the mixing is lifted from zero, no matter how small, the crossing cannot take place. The resonance condition corresponds to the minimum level-splitting point.

6.7 Neutrino oscillations in variable matter

In the sun the density of electrons is not constant. However, if the variation is sufficiently slow, the eigenstates will change slowly with the density, and we can assume that the neutrino produced in an

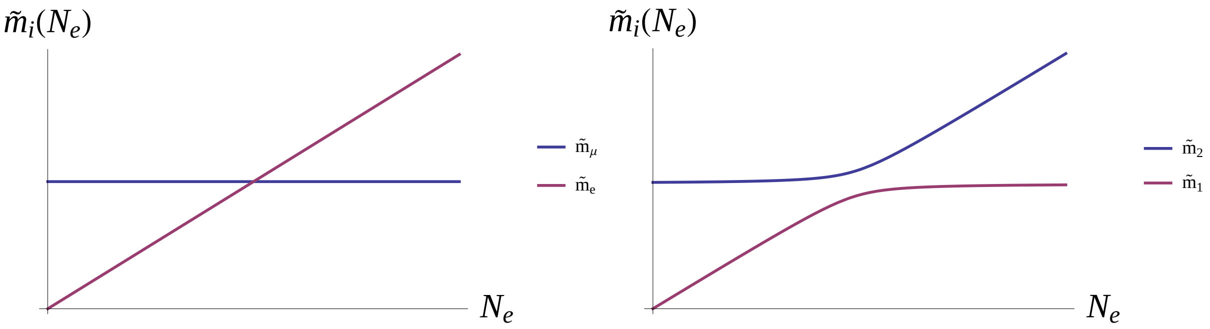


Fig. 19: Mass eigenstates as a function of the electron number density at fixed neutrino energy for $\theta = 0$ (left) and $\theta \neq 0$ (right).

eigenstate in the center of the sun, remains in the same eigenstate along the trajectory. This is the so-called *adiabatic approximation*.

We consider here two-family mixing for simplicity. At any point in the trajectory, it is possible to diagonalize the Hamiltonian fixing the matter density to that at the given point. The resulting eigenstates can be written as

$$|\tilde{\nu}_1\rangle = |\nu_e\rangle \cos\tilde{\theta} - |\nu_\mu\rangle \sin\tilde{\theta}, \tag{6.37}$$

$$|\tilde{\nu}_2\rangle = |\nu_e\rangle \sin\tilde{\theta} + |\nu_\mu\rangle \cos\tilde{\theta}. \tag{6.38}$$

Neutrinos are produced close to the centre $x = 0$ where the electron density is $N_e(0)$. Let us suppose that it satisfies

$$2\sqrt{2}G_F N_e(0) \gg \Delta m^2 \cos 2\theta. \tag{6.39}$$

Then the diagonalization of the mass matrix at this point gives

$$\tilde{\theta} \simeq \frac{\pi}{2} \Rightarrow |\nu_e\rangle \simeq |\tilde{\nu}_2\rangle, \tag{6.40}$$

in such a way that an electron neutrino is mostly the second mass eigenstate. When neutrinos exit the sun, at $x = R_\odot$, the matter density falls to zero, $N_e(R_\odot) = 0$, and the local effective mixing angle is the one in vacuum, $\tilde{\theta} = \theta$. If θ is small, the eigenstate $\tilde{\nu}_2$ is mostly ν_μ according to Eq. (6.38).

Therefore an electron neutrino produced at $x = 0$ is mostly the eigenstate $\tilde{\nu}_2$, but this eigenstate outside the sun is mostly ν_μ. There is maximal $\nu_e \to \nu_\mu$ conversion if the adiabatic approximation is a good one. This is the famous MSW effect [23, 24]. The conditions for this to happen are:

– *Resonant condition*: the density at the production is above the critical one

$$N_e(0) > \frac{\Delta m^2 \cos 2\theta}{2\sqrt{2}EG_F}. \tag{6.41}$$

– *Adiabaticity*: the splitting of the levels is large compared to energy injected in the system by the

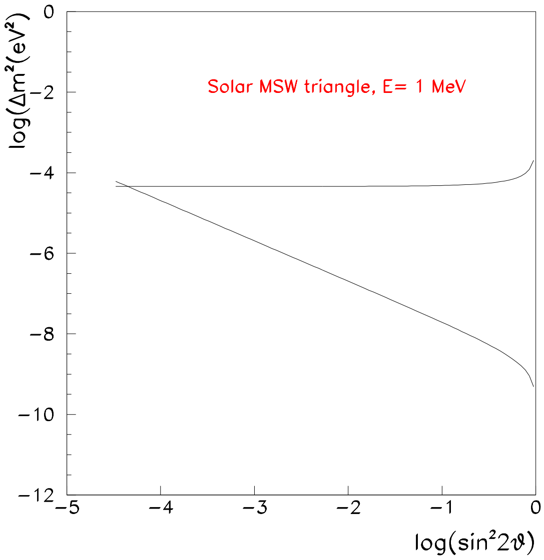


Fig. 20: MSW triangle: in the region between the two lines the resonance and adiabaticity conditions are both satisfied for neutrinos of energy 1 MeV.

variation of $N_e(r)$. A measurement of this is given by γ which should be much larger than one:

$$\gamma = \frac{\sin^2 2\theta}{\cos 2\theta} \frac{\Delta m^2}{2E} \frac{1}{|\nabla \log N_e(r)|} > \gamma_{\min} > 1, \tag{6.42}$$

where $\nabla = \partial/\partial r$.

At fixed energy both conditions give the famous MSW triangles, if plotted on the plane $(\log(\sin^2 2\theta), \log(\Delta m^2))$:

$$\log(\Delta m^2) < \log\left(\frac{2\sqrt{2} G_F N_e(0) E}{\cos 2\theta}\right) \tag{6.43}$$

$$\log(\Delta m^2) > \log\left(\gamma_{\min} 2E \nabla \log N_e \frac{\cos 2\theta}{\sin^2 2\theta}\right). \tag{6.44}$$

For example, taking $N_e(r) = N_c \exp(-r/R_0), R_0 = R_\odot/10.54, N_c = 1.6 \times 10^{26}$ cm^{-3}, $E = 1$ MeV, these curves are shown in Fig. 20.

It should be stressed that neutrino oscillations are not responsible for the flavour transition of solar neutrinos. The survival probability of the solar ν_e in the adiabatic approximation is the incoherent sum of the contribution of each of the mass eigenstates:

$$P(\nu_e \to \nu_e) = \sum_i |\langle \nu_e | \tilde{\nu}_i(R_\odot)\rangle|^2 |\langle \tilde{\nu}_i(0)|\nu_e\rangle|^2, \tag{6.45}$$

where $\tilde{\nu}_i(r)$ is the i-th mass eigenstate for the electron number density, $N_e(r)$, at a distance r from the center of the sun. If the mass eigenstates contribute incoherently, how can we measure the neutrino mass splitting? The answer is that the resonance condition of Eq. (6.41) depends on the neutrino energy. If we

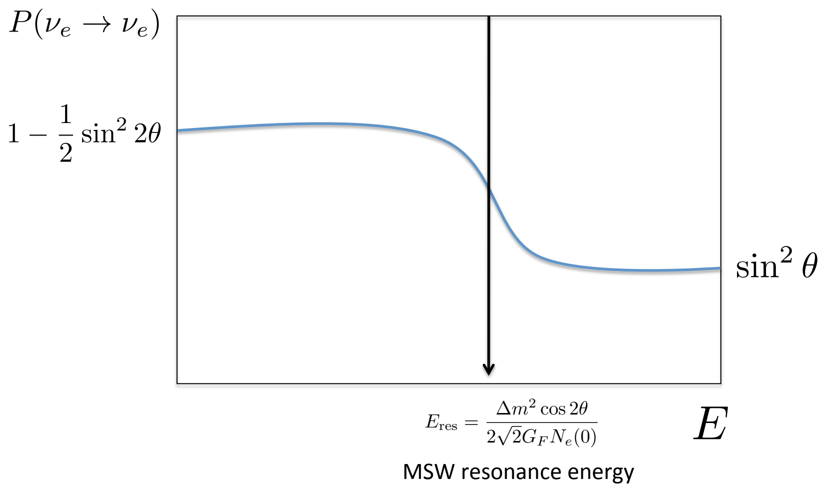


Fig. 21: Schematic survival probability of solar neutrinos as a function of the energy.

define

$$E_{\rm res} \equiv \frac{\Delta m^2 \cos 2\theta}{2\sqrt{2}G_F N_e(0)}, \tag{6.46}$$

the MSW effect will affect neutrinos with $E > E_{\rm res}$, while for $E < E_{\rm res}$, the oscillation probability is close to that in vacuum for averaged oscillations. The spectrum of the solar neutrino flux includes energies both above and below $E_{\rm res}$:

$$\begin{aligned} P(\nu_e \to \nu_e) &\simeq 1 - \tfrac{1}{2}\sin^2 2\theta, \quad E \ll E_{\rm res} \\ P(\nu_e \to \nu_e) &\simeq \sin^2\theta, \qquad\quad E \gg E_{\rm res} \end{aligned} \tag{6.47}$$

The sensitivity to Δm^2 relies on the ability to locate the resonant energy. This behaviour is schematically depicted in Fig. 21.

7 Evidence for neutrino oscillations

Nature has been kind enough to provide us with two natural sources of neutrinos (the sun and the atmosphere) where neutrino flavour transitions have been observed in a series of ingenious experiments, that started back in the 1960s with the pioneering experiment of R. Davies. This effort was rewarded with the Nobel prize of 2002 to R. Davies and M. Koshiba *for the detection of cosmic neutrinos*.

7.1 Solar neutrinos

The sun, like all stars, is an intense source of neutrinos produced in the chain of nuclear reactions that burn hydrogen into helium:

$$4p \longrightarrow {}^4\text{He} + 2e^+ + 2\nu_e. \tag{7.1}$$

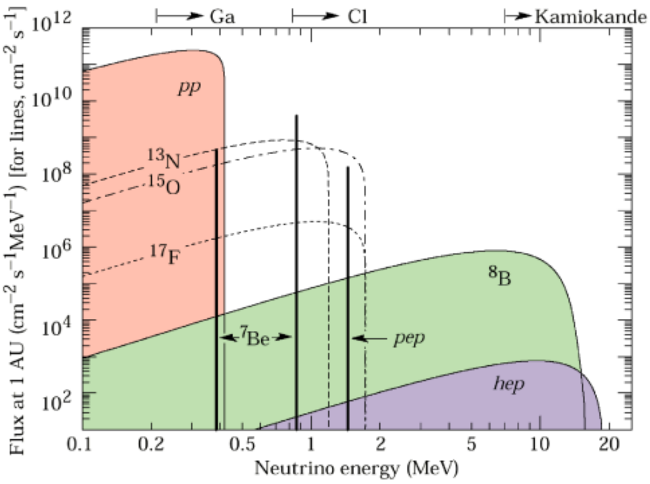


Fig. 22: Spectrum of solar neutrinos [26]. The arrows indicate the threshold of the different detection techniques.

The theory of stellar nucleosynthesis was established at the end of the 30's by H. Bethe [25]. The spectrum of the solar ν_e, for massless neutrinos, is shown in Fig. 22. The prediction of this flux, obtained by J. Bahcall and collaborators [26], is the result of a detailed simulation of the solar interior and has been improved over many years. It is the so-called standard solar model (SSM).

Neutrinos coming from the sun have been detected with several experimental techniques that have a different neutrino energy threshold as indicated in Fig. 22. On the one hand, the radiochemical techniques, used in the experiments Homestake (chlorine, ^{37}Cl) [27], Gallex/GNO [28] and Sage [29] (using gallium, ^{71}Ga, and germanium, ^{71}Ge, respectively), can count the total number of neutrinos with a rather low threshold ($E_\nu > 0.81$ MeV in Homestake and $E_\nu > 0.23$ MeV in Gallex and Sage), but they cannot get any information on the directionality, the energy of the neutrinos, nor the time of the event.

On the other hand, Kamiokande [30] pioneered a new technique to observe solar neutrinos using water Cherenkov detectors that can measure the recoil electron in elastic neutrino scattering on electrons: $\nu_e \ + \ e^- \ \rightarrow \ \nu_e \ + e^-$. This is a real-time experiment that provides information on the directionality and the energy of the neutrinos. The threshold on the other hand is much higher, ~ 5 MeV. All these experiments have consistently observed a number of solar neutrinos between 1/3 and 1/2 of the number expected in the SSM and for a long time this was referred to as the *solar neutrino problem or deficit*.

The progress in this field over the last two decades has been enormous culminating in a solution to this puzzle that no longer relies on the predictions of the SSM. There have been three milestones.

1998: The experiment Super-Kamiokande [31] measured the solar neutrino deficit with unprecedented precision, using the elastic reaction (ES):

$$\text{(ES)} \qquad \nu_e \ + \ e^- \ \rightarrow \ \nu_e \ + e^- \qquad E_{\text{thres}} > 5 \text{ MeV}. \tag{7.2}$$

The measurement of the direction of the events demonstrated that the neutrinos measured definitely come from the sun: the left plot of Fig. 23 shows the distribution of the events as a function of the zenith angle

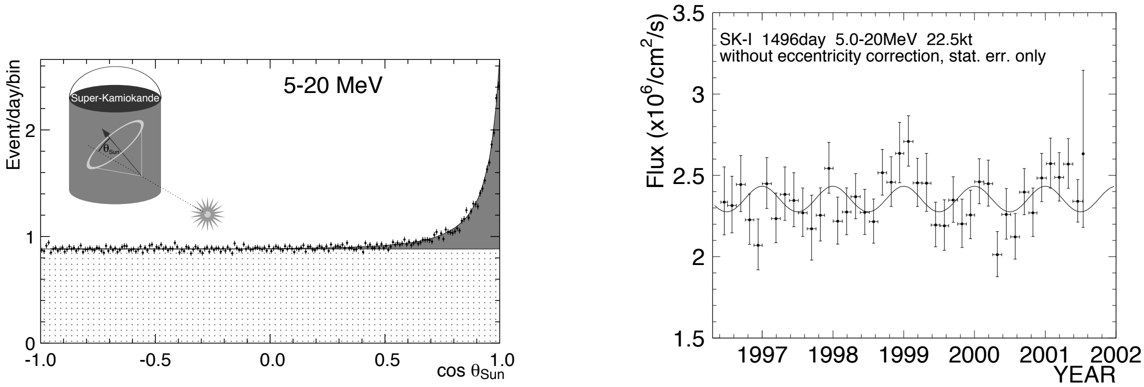


Fig. 23: Left: distribution of solar neutrino events as a function of the zenith angle of the sun. Right: seasonal variation of the solar neutrino flux in Super-Kamiokande (from Ref. ([32])).

of the sun. A seasonal variation of the flux is expected since the distance between the earth and the sun varies seasonally. The right plot of Fig. 23 shows that the measured variation is in perfect agreement with that expectation.

2001: The SNO experiment [33, 34] measured the flux of solar neutrinos using also the two reactions:

$$\text{(CC)} \qquad \nu_e + d \to p + p + e^- \qquad\qquad E_{\text{thres}} > 5\ \text{MeV} \tag{7.3}$$

$$\text{(NC)} \qquad \nu_x + d \to p + n + \nu_x \quad x = e, \mu, \tau \qquad E_{\text{thres}} > 2.2\ \text{MeV} \tag{7.4}$$

Since the CC reaction is only sensitive to electron neutrinos, while the NC one is sensitive to all the types that couple to the Z^0 boson, the comparison of the fluxes measured with both reactions can establish if there are ν_μ and ν_τ in the solar flux independently of the normalization given by the SSM. The result is shown on the Nobel-prize-winning plot Fig. 24. These measurements demonstrate that the sun shines (ν_μ, ν_τ) about twice more than it shines ν_e, which constitutes the first direct demonstration of flavour transitions in the solar flux! Furthermore the NC flux that measures all active species in the solar flux, is compatible with the total ν_e flux expected according to the SSM.

All solar neutrino data can be interpreted in terms of neutrino masses and mixings. The solar ν_e deficit can be explained for a $\Delta m^2_{\text{solar}} \simeq 7$–$8 \times 10^{-5}$eV and a relatively large mixing angle. The fortunate circumstance that

$$\Delta m^2_{\text{solar}} \sim \langle E_\nu(1\ \text{MeV})\rangle / L(100\ \text{km}) \tag{7.5}$$

implies that one could look for this oscillation measuring reactor neutrinos at baselines of ~ 100 km. This was the third milestone.

2002: The solar oscillation is confirmed with reactor neutrinos in the KamLAND experiment [35]. This has 1 kilo ton of liquid scintillator which measures the flux of reactor neutrinos produced in a cluster of nuclear plants around the Kamioka mine in Japan. The average distance is $\langle L \rangle = 175$ km. Neutrinos

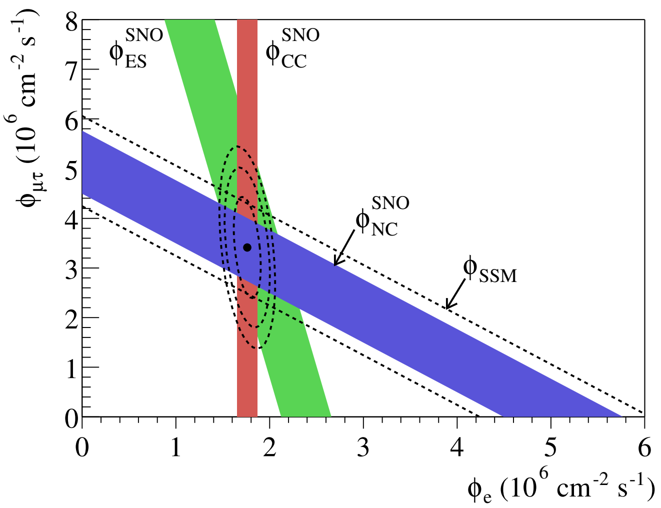


Fig. 24: Flux of ν_μ and ν_τ versus the flux of ν_e in the solar neutrino flux as measured from the three reactions observable in the SNO experiment. The dashed band shows the prediction of the SSM, which agrees perfectly with the flux measured with the NC reaction (from Ref. [34]).

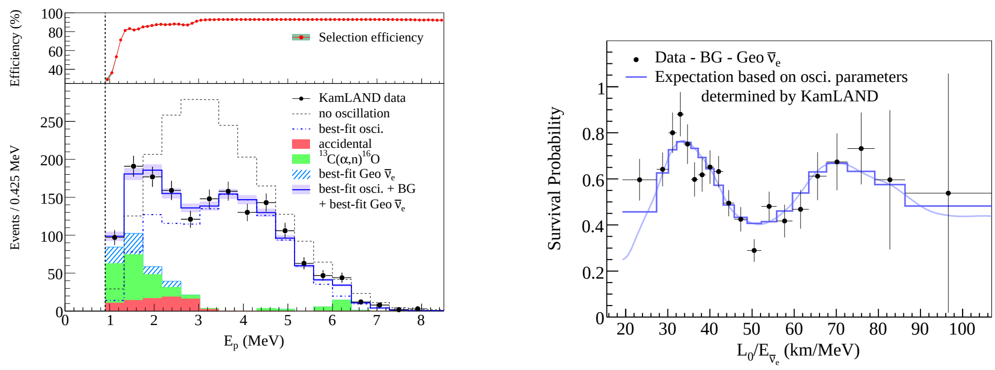


Fig. 25: Spectral distribution of the $\bar{\nu}_e$ events in KamLAND (left) and E_ν/L dependence (right). The data are compared to the expectation in the absence of oscillations and to the best fit oscillation hypothesis (from Ref. [36]).

are detected via inverse β-decay which has a threshold energy of about 2.6 MeV:

$$\bar{\nu}_e + p \to e^+ + n \qquad E_{\rm th} > 2.6\ \text{MeV}\,. \tag{7.6}$$

Figure 25 shows the KamLAND results [36] on the antineutrino spectrum, as well as the survival probability as a function of the ratio E_ν/L.

The low-energy contribution of geo-neutrinos is clearly visible. This measurement could have important implications in geophysics.

Concerning the sensitivity to the oscillation parameters, Fig. 26 shows the present determination of the solar oscillation parameters from KamLAND and other solar experiments. The precision in the determination of $\Delta m^2_{\rm solar}$ is spectacular and shows that solar neutrino experiments are entering the era of

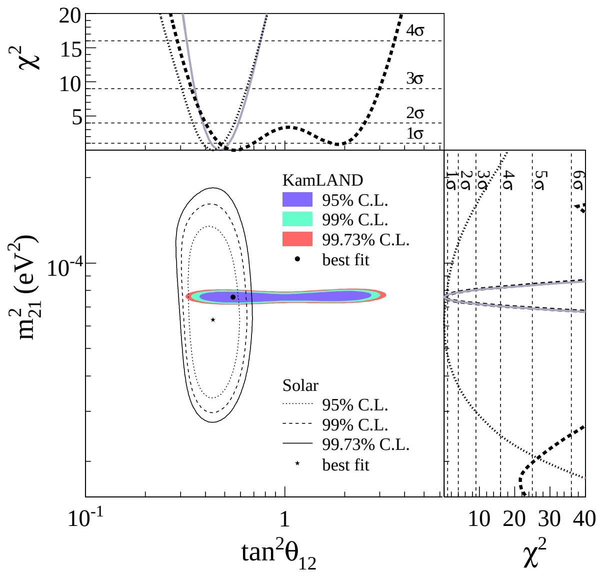


Fig. 26: Analysis of all solar and KamLAND data in terms of oscillations (from Ref. [36]).

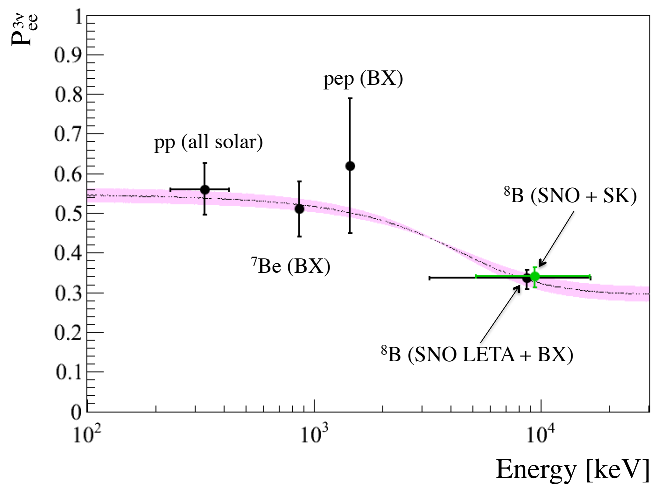


Fig. 27: Comparison of solar neutrino fluxes measured by the different solar neutrino experiments (from Ref. [37]).

precision physics.

The last addition to this success story is the Borexino experiment [37]. This is the lowest-threshold real-time solar neutrino experiment and the only one capable of measuring the flux of the monochromatic ^{7}Be neutrinos and pep neutrinos. Their recent results are shown in Fig. 27. The result is in agreement with the oscillation interpretation of other solar and reactor experiments and it adds further information to disfavour alternative exotic interpretations of the data.

In summary, solar neutrinos experiments have made fundamental discoveries in particle physics and are now becoming useful for other applications, such as a precise understanding of the sun and the

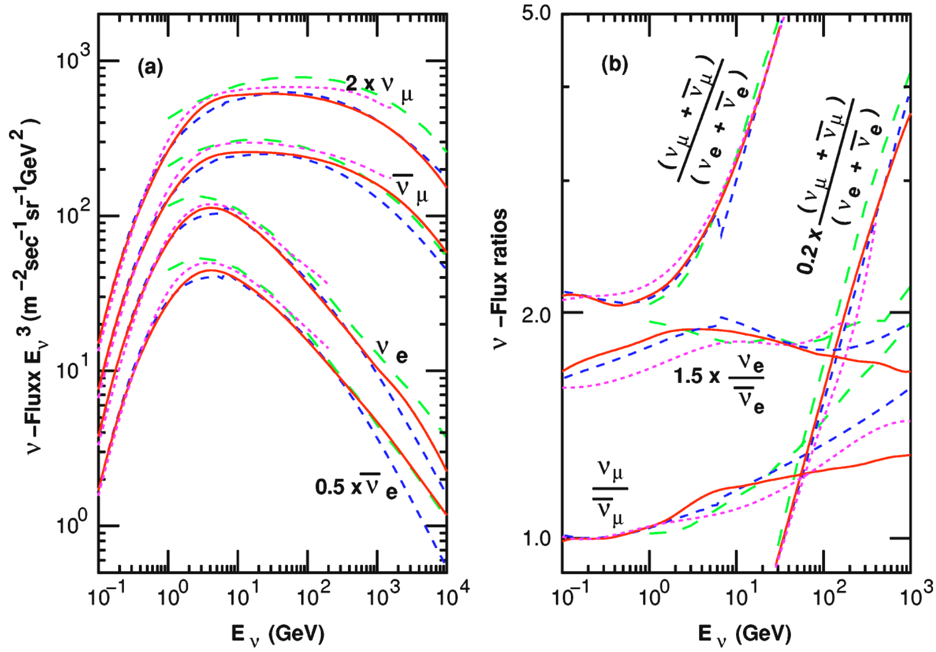


Fig. 28: Comparison of the predictions of different Monte Carlo simulations of the atmospheric neutrino fluxes averaged over all directions (left) and of the flux ratios $(\nu_\mu + \bar{\nu}_\mu)/(\nu_e + \bar{\nu}_e)$, $\nu_\mu/\bar{\nu}_\mu$, and $\nu_e/\bar{\nu}_e$ (right). The solid line corresponds to a recent full 3D simulation. Taken from the last reference in Ref. [38].

earth.

7.2 Atmospheric neutrinos

Neutrinos are also produced in the atmosphere when primary cosmic rays impinge on it producing K, π that subsequently decay. The fluxes of such neutrinos can be predicted within a 10–20$\%$ accuracy to be those in the left plot of Fig. 28.

Clearly, atmospheric neutrinos are an ideal place to look for neutrino oscillation since the E_ν/L span several orders of magnitude, with neutrino energies ranging from a few hundred MeV to 10^3 GeV and distances between production and detection varying from 10–10^4 km, as shown in Fig. 29 (right).

Many of the uncertainties in the predicted fluxes cancel when the ratio of muon to electron events is considered. The first indication of a problem was found when a deficit was observed precisely in this ratio by several experiments: Kamiokande, IMB, Soudan2 and Macro.

In 1998, Super-Kamiokande clarified the origin of this anomaly [39]. This experiment can distinguish muon and electron events, measure the direction of the outgoing lepton (the zenith angle with respect to the earth's axis) which is correlated to that of the neutrino (the higher the energy the higher the correlation), in such a way that they could measure the variation of the flux as a function of the distance travelled by the neutrinos. Furthermore, they considered different samples of events: sub-GeV (lepton with energy below 1 GeV), multi-GeV (lepton with energy above 1 GeV), together with stopping and through-going muons that are produced on the rock surrounding Super-Kamiokande. The different

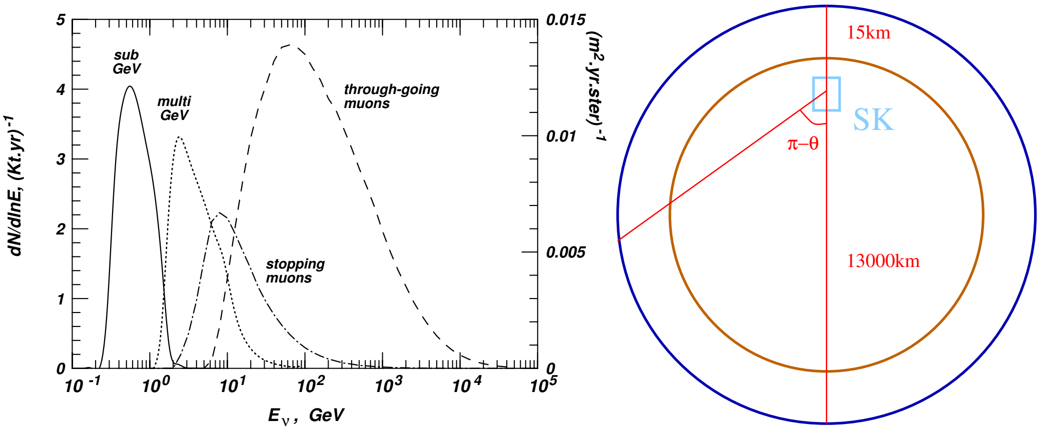


Fig. 29: Left: Parent neutrino energies of the different samples considered in Super-Kamiokande: sub-GeV, multi-GeV, stopping and through-going muons. Right: Distances travelled by atmospheric neutrinos as a function of the zenith angle.

samples correspond to different parent neutrino energies as can be seen in Fig. 29 (left).

The number of events for the different samples as a function of the zenith angle of the lepton are shown in the Nobel-prize-winning plot Fig. 30.

While the electron events observed are in rough agreement with predictions, a large deficit of muon events was found with a strong dependence on the zenith angle: the deficit was almost 50% for those events corresponding to neutrinos coming from below $\cos\theta = -1$, while there is no deficit for those coming from above. The perfect fit to the oscillation hypothesis is rather non-trivial given the sensitivity of this measurement to the E_ν (different samples) and L (zenith angle) dependence. The significance of the E_ν/L dependence has also been measured by the Super-Kamiokande Collaboration [41], as shown in Fig. 31. The best fit value of the oscillation parameters indicate $\Delta m^2 \simeq 3\times 10^{-3}$ eV^2 and maximal mixing.

Appropriate neutrino beams to search for the atmospheric oscillation can easily be produced at accelerators if the detector is located at a long baseline of a few hundred kilometres, and also with reactor neutrinos in a baseline of $\mathcal{O}$(1km), since

$$|\Delta m^2_{\text{atmos}}| \sim \frac{E_\nu(1-10\text{ GeV})}{L(10^2-10^3\text{ km})} \sim \frac{E_\nu(1-10\text{ MeV})}{L(0.1-1\text{ km})}. \tag{7.7}$$

A *conventional* accelerator neutrino beam, as the one used in the LSS experiment, is produced from protons hitting a target and producing π and K:

$$p \;\rightarrow\; \text{Target} \rightarrow \pi^+, K^+ \rightarrow \;\nu_\mu(\%\nu_e, \bar{\nu}_\mu, \bar{\nu}_e) \tag{7.8}$$
$$\nu_\mu \rightarrow \nu_x. \tag{7.9}$$

Those of a selected charge are focused and are left to decay in a long decay tunnel producing a neutrino beam of mostly muon neutrinos (or antineutrinos) with a contamination of electron neutrinos of a few per cent. The atmospheric oscillation can be established by studying, as a function of the energy, either the disappearance of muon neutrinos, the appearance of electron neutrinos or, if the energy of the beam

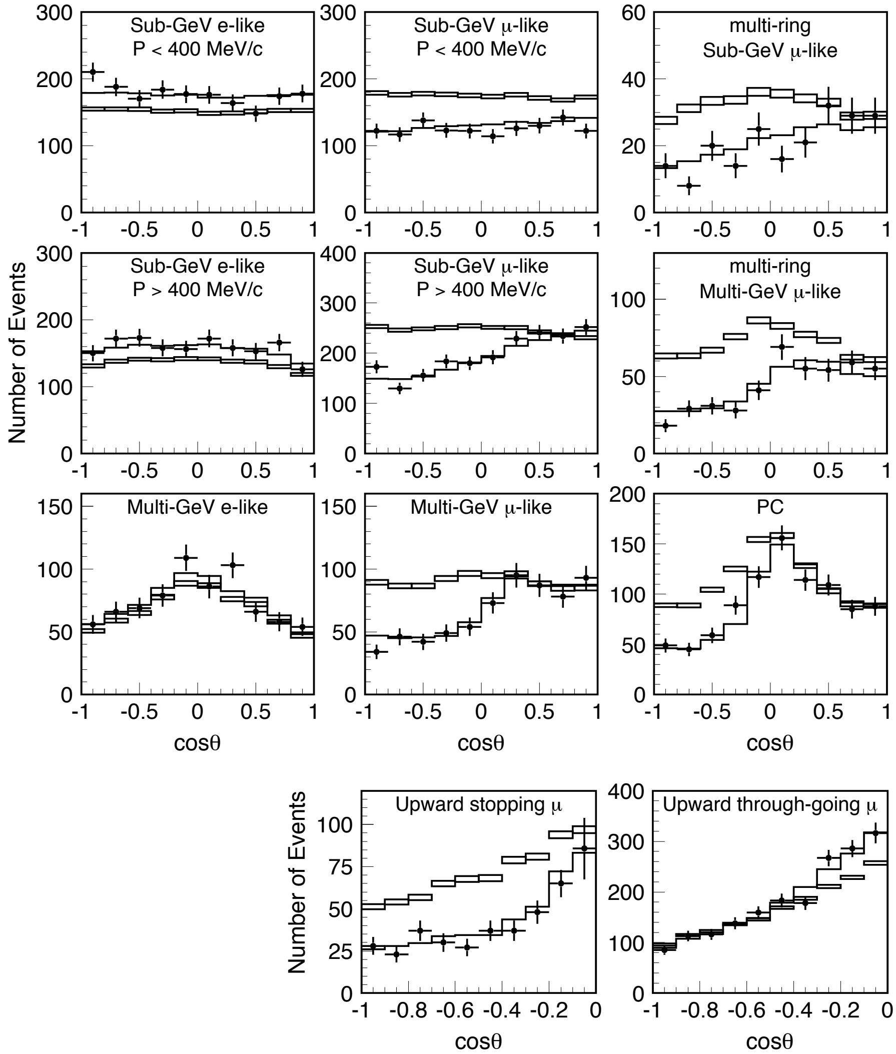


Fig. 30: Zenith angle distribution for fully-contained single-ring e-like and μ-like events, multi-ring μ-like events, partially contained events, and upward-going muons. The points show the data and the boxes show the Monte Carlo events without neutrino oscillations. The solid lines show the best-fit expectations for $\nu_\mu \leftrightarrow \nu_\tau$ oscillations (from Ref. [40]).

is large enough, the appearance of τ neutrinos.

Three conventional beams confirmed the atmospheric oscillation from the measurement of the disappearance of ν_μ neutrinos: K2K (L = 235 km) [42], MINOS (L = 730 km) [43] and from the appearance of ν_τ, OPERA (L = 730 km) [44]. Fig. 32 shows the measurement of the ν_μ survival probability as a function of the reconstructed neutrino energy in the MINOS experiment.

Three reactor neutrino experiments, Daya Bay [46], RENO [47] and Double Chooz [48], have discovered that the electron neutrino flavour also oscillates with the atmospheric wavelength: electron antineutrinos from reactors disappear at distances of $\mathcal{O}(1\text{ km})$, but with a small amplitude. See Fig. 33.

Finally the T2K and NOVA experiments have measured the appearance of ν_e and $\bar{\nu}_e$ in an accelerator $\nu_\mu/\bar{\nu}_\mu$ beam [49, 50] in the atmospheric range. The agreement of all these measurements with the

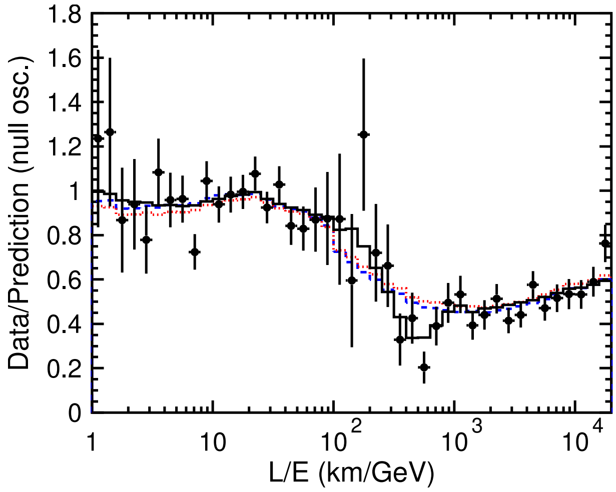


Fig. 31: Ratio of the data to the non-oscillated Monte Carlo events (points) with the best-fit expectation for 2-flavour $\nu_\mu \leftrightarrow \nu_\tau$ oscillations (solid line) as a function of E_ν/L (from Ref. [41]).

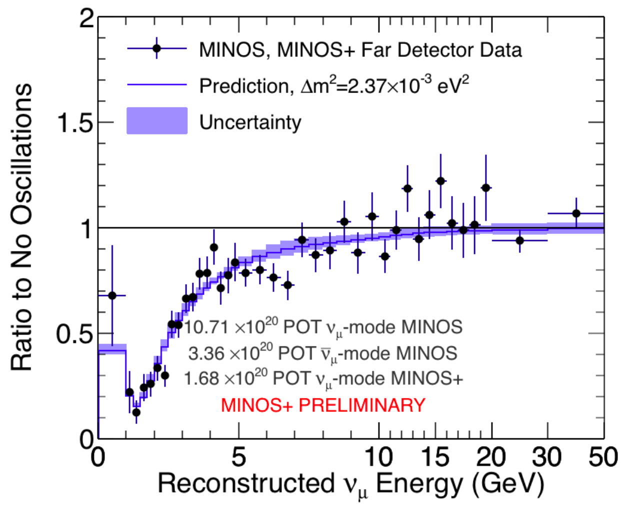


Fig. 32: Ratio of measured to expected (in absence of oscillations) neutrino events in MINOS as a functions of neutrino energy compared to the best fit oscillation solution (from Ref. [45]).

original atmospheric oscillation signal is excellent.

8 The three-neutrino mixing scenario

As we have seen, the evidence summarized in the previous section points to two distinct neutrino mass square differences related to the solar and atmospheric oscillation frequencies:

$$\underbrace{|\Delta m^2_{\text{solar}}|}_{\sim 8\cdot 10^{-5}\ \text{eV}^2} \ll \underbrace{|\Delta m^2_{\text{atmos}}|}_{\sim 2.5\cdot 10^{-3}\ \text{eV}^2} \tag{8.1}$$

The mixing of the three standard neutrinos ν_e, ν_μ, ν_τ can accommodate both. The two independent neutrino mass square differences are conventionally assigned to the solar and atmospheric ones in the

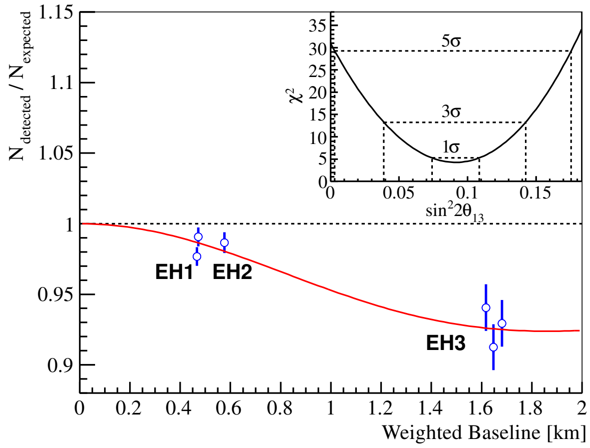


Fig. 33: Ratio of measured to expected reactor neutrino events as function of the baseline in the Daya Bay experiment (from Ref. [46]).

following way:

$$\Delta m^2_{13} = m^2_3 - m^2_1 = \Delta m^2_{\rm atmos}, \qquad \Delta m^2_{12} = m^2_2 - m^2_1 = \Delta m^2_{\rm solar}\,. \tag{8.2}$$

The PMNS mixing matrix depends on three angles and one or more CP phases (see Eq. (4.7) for the standard parametrization). Only one CP phase, the so-called Dirac phase δ, appears in neutrino oscillation probabilities.

With this convention, the mixing angles θ_{23} and θ_{12} in the parametrization of Eq. (4.7) correspond approximately to the ones measured in atmospheric and solar oscillations, respectively. This is because solar and atmospheric anomalies approximately decouple as independent 2-by-2 mixing phenomena thanks to the hierarchy between the two mass splittings, $|\Delta m^2_{\rm atmos}| \gg |\Delta m^2_{\rm solar}|$, on the one hand, and the fact that the angle θ_{13}, which measures the electron component of the third mass eigenstate element $\sin\theta_{13} = (U_{\rm PMNS})_{e3}$, is small.

To see this, let us first consider the situation in which $E_\nu/L \sim |\Delta m^2_{\rm atmos}|$. We can thus neglect the solar mass square difference in front of the atmospheric one and E_ν/L. The oscillation probabilities obtained in this limit are given by

$$P(\nu_e \to \nu_\mu) \simeq s^2_{23}\,\sin^2 2\theta_{13}\,\sin^2\left(\frac{\Delta m^2_{13} L}{4E_\nu}\right), \tag{8.3}$$

$$P(\nu_e \to \nu_\tau) \simeq c^2_{23}\,\sin^2 2\theta_{13}\,\sin^2\left(\frac{\Delta m^2_{13} L}{4E_\nu}\right), \tag{8.4}$$

$$P(\nu_\mu \to \nu_\tau) \simeq c^4_{13}\,\sin^2 2\theta_{23}\,\sin^2\left(\frac{\Delta m^2_{13} L}{4E_\nu}\right). \tag{8.5}$$

The results for antineutrinos are the same (there is no CP violation if one mass difference is neglected). All flavours oscillate therefore with the atmospheric frequency, but only two angles enter these formulae: θ_{23} and θ_{13}. The latter is the only one that enters the disappearance probability for ν_e or $\bar\nu_e$ in this regime since

$$P(\nu_e \to \nu_e) = P(\bar{\nu}_e \to \bar{\nu}_e) = 1 - P(\nu_e \to \nu_\mu) - P(\nu_e \to \nu_\tau) \simeq \sin^2 2\theta_{13} \sin^2\left(\frac{\Delta m^2_{13} L}{4E_\nu}\right) . \quad (8.6)$$

This is precisely the measurement of reactor neutrino experiments like Chooz, Daya Bay, RENO and Double Chooz. Therefore the oscillation amplitude of these experiments is a direct measurement of the angle θ_{13}, which has been measured to be small.

Note that in the limit $\theta_{13} \to 0$, the only probability that survives in Eq. (8.5) is the $\nu_\mu \to \nu_\tau$ one, which has the same form as a 2-family mixing formula Eq. (6.22) if we identify

$$(\Delta m^2_{\text{atmos}}, \theta_{\text{atmos}}) \to (\Delta m^2_{13}, \theta_{23}) . \quad (8.7)$$

Therefore the close-to-maximal mixing angle observed in atmospheric neutrinos and the accelerator neutrino experiments like MINOS is identified with θ_{23}.

Instead if we consider experiments in the solar range, $E_\nu/L \sim \Delta m^2_{\text{solar}}$, the atmospheric oscillation its too rapid and gets averaged out. The survival probability for electrons in this limit is given by:

$$P(\nu_e \to \nu_e) = P(\bar{\nu}_e \to \bar{\nu}_e) \simeq c^4_{13}\left(1 - \sin^2 2\theta_{12} \sin^2\left(\frac{\Delta m^2_{12} L}{4E_\nu}\right)\right) + s^4_{13}. \quad (8.8)$$

Again it depends only on two angles, θ_{12} and θ_{13}, and in the limit in which the latter is zero, the survival probability measured in solar experiments has the form of two-family mixing if we identify

$$(\Delta m^2_{\text{solar}}, \theta_{\text{solar}}) \to (\Delta m^2_{12}, \theta_{12}) . \quad (8.9)$$

The results that we have shown in the previous section of solar and atmospheric experiments have been analysed in terms of 2-family mixing. The previous argument indicates that when fits are done in the context of 3-family mixing nothing changes too much.

On the other hand, the fact that reactor experiments have already measured the disappearance of reactor $\bar{\nu}_e$ in the atmospheric range implies that the effects of $\theta_{13} \simeq 9^\circ$ are not negligible, and therefore a proper analysis of all the oscillation data requires performing global fits in the 3-family scenario. Figure 34 shows the $\Delta\chi^2$ as a function of each of the six parameters from one recent global analysis [51]. See also Refs. [52, 53].

There are two parameters in which we observe two distinct minima, these corresponds to degeneracies that cannot be resolved with present data. The first corresponds to the neutrino mass ordering or hierarchy: present data cannot distinguish between the normal (NH or NO) and inverted ordering (IH or IO) represented in Fig. 35.

Note that we denote by $\Delta m^2_{13} = \Delta m^2_{\text{atmos}}$ the atmospheric splitting for NO and $\Delta m^2_{23} = -\Delta m^2_{\text{atmos}}$ for IO. The second degeneracy corresponds to the octant choice of θ_{23}. Present data are mostly sensitive to $\sin^2 2\theta_{23}$. If this angle is not maximal, there are two possible choices that are roughly equivalent $\theta_{23} \leftrightarrow \pi/4 - \theta_{23}$. Due to this degeneracy, the largest angle is also the one less accurate. The

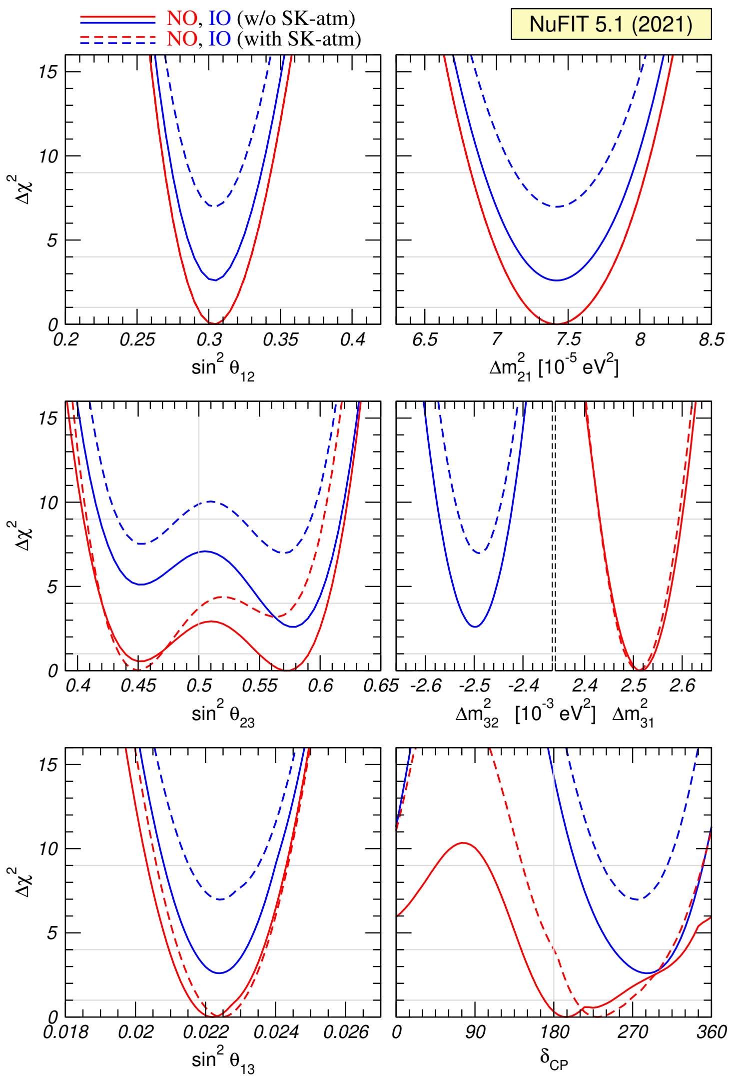


Fig. 34: $\Delta\chi^2$ of the fits to the standard 3ν-mixing scenario including all available neutrino oscillation data (from Ref. [51]). The solid lines do not include SK atmospheric data, while the dashed ones do.

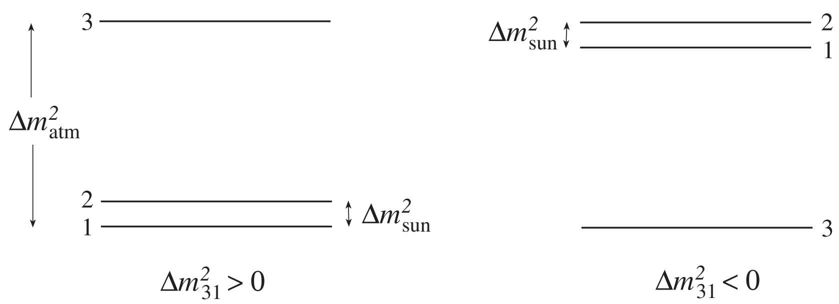


Fig. 35: Possible neutrino spectra consistent with solar and atmospheric data.

1σ limits for NO are:

$$\begin{aligned}\theta_{23}/^\circ = 49.2^{+1}_{-1.3}, \qquad \theta_{12}/^\circ = 33.4^{+0.77}_{-0.74}, \qquad \theta_{13}/^\circ = 8.57(13),\\ \Delta m^2_{12} = 7.42(21)\times 10^{-5}\ \mathrm{eV}^2, \qquad \Delta m^2_{13} = 2.515(28)\times 10^{-3}\ \mathrm{eV}^2.\end{aligned} \tag{8.10}$$

The CP phase δ remains roughly unconstrained at 3σ, while there is about half of the region excluded at 2σ. As we will see, the dependence on the phase requires sensitivity to both frequencies simultaneously.

9 Prospects in determining unknown neutrino parameters

An ambitious experimental program is underway to pin down the remaining unknowns and reach a 1% precision in the lepton flavour parameters. The neutrino ordering, the octant of θ_{23} and the CP violating phase, δ, can be searched for in neutrino oscillation experiments with improved capabilities. The determination of the absolute neutrino mass scale relies on tritium beta decay experiments or cosmology.

9.1 Neutrino ordering

Concerning the neutrino ordering, the best hope to identify the spectrum exploits the MSW effect in the propagation of GeV neutrinos through earth's matter. In the case of three neutrinos propagating in matter, the ν mass eigenstates as a function of the electron density for vanishing θ_{12}, θ_{13} are depicted in Fig. 36 for NO and IO. For NO we see that there are two level crossings giving rise to two MSW resonances. The first one is essentially the one relevant for solar neutrinos, as it affects the smallest mass splitting, with the resonance condition:

$$E^{(1)}_{\mathrm{res}} = \frac{\Delta m^2_{12}\cos 2\theta_{12}}{2\sqrt{2}G_F N_e}. \tag{9.1}$$

The second one affects the largest mass splitting

$$E^{(2)}_{\mathrm{res}} = \frac{\Delta m^2_{13}\cos 2\theta_{13}}{2\sqrt{2}G_F N_e}. \tag{9.2}$$

For IO, only the first resonance appears in the ν channel.

For $\bar{\nu}$ the dependence on N_e of the first eigenstate has a negative slope and therefore there is no resonance for NO and only the atmospheric resonance appears for IO.

The existence of the atmospheric resonance implies a large enhancement of the oscillation probability $P(\nu_e \leftrightarrow \nu_\mu)$ for NO for energies near the resonant energy and at sufficiently long baseline. For IO the enhancement occurs in $P(\bar{\nu}_e \leftrightarrow \bar{\nu}_\mu)$ instead. For the typical matter densities of the earth's crust and mantle and the value of the atmospheric mass splitting, the resonant energy for neutrinos travelling through earth is $\simeq 6$ GeV, an energy that can be reached in accelerator neutrino beams. The measurement of the neutrino ordering becomes almost a digital measurement sending a conventional ν beam sufficiently far as shown in Fig. 37, which shows the oscillation probability $P(\nu_\mu \to \nu_e)$ as a function of the neutrino energy at a distance corresponding to the baseline from CERN-Kamioka (8770 km).

The first experiment that will be sensitive to this effect is the NOvA experiment, optimized like

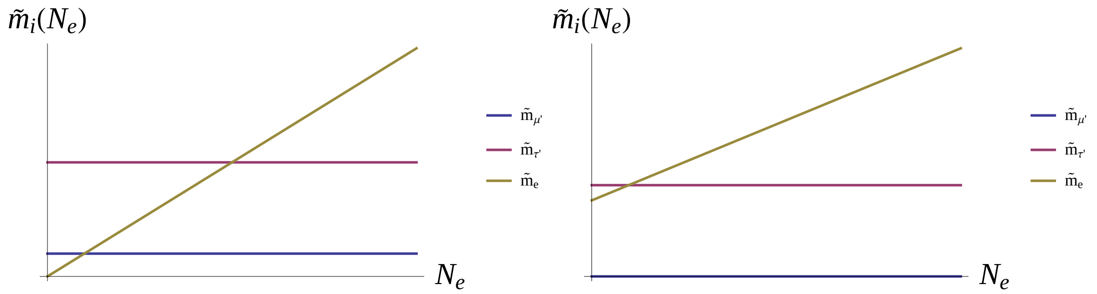


Fig. 36: Level crossings for ν in the three neutrino scenario for NO (left) and IO (right) at vanishing θ_{12} and θ_{13}.

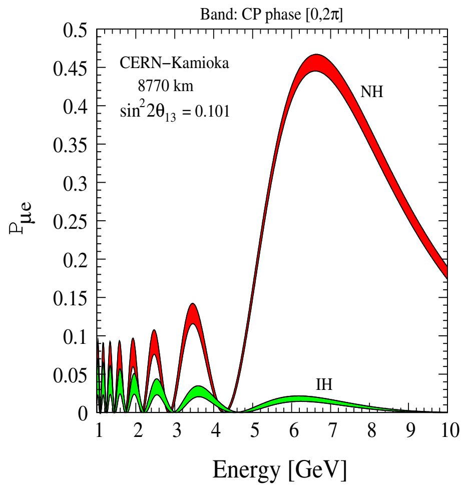


Fig. 37: Resonant increase of the $P_{\mu e}$ for NH as a function of neutrino energy for L corresponding to the distance CERN-Kamioka for NH/IH. The bands corresponds to the uncertainty in δ (from Ref. [54]).

T2K to see the ν_e appearance signal, with a baseline of 810km, which is however a bit short to see a large enhancement. Nevertheless if lucky NOvA could discriminate the ordering at 3σ.

The atmospheric resonance must also affect atmospheric neutrinos at the appropriate energy and baseline. Unfortunately the atmospheric flux contains both neutrinos and antineutrinos in similar numbers, and the corresponding events cannot be told apart, because present atmospheric neutrino detectors cannot measure the lepton charge. If we superimpose the neutrino and antineutrino signals, both orderings will give rise to an enhancement in the resonance region, since either the neutrino or antineutrino channel will have a resonance. Nevertheless with sufficient statistics, there is some discrimination power and in fact the biggest neutrino telescopes, IceCube and KM3NeT have proposed to instrument more finely some part of their detectors (PINGU and ORCA projects) to perform this measurement. Also the next generation of atmospheric neutrino detectors, such as Hyper-Kamiokande, with a factor $\mathcal{O}(20)$ more mass than the present Super-Kamiokande, or the INO detector that is designed to measure the muon charge in atmospheric events, could discriminate between the two orderings.

A very different strategy has been proposed for reactor neutrino experiments (e.g. JUNO project). The idea is to measure very precisely the reactor neutrinos at a baseline of roughly 50 km, where the depletion of the flux due to the solar oscillation is maximal. At this optimal distance, one can get a superb

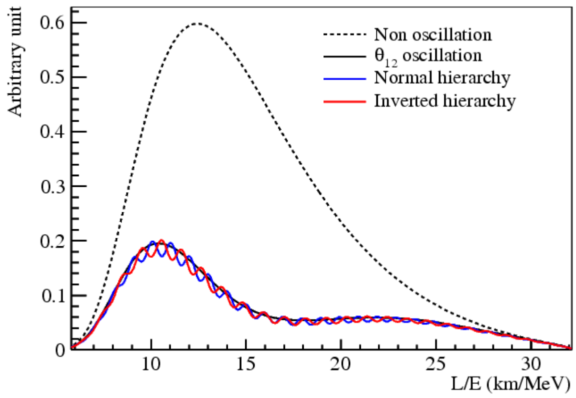


Fig. 38: Reactor neutrino spectrum in JUNO for NO/IO (from Ref. [57]).

measurement of the solar oscillation parameters, $(\theta_{12}, \Delta m^2_{12})$, and, with sufficient energy resolution, one could detect the modulation of the signal due to the atmospheric oscillation [55, 56]. Figure 38 shows how this modulation is sensitive to the neutrino ordering. A leap ahead is however needed to reach the required energy resolution that would enable this measurement.

9.2 Leptonic CP violation

As we have seen, the CP phase, δ, in the mixing matrix induces CP violation in vacuum neutrino oscillations, that is a difference between $P(\nu_\alpha \to \nu_\beta)$ and $P(\bar{\nu}_\alpha \to \bar{\nu}_\beta)$, for $\alpha \neq \beta$. As we saw in the general expression of Eq. (6.20), CP violation is possible if there are imaginary entries in the mixing matrix that make $\text{Im}[W^{jk}_{\alpha\beta}] \neq 0$. By CPT, disappearance probabilities cannot violate CP however, because under CPT

$$P(\nu_\alpha \to \nu_\beta) = P(\bar{\nu}_\beta \to \bar{\nu}_\alpha)\,, \tag{9.3}$$

so in order to observe a CP or T-odd asymmetry the initial and final flavour must be different, $\alpha \neq \beta$:

$$A^{CP}_{\alpha\beta} \equiv \frac{P(\nu_\alpha \to \nu_\beta) - P(\bar{\nu}_\alpha \to \bar{\nu}_\beta)}{P(\nu_\alpha \to \nu_\beta) + P(\bar{\nu}_\alpha \to \bar{\nu}_\beta)}, \quad A^{T}_{\alpha\beta} \equiv \frac{P(\nu_\alpha \to \nu_\beta) - P(\nu_\beta \to \nu_\alpha)}{P(\nu_\alpha \to \nu_\beta) + P(\nu_\beta \to \nu_\alpha)}\,. \tag{9.4}$$

In the case of 3-family mixing it is easy to see that the CP(T)-odd terms in the numerator are the same for all transitions $\alpha \neq \beta$:

$$A^{\text{CP(T)-odd}}_{\nu_\alpha\nu_\beta} = \frac{\sin\delta c_{13} \sin 2\theta_{13} \overbrace{\sin 2\theta_{12} \frac{\Delta m^2_{12} L}{4E_\nu}}^{\text{solar}} \overbrace{\sin 2\theta_{23} \sin^2 \frac{\Delta m^2_{13} L}{4E_\nu}}^{\text{atmos}}}{P^{\text{CP-even}}_{\nu_\alpha\nu_\beta}}\,. \tag{9.5}$$

As expected, the numerator is GIM suppressed in all the Δm^2_{ij} and all the angles, because if any of them is zero, the CP-odd phase becomes unphysical. Therefore an experiment which is sensitive to CP violation must be sensitive to both mass splittings simultaneously. In this situation, it is not clear a priori what the optimization of E/L should be.

It can be shown that including only statistical errors, the signal-to-noise ratio for this asymmetry is maximized for $\langle E_\nu \rangle / L \sim \Delta m^2_{\rm atmos}$. In this case, only two small parameters remain in the CP-odd terms: the solar splitting, $\Delta m^2_{\rm solar}$ (i.e., compared to the other scales, $\Delta m^2_{\rm atmos}$ and $\langle E_\nu \rangle / L$), and the angle θ_{13}. The asymmetry is then larger in the sub-leading transitions: $\nu_e \to \nu_\mu(\nu_\tau)$, because the CP-even terms in the denominator are also suppressed by the same small parameters. A convenient approximation for the $\nu_e \leftrightarrow \nu_\mu$ transitions is obtained expanding to second order in both small parameters [58]:

$$
\begin{aligned}
P_{\nu_e\nu_\mu(\bar\nu_e\bar\nu_\mu)} \;&=\; s^2_{23}\, \sin^2 2\theta_{13}\, \sin^2\left(\frac{\Delta m^2_{13}\, L}{4E_\nu}\right) && \equiv P^{\rm atmos}\\
&+\; c^2_{23}\, \sin^2 2\theta_{12}\, \sin^2\left(\frac{\Delta m^2_{12}\, L}{4E_\nu}\right) && \equiv P^{\rm solar}\\
&+\; \tilde J \cos\left(\pm\delta - \frac{\Delta m^2_{13}\, L}{4E_\nu}\right) \frac{\Delta m^2_{12}\, L}{4E_\nu} \sin\left(\frac{\Delta m^2_{13}\, L}{4E_\nu}\right) && \equiv P^{\rm inter}, \qquad (9.6)
\end{aligned}
$$

where $\tilde J \equiv c_{13}\ \sin 2\theta_{13}\ \sin 2\theta_{12}\ \sin 2\theta_{23}$. The first term corresponds to the atmospheric oscillation, the second one is the solar one and there is an interference term which has the information on the phase δ and depends on both mass splittings.

These results correspond to vacuum propagation, but usually these experiments require the propagation of neutrinos in the earth's matter. The oscillation probabilities in matter can also be approximated by a similar series expansion [58]. The result has the same structure as in vacuum:

$$
\begin{aligned}
P_{\nu_e\nu_\mu(\bar\nu_e\bar\nu_\mu)} \;&=\; s^2_{23}\, \sin^2 2\theta_{13} \left(\frac{\Delta_{13}}{B_\pm}\right)^2 \sin^2\left(\frac{B_\pm L}{2}\right)\\
&+\; c^2_{23}\, \sin^2 2\theta_{12} \left(\frac{\Delta_{12}}{A}\right)^2 \sin^2\left(\frac{A\, L}{2}\right)\\
&+\; \tilde J\, \frac{\Delta_{12}}{A}\, \sin\left(\frac{AL}{2}\right) \frac{\Delta_{13}}{B_\pm}\, \sin\left(\frac{B_\pm L}{2}\right) \cos\left(\pm\delta - \frac{\Delta_{13}\, L}{2}\right), \qquad (9.7)
\end{aligned}
$$

where

$$
B_\pm = |A \pm \Delta_{13}| \qquad , \Delta_{ij} = \frac{\Delta m^2_{ij}}{2E_\nu} \qquad , A = \sqrt{2} G_F N_e\,. \qquad (9.8)
$$

The oscillation probability for neutrinos and antineutrinos now differ not just because of leptonic CP violation, but also due to the matter effects, that as we have seen can be resonant. In particular, the atmospheric term which is the dominant one, shows the expected resonant enhancement in the neutrino or antineutrino oscillation probability (depending on the ordering).

The sensitivity to the interference term requires very good knowledge of the leading atmospheric term and the present degeneracies (the octant and the neutrino ordering) directly affect the leading term compromising therefore the δ sensitivity. Either both uncertainties are solved before this measurement, or there must be sufficient sensitivity from the energy dependence of the signal to resolve all unknowns simultaneously.

A rough optimization of L for fixed E/L for discovering CP violation is shown in Fig. 39. It shows the signal-to-noise as a function of the true value of δ, assuming only statistical errors, but including the expected dependence of the cross sections and fluxes. At very short baselines, the sensitivity is compromised due to the lack of knowledge of the neutrino ordering. In a wide intermediate region

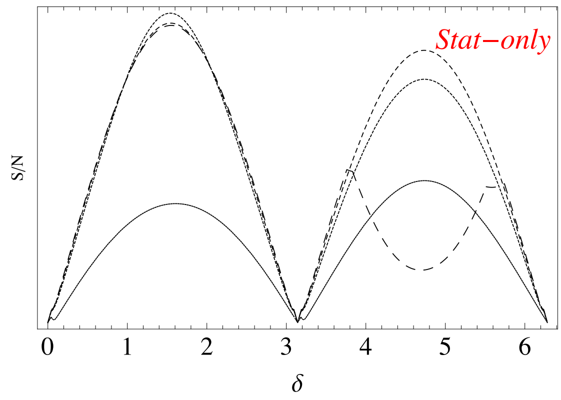


Fig. 39: Signal-to-noise for the discovery of CP violation at fixed $E/L \sim \Delta m^2_{\rm atm}$ as a function of the true value of δ for $L = 295$km (long-dashed), $L = 650$km (short-dashed), $L = 1300$km (dotted), $L = 2300$km (solid). The ordering is assumed to be unknown.

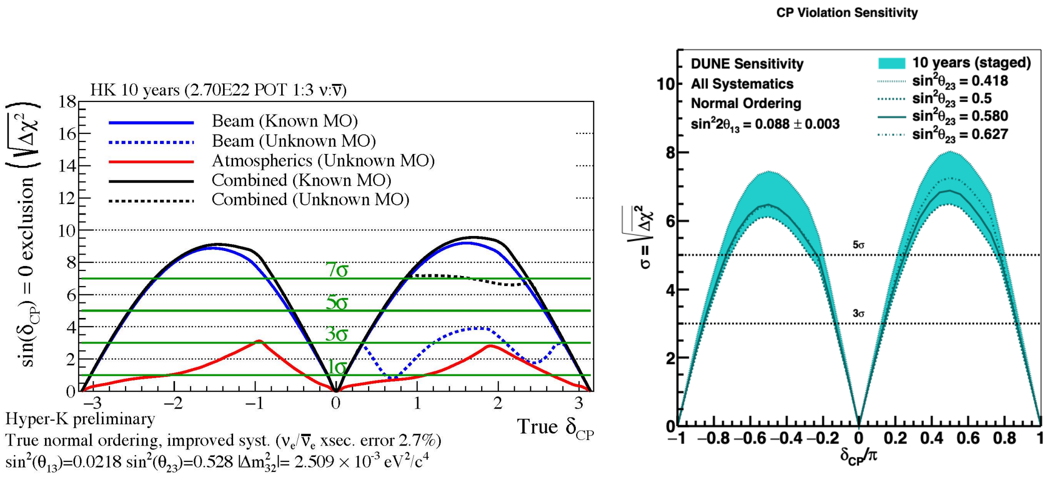


Fig. 40: Sensitivity to CP violation as a function of the true value of δ in Hyper-Kamiokande (left) [59] and DUNE (right) [60]. Solid (dashed) lines on the left plot correspond to the mass ordering (MO) known(unknown).

around $\mathcal{O}(1000)$km the sensitivity is optimal, and at much larger baselines the sensitivity deteriorates because the matter effects completely hide CP-violation.

Several projects have been proposed to search for leptonic CP violation, including conventional beams, but also novel neutrino beams from muon decays (neutrino factories), from radioactive ion decays (β-beams) or from spalation sources (ESS). The relatively large value of θ_{13} has refocused the interest in using the less challenging conventional beams and two projects are presently being developed: the Hyper-Kamiokande detector, an up-scaled version of Super-Kamiokande that will measure atmospheric neutrinos with unprecedented precision, and also intercept a neutrino beam from JPARC at a relatively short baseline $L = 295$km, and the DUNE project that involves a liquid argon neutrino detector and a neutrino beam from Fermilab to the Soudan mine at a baseline of $L = 1500$km. The expected sensitivities to CP violation of both projects are shown in Fig. 40.

9.3 Absolute neutrino mass scale

Neutrino oscillation experiments are only sensitive to neutrino mass differences, so at present we do not have information on the absolute neutrino mass scale, only upper limits. The sum of all neutrino masses is tightly constrained by cosmological measurements of the cosmic microwave background (CMB) [61]:

$$\sum_i m_i \leq 0.12\,\text{eV}. \tag{9.9}$$

As we have seen the kinematical effects of neutrino masses in this range can also modify the end-point spectrum of beta decay. More precisely, this measurement can constrain the combination

$$m_{\nu_e} \equiv \sqrt{\sum_i |U_{ei}|^2 m_i^2}. \tag{9.10}$$

The strongest upper limit of 0.8 eV as we saw has been set by the Katrin experiment [7].

In Fig. 41 we show the allowed regions on the plane m_{ν_e} vs $\sum_i m_i$ from the known neutrino masses and mixings. The limit from cosmology on the right axis is already more stringent (although cosmological model dependent) than the present and future expected sensitivity of the Katrin experiment.

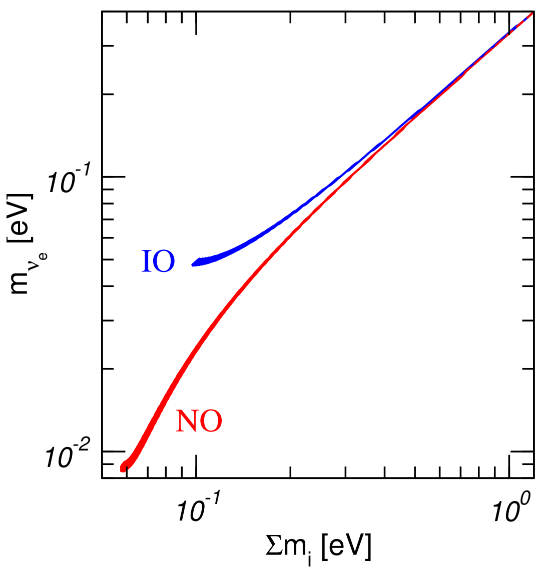


Fig. 41: Allowed region for m_{ν_e} for IO (blue contour) and NO (red contour) from a global analysis of neutrino data (from Ref. [51]) on the plane m_{ν_e} vs the sum of all neutrino masses.

10 Outliers: the LSND anomaly

The long-standing puzzle brought by the LSND experiment is still unresolved. This experiment [62] observed a surplus of electron events in a muon neutrino beam from π^+ decaying in flight (DIF) and a surplus of positron events in a neutrino beam from μ^+ decaying at rest (DAR). The interpretation of this data in terms of neutrino oscillations, that is a non-vanishing $P(\nu_\mu \to \nu_e)$, gives the range shown by a coloured band in Fig. 43.

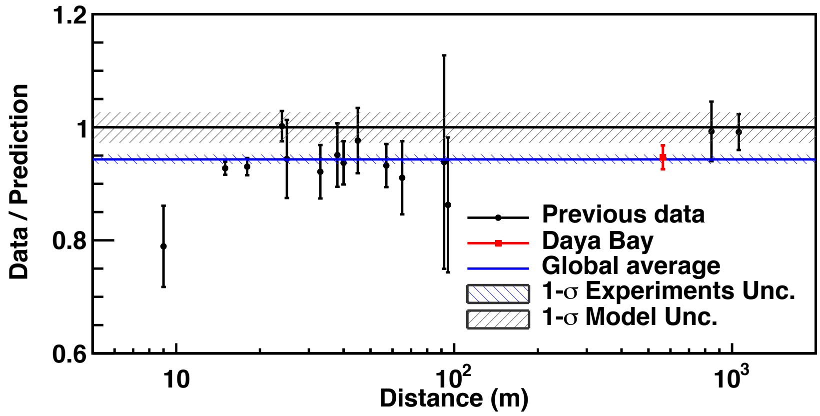


Fig. 42: Reactor neutrino flux measured by various near detectors compared with the recent flux predictions (from Ref. [70]).

$$
\begin{array}{llll}
\pi^+ \to & \mu^+ & \nu_\mu & \\
& & \nu_\mu \to \nu_e & \text{DIF } (28 \pm 6/10 \pm 2) \\
& \mu^+ & \to e^+ \nu_e \bar{\nu}_\mu & \\
& & & \bar{\nu}_\mu \to \bar{\nu}_e \quad \text{DAR } (64 \pm 18/12 \pm 3)
\end{array}
$$

A significant fraction of this region was already excluded by the experiment KARMEN [63] that has unsuccessfully searched for $\bar{\nu}_\mu \to \bar{\nu}_e$ in a similar range.

The experiment MiniBOONE was designed to further investigate the LSND signal, with inconclusive results [64]. They did not confirm the LSND anomaly, but found a significant excess at lower energies [65]. Recently the MicroBoone experiment [66], designed to have improved discrimination capabilities of NC background, did not find evidence for the MiniBOONE anomaly.

On the other hand, the results of various short baseline (tens of meters) reactor neutrino experiments were revised, after an update on the reactor neutrino flux predictions [67–69], which increased these fluxes by a few per cent. While the measured neutrino flux was found to be in agreement with predictions before, after this revision some reactor neutrinos seem to disappear before reaching near detectors, $L = \mathcal{O}(10)$m. This is the so-called reactor anomaly shown in Fig. 42. This result brought some excitement because if this disappearance is due to oscillations, it might reinforce the oscillation interpretation of the LSND anomaly.

The required mass splitting to describe both anomalies is $\Delta m^2_{\rm LSND} \simeq 1{\rm eV}^2$, which is much larger than the solar and atmospheric, and therefore requires the existence of at least a fourth neutrino mass eigenstate, i. If such a state can explain the LSND anomaly, it must couple to both electrons and muons. Unfortunately the smoking gun would require that also accelerator ν_μ disappear with the same wavelength and this has not been observed:

$$
\begin{array}{rcll}
P(\nu_\mu \to \nu_e) & \propto & |U_{ei} U_{\mu i}|^2 & \quad \text{LSND} \\
1 - P(\nu_e \to \nu_e) & \propto & |U_{ei}|^4 & \quad \text{reactor} \\
1 - P(\nu_\mu \to \nu_\mu) & \propto & |U_{\mu i}|^4 & \quad \text{not observed}
\end{array}
$$

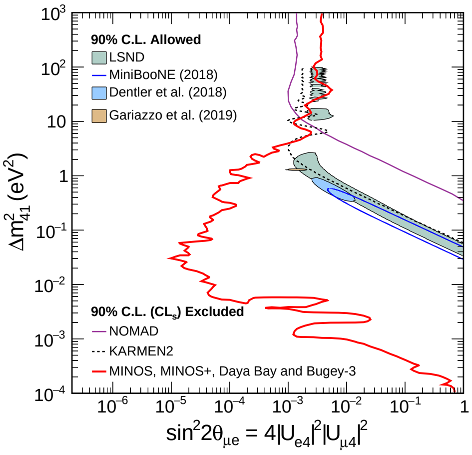


Fig. 43: Sterile neutrino search combining disappearance of ν_μ's and ν_e (from Ref. [71]). At 90% CL only the region to the left of the red line is allowed, excluding most of the regions favoured by LSND, MiniBoone and the global fits.

The strongest constraint on the disappearance of ν_μ in the LSND range has been recently set by MINOS+ and the tension between appearance and disappearance measurements is shown in Fig. 43.

Very recently a new update on the flux predictions has been presented and the significance of the reactor anomaly has decreased. In parallel a plethora of new short baseline reactor neutrino experiments (Prospect, DANSS, Stereo, NEOS, NEUTRINO-4) have taken data exploiting the L dependence of a putative oscillation signal. The results have for the most part not confirmed the oscillation of reactor neutrinos. A global analysis of all the reactor data results shows that at 2.6σ the results are compatible with the non-oscillation hypothesis. See Ref. [72] for a recent status and references.

11 Neutrinos and BSM physics

The new lepton flavour sector of the SM has opened new perspectives into the flavour puzzle. As we have seen neutrinos are massive but significantly lighter than the remaining charged fermions. Clearly the gap of Fig. 11 calls for an explanation. The leptonic mixing matrix is also very different to that in the quark sector. The neutrino mixing matrix is approximately given in Ref. [51]

$$|U_{\rm PMNS}|_{3\sigma} \simeq \begin{pmatrix} 0.80-0.84 & 0.51-0.58 & 0.14-0.16 \\ 0.23-0.50 & 0.46-0.69 & 0.63-0.78 \\ 0.26-0.52 & 0.47-0.70 & 0.61-0.76 \end{pmatrix} . \tag{11.1}$$

The CKM matrix is presently constrained [73] to be:

$$|V_{\rm CKM}| \simeq \begin{pmatrix} 0.97435(16) & 0.22500(67) & 0.00369(11) \\ 0.22486(67) & 0.97349(16) & 0.04182(85) \\ 0.00857(20) & 0.04110(83) & 0.999118(31) \end{pmatrix} . \tag{11.2}$$

There is a striking difference between the two (and not only in the precision of the entries). The CKM matrix is close to the unit matrix:

$$V_{\text{CKM}} \simeq \begin{pmatrix} 1 & O(\lambda) & O(\lambda^3) \\ O(\lambda) & 1 & O(\lambda^2) \\ O(\lambda^3) & O(\lambda^2) & 1 \end{pmatrix}, \quad \lambda \sim 0.2, \tag{11.3}$$

while the leptonic one has large off-diagonal entries. With a similar level of precision, it is close to the tri-bimaximal mixing pattern [74]

$$U_{\text{PMNS}} \simeq V_{\text{tri-bi}} \simeq \begin{pmatrix} \sqrt{\frac{2}{3}} & \sqrt{\frac{1}{3}} & 0 \\ -\sqrt{\frac{1}{6}} & \sqrt{\frac{1}{3}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{6}} & -\sqrt{\frac{1}{3}} & \sqrt{\frac{1}{2}} \end{pmatrix}.$$

Discrete flavour symmetries have been extensively studied as the possible origin of this pattern.

While we do not have yet a compelling explanation of the different mixing patterns, we do have one for the gap between neutrino and other fermion masses. We saw that if the light neutrinos are Majorana particles and get their mass via the Weinberg interaction of Fig. 12, they are signalling BSM physics. As we have seen neutrino masses are then

$$m_\nu = \lambda \frac{v^2}{\Lambda}, \tag{11.4}$$

where Λ represents the mass of the neutrino mass mediators, i.e. the heavy particles that give rise to the Weinberg interaction. The more massive these particles are, the lighter neutrinos become. This is the famous *seesaw* mechanism depicted in Fig. 44.

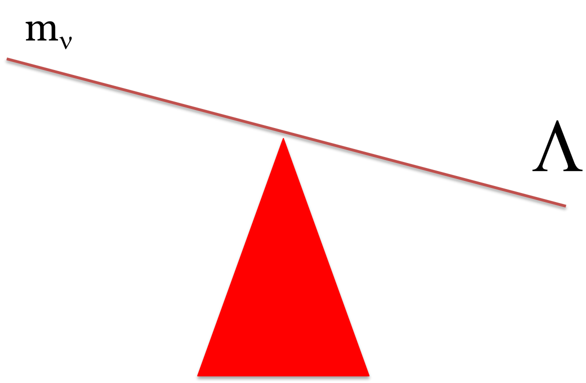


Fig. 44: Seesaw mechanism: the higher the scale Λ of new physics is, the lighter neutrino masses become.

λ on the other hand is the strength of the coupling of the new states with the lepton and Higgs doublets. Both parameters are in principle undetermined and only the combination $\frac{\lambda}{\Lambda}$ is fixed by neutrino masses. If we assume that the *natural* choice for λ is $\mathcal{O}(1)$, then neutrino masses require $\Lambda \sim M_{\text{GUT}}$, that is a grand unification scale. This is an intriguing fact, however it leads to the famous hierarchy problem [75, 76]:

$$m_H^2 \sim \Lambda^2. \tag{11.5}$$

The recent discovery of the Higgs field and in particular the value of its mass $m_H = 125$ GeV [77] suggests that the SM is as healthy as ever. In spite of the Landau poles present in the theory, the value of the SM couplings surprisingly conspire to make the model consistent up to the Planck scale [78].

On the other hand, the SM contains other small couplings, for example the electron Yukawa coupling is $Y_e \sim \mathcal{O}(10^{-6})$. It is then a fair question to ask how small can λ be not to worsen the flavour hierarchies in the charged lepton and quark sectors. Unfortunately the answer to this question depends on the underlying model. We can for example consider the three types of seesaw models, which correspond to the models that give rise to the Weinberg operator from the exchange of a massive particle, as depicted in Fig. 45:

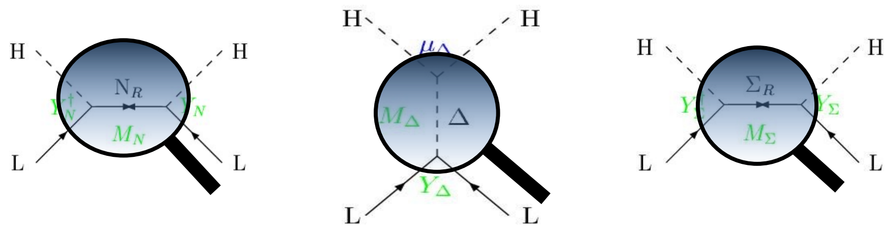


Fig. 45: Magnifying-glass view of the Weinberg operator in seesaw models of Type I (left), Type II (middle), Type III (right).

- type I see-saw: SM+ heavy singlet fermions, N, with mass M_N [79–82],
- type II see-saw: SM + heavy triplet scalar, Δ, with mass M_Δ [83–87],
- type III see-saw: SM + heavy triple fermions, Σ with mass M_Σ [88, 89],

In each of these cases $\Lambda = M_{N/\Delta/\Sigma}$ and the matching of the underlying theory to the Weinberg interaction fixes λ. For Type I and III:

$$\text{TypeI/III}: \ \lambda = \mathcal{O}(Y^2_{N/\Sigma}), \tag{11.6}$$

where $Y_{N,\Sigma}$ is the neutrino Yukawa coupling. In the case of Type II also the scalar trilinear coupling enters. If we now plot the hierarchies in the Yukawa couplings as opposed to the masses for Type I and III, we see that assuming a $Y_{N,\Sigma} \sim Y_e$, the scale Λ can be close to the electroweak scale, as shown in Fig. 46.

It is also possible that Weinberg's interaction is generated by new physics at higher orders, such as in the famous Zee model [90] and related ones [91, 92]. In this case, neutrino masses have an additional suppression by loop factors $1/(16\pi^2)$ and generically higher powers of the couplings of the underlying theory.

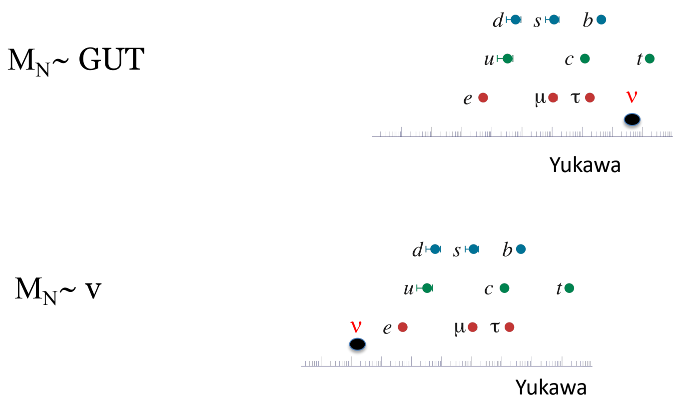


Fig. 46: Yukawa hierarchies in the Type I and III seesaw model if $M_N \sim M_{\rm GUT}$ or $\sim v$.

Summarizing, for $\Lambda \in [v, M_{\rm GUT}]$, neutrino masses do not imply larger hierarchies than already present in the minimal SM. Determining the scale Λ is one of the crucial problems in neutrino physics that we will try to elucidate in the future.

If $\Lambda \gg 100\,\text{MeV}$, there is a model-independent prediction: neutrinoless double-beta decay is possible with an amplitude proportional to the combination

$$m_{ee} = \sum_{i=1,3} (U_{\rm PMNS})^2_{ei} m_i. \tag{11.7}$$

The information we already have about neutrino masses and mixings constrains this quantity to be in any of the bands in Fig. 47 depending on the neutrino mass ordering.

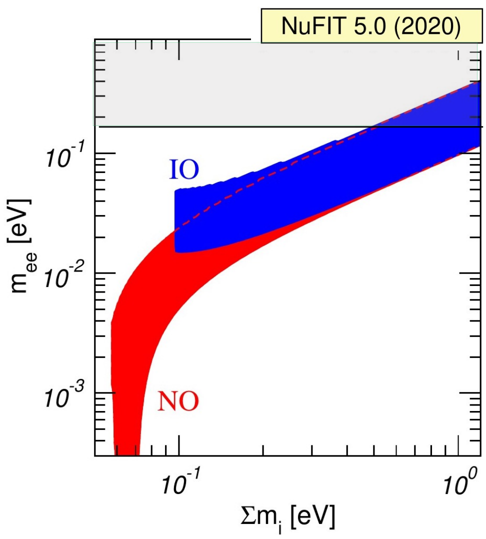


Fig. 47: Allowed region for m_{ee} for IO (blue contour) and NO (red contour) from a global analysis of neutrino data (from Ref. [51]) on the plane m_{ee} vs the sum of all neutrino masses. We have added by the shaded region the exclusion from present neutrinoless double-beta decay searches.

Obviously if Λ is below the energies of present colliders, the new particles may be directly accessible. The dynamics of this new physics sector breaks lepton number and generically might induce the generation of the baryon asymmetry in the universe or may be connected to dark matter. Unfortunately both predictions: the production of these new states in colliders and their connection to baryogenesis or dark matter are model dependent. The type I seesaw model is the better studied case so we will consider this scenario in the following discussion.

11.1 One example: Type I seesaw model

It is arguably the most minimal extension of the SM explaining neutrino masses [79–82]. It involves the addition of $n_R \geq 2$ singlet Weyl fermions, ν_R, to the SM. With $n_R = 2$ two light neutrinos can be massive, which is the minimum compatible with neutrino mass measurements, i.e. two neutrino mass differences. The minimum number of singlets required to give non-zero mass to the three light neutrinos is $n_R = 3$, as shown in Fig. 48.

$(\mathbf{1},\mathbf{2})_{-\frac{1}{2}}$	$(\mathbf{3},\mathbf{2})_{-\frac{1}{6}}$	$(\mathbf{1},\mathbf{1})_{-1}$	$(\mathbf{3},\mathbf{1})_{-\frac{2}{3}}$	$(\mathbf{3},\mathbf{1})_{-\frac{1}{3}}$	$(\mathbf{1},\mathbf{1})_0$
$\begin{pmatrix}\nu_e\\ e\end{pmatrix}_L$	$\begin{pmatrix}u^i\\ d^i\end{pmatrix}_L$	e_R	u^i_R	d^i_R	ν_R^1
$\begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}_L$	$\begin{pmatrix}c^i\\ s^i\end{pmatrix}_L$	μ_R	c^i_R	s^i_R	ν_R^2
$\begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}_L$	$\begin{pmatrix}t^i\\ b^i\end{pmatrix}_L$	τ_R	t^i_R	b^i_R	ν_R^3

Fig. 48: Particle content of the SM+Type I seesaw model with three light massive neutrinos.

The most general renormalizable Lagrangian which satisfies Lorentz and the gauge symmetries is given by:

$$\mathcal{L}_{\rm TypeI} \;=\; \mathcal{L}_{\rm SM} - \sum_{\alpha,i} \bar{L}^\alpha Y_\nu^{\alpha i}\, \tilde{\Phi}\, \nu_R^i - \sum_{i,j}^{n_R} \frac{1}{2} \bar{\nu}_R^{ic}\, M_N^{ij}\, \nu_R^j + \text{h.c.}\,, \tag{11.8}$$

where the new parameters involved are a $3 \times n_R$ neutrino Yukawa matrix and a $n_R \times n_R$ symmetric Majorana mass matrix for the singlet fields. Upon spontaneous symmetry breaking these couplings become mass terms, that can be written in the Majorana basis (ν_L^c, ν_R) as

$$\mathcal{L}_{\rm TypeI} \;\rightarrow\; \mathcal{L}_{\rm SM} - \frac{1}{2}\begin{pmatrix}\bar{\nu}_L & \bar{\nu_R^c}\end{pmatrix}\begin{pmatrix}0 & m_D\\ m_D^T & M_N\end{pmatrix}\begin{pmatrix}\nu_L^c\\ \nu_R\end{pmatrix} + h.c. + ... \tag{11.9}$$

where

$$m_D = Y_\nu \frac{v}{\sqrt{2}}. \tag{11.10}$$

Note that Dirac neutrinos are a particular case of the model for $n_R = 3$. If we invoke a global lepton number symmetry, under which ν_R have charge $+1$, this forces $M_N = 0$, the singlets are exactly equivalent to the right-handed neutrinos in the Dirac case described in sec. 3.1. In the opposite limit $M_N \gg v$, the singlets can be integrated out and give rise to the Weinberg interaction as well as others at $d = 6$, etc. For intermediate M_N, the spectrum of this theory contains in general $3 + n_R$ Majorana neutrinos, which are admixtures of the active ones and the extra singlets.

It is easy to diagonalize the mass matrix in Eq. (11.9) in an expansion in m_D/M_N. The result to leading order in this expansion is

$$U^T \begin{pmatrix} 0 & m_D \\ m_D^T & M_N \end{pmatrix} U \simeq \begin{pmatrix} -m_D \frac{1}{M_N} m_D^T & 0 \\ 0 & M_N \end{pmatrix} + \mathcal{O}(\theta^2), \quad U = \begin{pmatrix} 1 & \theta \\ -\theta^\dagger & 1 \end{pmatrix}, \tag{11.11}$$

where

$$\theta = m_D^* \frac{1}{M_N}. \tag{11.12}$$

The matrix represents the active component of the heavy neutrino states and therefore controls their gauge interactions. To this order therefore the light neutrino and heavy neutrino masses are given by

$$m_l = \mathrm{Diag}\left[-m_D \frac{1}{M_N} m_D^T\right], \quad M_h = \mathrm{Diag}[M_N]. \tag{11.13}$$

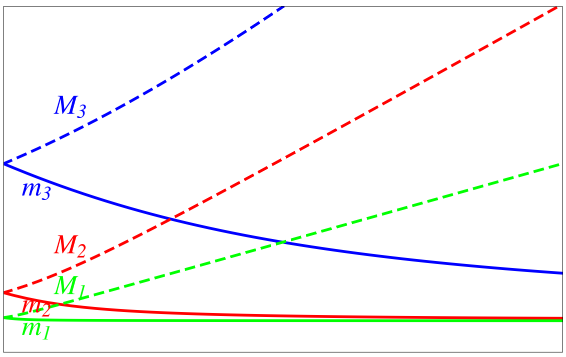


Fig. 49: Spectrum of the type I seesaw model for $n_R = 3$ as a function of a common M_N.

Figure 49 depicts the spectrum for the case of $n_R = 3$ as a function of a common M_N. In the limit $M_N \to 0$ the states degenerate in pairs to form Dirac fermions. As M_N increases three states get more massive proportional to M_N. These are often referred to as heavy neutral leptons (HNL), while three get lighter proportional to M_N^{-1}, as expected from the seesaw mechanism. The number of new free parameters is large. For the case $n_R = 3$ there are 18 fundamental parameters in the lepton sector: six of them are masses, six mixing angles and six phases. The counting of parameters for general n_R is shown in Table 4. Out of these 18 parameters we have determined only five: two mass differences and three neutrino mixing angles.

A very convenient parametrization in this model was introduced by Casas–Ibarra [93], which

Table 4: Number of physical parameters in the see-saw model with n families and the same number of right-handed Majorana neutrinos at high and low energies

	Yukawas	Field redefinitions	$No.\ m$	$No.\ \theta$	$No.\ \phi$
see-saw $E \geq M_i$	$Y_l, Y_\nu, M_R = M_R^T$ $5n^2 + n$	$U(n)^3$ $\frac{3(n^2-n)}{2}, \frac{3(n^2+n)}{2}$	$3n$	$n^2 - n$	$n^2 - n$
see-saw $E \ll M_i$	$Y_l, \alpha_\nu^T = \alpha_\nu$ $3n^2 + n$	$U(n)^2$ $n^2 - n, n^2 + n$	$2n$	$\frac{n^2-n}{2}$	$\frac{n^2-n}{2}$

allows to write in all generality (up to corrections of $\mathcal{O}(\theta^2)$) the Lagrangian parameters in terms of those of the light neutrino masses and mixings, and others related to the HNLs. In particular the phenomenology of this model depends on the spectrum of neutrino mass eigenstates, that we denote by $(\nu_1, \nu_2, \nu_3, N_1, N_2, ...N_{n_R})$, and their admixture in the flavour neutrino states :

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U_{ll} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} + U_{lh} \begin{pmatrix} N_1 \\ N_2 \\ .. \\ N_{n_R} \end{pmatrix} . \tag{11.14}$$

In the Casas–Ibarra parametrization we have

$$\begin{aligned} U_{ll} &= U_{\rm PMNS} + \mathcal{O}(\theta^2), \\ U_{lh} &= i U_{\rm PMNS} \sqrt{m_l} R \frac{1}{\sqrt{M_h}} + \mathcal{O}(\theta^2), \end{aligned} \tag{11.15}$$

where R is a general complex orthogonal matrix, $R^T R = 1$, which together with the heavy neutrino masses, M_h, parametrizes the parameter space inaccessible to neutrino oscillation experiments. Note that U_{ll} is the mixing matrix that we measure in neutrino oscillation experiments, assuming the heavy states are too heavy to play a role. This matrix is however no longer unitary,[3] but the unitarity violations are parametrically of $\mathcal{O}(\theta^2) \sim m_l/M_h$.

Equations (11.15) indicate that in this model there is a strong correlation between flavour mixings of the heavy states, U_{lh}, and the ratio of light-to-heavy neutrino masses. However the presence of the unknown matrix R, which is not bounded, implies that the naive seesaw scaling, $|U_{lh}|^2 \sim m_l/M_h$, that would hold exactly for one neutrino family, is far too naive for $n_R > 1$. In fact there are regions of parameter space where these mixings can be much larger than suggested by the naive scaling, and these are precisely the regions with more phenomenological interest, as we will see below.

Let us discuss some phenomenological implications of the different choices of the scale M_N.

[3]The Casas–Ibarra parametrization needs to be modified in the presence of large unitarity violations. A similar parametrization valid to all orders in θ is given in Ref. [94].

11.1.1 Neutrinoless double-beta decay

The amplitude for this process receives contributions from the light and heavy states:

$$m_{ee} \equiv \sum_{i=1}^{3} (U_{\rm PMNS})^2_{ei} m_i + \sum_{j=1}^{n_R} (U_{lh})^2_{ej} M_j \frac{\mathcal{M}^{\beta\beta0\nu}(M_j)}{\mathcal{M}^{\beta\beta0\nu}(0)}, \tag{11.16}$$

where the ratio of matrix elements $\mathcal{M}^{\beta\beta0\nu}$ for heavy and light mediators satisfy [95]:

$$\frac{\mathcal{M}^{\beta\beta0\nu}(M_j)}{\mathcal{M}^{\beta\beta0\nu}(0)} \propto \left(\frac{100\text{MeV}}{M_j}\right)^2, \quad M_j \to \infty. \tag{11.17}$$

If all the heavy state masses $\gg$ 100 MeV, the second term is suppressed and the amplitude contains only the light neutrino masses and mixings, which is constrained as shown before in Fig. 47. A plethora of experiments using different technologies have been proposed to reach a sensitivity in m_{ee} in the range of 10^{-2} eV , which could be sufficient to explore the full parameter space in the case of the IO. The importance of this measurement can hardly be overstated. A non-zero m_{ee} will imply that neutrinos are Majorana and therefore a new physics scale must exist, that lepton number is violated, and might give very valuable information on the lightest neutrino mass, and even help establishing the neutrino mass ordering. On the other hand, if the heavy states are not too heavy, within 100 MeV–few GeV, they could also contribute to the process significantly and even dominate over the light neutrino contribution for both orderings [96–98].

11.1.2 Cosmology and the seesaw scale

For $M_N \leq 100$ MeV, the heavy states in seesaw models can sizeably modify the history of the Universe: the abundance of light elements, the fluctuations in the CMB and the galaxy distribution at large scales. This is the case because these extra states contribute to the expansion either as a significant extra component of dark matter (Ω_m) or radiation ($\Delta N_{\rm eff}$).

The singlet states in this mass range are produced at T below the electroweak phase transition via mixing. The state i will reach thermal equilibrium if their interaction rate, $\Gamma_{s_i}(T)$, is larger than the Hubble parameter at some T. If this is the case, the extra species will contribute like one extra neutrino for $T > M_i$ or like an extra component of dark matter for $T < M_i$. The latest results from Planck strongly constrain an extra radiation component at CMB:

$$N_{\rm eff}(\text{CMB}) = 3.2 \pm 0.5. \tag{11.18}$$

and also measures the dark matter component to be $\Omega_m = 0.308 \pm 0.012$. Similar bounds are obtained from the abundance of light elements, BBN. These bounds exclude the possibility of having essentially any extra fully thermalized neutrino that is sufficiently long-lived to survive BBN. It can be shown that the ratio $\frac{\Gamma_{s_i}(T)}{H(T)}$ reaches a maximum at $T_{\rm max}$ [99, 100] and

$$\frac{\Gamma_{s_i}(T_{\rm max})}{H(T_{\rm max})} \sim \frac{\sum_\alpha |(U_{lh})_{\alpha i}|^2 M_i}{\sqrt{g_*(T_{\rm max})}}. \tag{11.19}$$

The naive seesaw scaling $U_{lh}^2 M_h \sim m_l$, would seem to imply that the thermalization condition depends only on the light neutrino masses and is independent on the seesaw scale. In fact a detailed study shows that indeed this naive expectation holds.

For $n_R = 2$, the heavy states must be $M_i \geq 100$ MeV [101], so that they might decay before BBN. For $n_R = 3$ two things can happen [102]. If the lightest neutrino mass, $m_{\text{lightest}} \geq 3 \times 10^{-3}$ eV, all the three heavy states thermalize and $M_i \geq 100$ MeV. If $m_{\text{lightest}} \leq 3 \times 10^{-3}$ eV two states must be above this limit, but one of the states with mass M_1 might not thermalize and therefore be sufficiently diluted. M_1 may take any value provided m_{lightest}, which is presently unconstrained, and is tuned accordingly.

11.1.3 Warm dark matter

For $m_{\text{lightest}} \leq 10^{-5}$ eV, M_1 might be $\mathcal{O}($ keV$)$, and a viable warm dark matter candidate [103, 104]. This scenario is the so-called νMSM model [104]. The most spectacular signal of this type of dark matter is a monochromatic X-ray line from the decay of this keV neutrino. There has been some evidence for an unexplained X-ray line in galaxy clusters that might be compatible with a 7 keV neutrino [105, 106]. These results are under intense scrutiny. If interpreted in terms of a keV neutrino, the mixing however is too small and some extra mechanism is needed to enhance the production so that it matches the required dark matter density, such as the presence of large primordial lepton asymmetries [107].

11.1.4 Direct searches for heavy neutral leptons

Naturalness arguments suggest that maybe the scale of M_N is not far from the electroweak scale. States with masses in this range could be produced in the lab [108]. The production of the HNL is mediated by charged or neutral currents or Higgs interactions with strength given by the U_{hl} coupling, see Fig. 50. The most important production mechanisms, from meson decays, at e^+e^- collisions at the Z peak or at hadron colliders, are shown in Fig. 51.

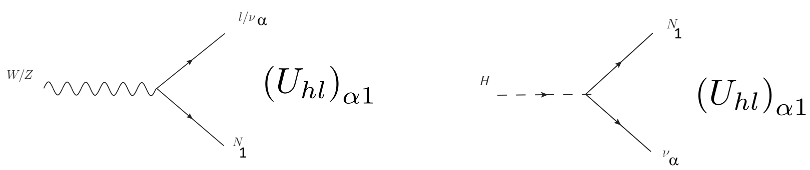


Fig. 50: Interactions of HNL in Type I seesaw model.

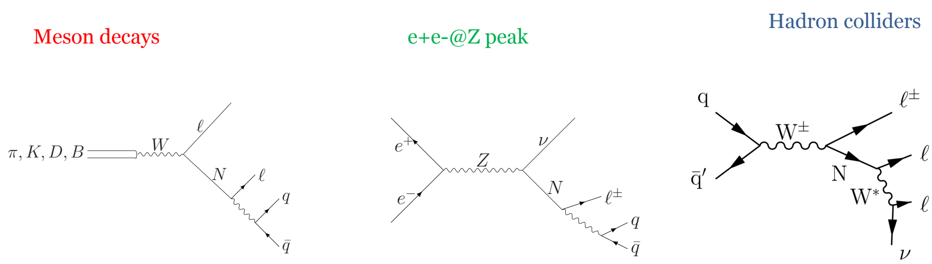


Fig. 51: Production processes of HNLs from meson decays, e^+e^- colliders and hadron colliders.

The present experimental bounds on the e mixings of these heavy states are shown in Figs. 52, on the plane $\sum_{\alpha=e,\mu,\tau} |(U_{hl})_{\alpha i}|^2$ versus M_i. The shaded regions correspond to existing constraints and the unshaded ones to prospects of various new experiments. For masses below a few GeV, the best constraints come from peak searches in meson decays. In particular the new beam dump experiment SHiP [109] can improve considerably the sensitivity in the region between the Kaon and B meson mass. Above the B meson mass and below the Z boson mass, searches in FCCee at the Z peak would improve present limits by several orders of magnitude [110]. The best existing limits in this range come from the LEP experiment DELPHI [111] and LHC searches from displaced vertices [112, 113]. The HNL in this range are very long lived and lead to displaced decays [114–116] that have a negligible SM background.

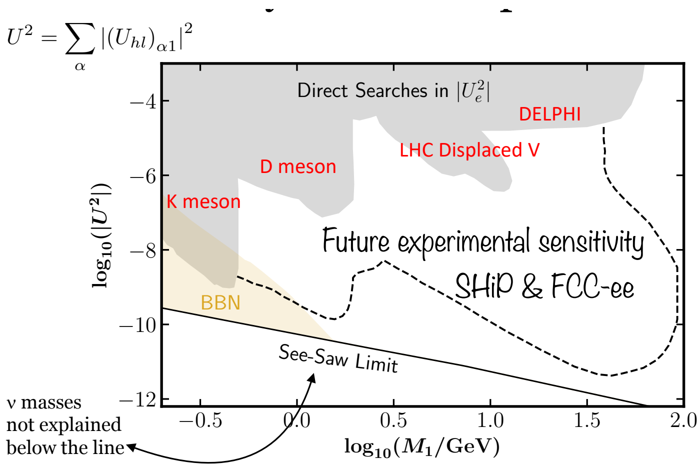


Fig. 52: Constraints from present and future experiments on a HNLs. Shaded regions are existing bounds on the HNL electron mixing as a function of the HNL mass, from the various processes that are sensitive to different mass ranges. The dashed line is the future sensitivity of SHiP at lower masses and FCCee at higher ones. Below the seesaw line neutrino masses cannot be explained. The exclusion from BBN is also added. Figure is courtesy of S. Sandner.

For masses above the W and Z masses, the best constraints are presently coming from LHC searches [117–119].

12 Low-scale leptogenesis

The Universe is made of matter. The matter–antimatter asymmetry is measured to be [61]

$$\eta_B \equiv \frac{N_b - N_{\bar{b}}}{N_\gamma} \sim 6.21(16) \times 10^{-10} \,. \tag{12.1}$$

One generic implication of neutrino mass models is that they provide a new mechanism to explain this asymmetry dynamically.

It has been known for a long time that all the ingredients to generate such an asymmetry from a symmetric initial state are present in the laws of particle physics. These ingredients were first put forward by Sakharov [120]:

1. Baryon number violation

$B + L$ is anomalous in the SM [121] both with and without massive neutrinos. At high T in the early

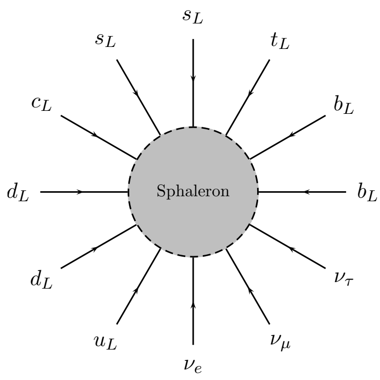


Fig. 53: Artistic view of a sphaleron.

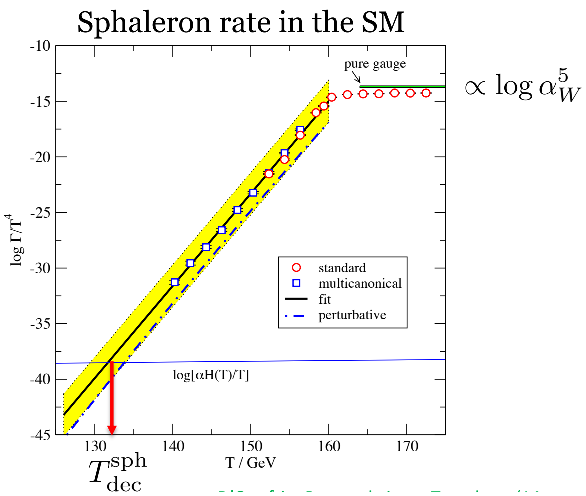


Fig. 54: Sphaleron rate in the SM normalized by T^4 as function of the temperature, from [123]. The horizontal line corresponds to the Hubble expansion rate.

Universe, $B + L$ violating transitions are in thermal equilibrium [122] due to the thermal excitation of configurations with topological charge called sphalerons, see Fig. 53.

These processes violate baryon and lepton numbers by the same amount:

$$\Delta B = \Delta L. \tag{12.2}$$

In seesaw models, there is generically an additional source of L violation (and $B - L$). If a lepton charge is generated at temperatures where the sphalerons are still in thermal equilibrium, a baryon charge can be generated.

The sphaleron rate in the SM has been computed accurately after the discovery of the Higgs boson [123]. The rate normalized to the fourth power of the temperature is shown in Fig. 54 around the electroweak phase transition. At $T \geq 160$GeV the rate is $\propto \alpha_W^5 T^4$, while it drops exponentially at lower temperatures. The Hubble rate is indicated by the horizontal line. The temperature where the sphaleron rate equals the Hubble expansion rate is the sphaleron decoupling temperature, $T_{\rm dec}^{\rm sph}$, below which no baryon number violation is possible.

2. C and CP violation

Any lepton or baryon asymmetry can only be generated if there is C and CP violation. Seesaw models generically include new sources of CP violation. As we have seen in type I seesaw model with $n_R = 3$ there are six new CP phases in the lepton sector. They can be absorbed in the Yukawa matrix, Y_ν of Eq. (11.8). Even though CP violation is connected to imaginary phases, CP violating observables such as the baryon asymmetry depends on many flavour parameters. A very useful concept is that of the flavour CP invariants [124]. Let us consider for example the minimal SM. Since quark Yukawa couplings are the only source of CP violation and they are small, we expect that any CP violating asymmetry generated at high temperatures (above the quark masses) can be expanded as a a polynomial in the up and down Yukawa couplings, Y_u and Y_d. Furthermore we expect that this polynomial is independent of the flavour basis used[4] and it is not real, that is it must have a non-zero imaginary part. The lowest order polynomial of Y_u and Y_d that satisfies these conditions is the famous Jarlskog invariant [124]

$$\Delta_{\rm CP}^{\rm quarks} = {\rm Im}\left[\det\left([Y_u Y_u^\dagger, Y_d Y_d^\dagger]\right)\right] \propto J \prod_{i<j}(m_{d_i}^2 - m_{d_j}^2)\prod_{i<j}(m_{u_i}^2 - m_{u_j}^2), \tag{12.3}$$

with

$$J \equiv {\rm Im}[V_{ij}^* V_{ii} V_{ji}^* V_{jj}] = c_{23}s_{23}c_{12}s_{12}c_{13}^2 s_{13}\sin\delta. \tag{12.4}$$

We can then naively estimate the baryon asymmetry generated at the EW transition in the SM as

$$Y_B \propto \frac{\Delta_{\rm CP}^{\rm quarks}}{T_{\rm EW}^{12}} \sim 10^{-20}, \tag{12.5}$$

where the denominator is fixed by dimensional analysis. This simple analysis shows that the CP violation in the minimal SM is far too small to explain the baryon asymmetry at the electroweak phase transition. A detailed computation arrives to the same conclusion [125].

In the case of the Type I seesaw extension of the SM we have also CP violation in the lepton sector encoded in the flavour parameters: Majorana mass matrix of the singlets, M_N, and the neutrino and charge lepton Yukawas, Y_ν and Y_l. The lowest order invariant involving Y_ν and M_N is [126, 127]:

$$\Delta_{\rm CP}^{\rm leptons} = {\rm Im}\left({\rm Tr}[Y_\nu^\dagger Y_\nu M^\dagger M M^* (Y_\nu^\dagger Y_\nu)^* M]\right), \tag{12.6}$$

or including also the lepton Yukawa

$$\tilde\Delta_{\rm CP}^{\rm leptons} = {\rm Im}\left({\rm Tr}[Y_\nu^\dagger Y_\nu M^\dagger M Y_\nu^\dagger Y_l Y_l^\dagger Y_\nu]\right) \equiv \sum_\alpha y_{l\alpha}^2 \Delta_\alpha. \tag{12.7}$$

Even at low scales, these invariants are potentially much larger that those in the quark sector [128].

3. Departure from thermal equilibrium

In order for a CP asymmetry to arise, it is necessary that the relevant processes occur out of thermal equilibrium since otherwise the abundances are fixed by the thermal Fermi–Dirac distributions and are

[4] A unitary rotation in flavour space of left and right chiral fields leaves all the terms in the Lagrangian invariant except the Yukawa couplings.

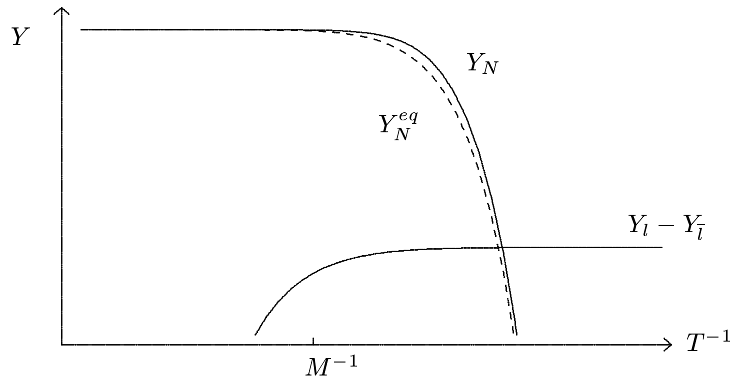


Fig. 55: High scale seesaw: abundance of the heavy Majorana singlets at the decoupling temperature and the lepton number generated in the decay.

equal for particles and antiparticles. Out-of-equilibrium conditions can happen in the evolution of the universe in the presence of first-order phase transitions, or due to the presence of sufficiently weakly coupled sectors that cannot keep up with the expansion of the universe. This happens when the interaction rates become smaller than the Hubble expansion rate, $\Gamma(T) \leq H(T)$. This must happen at T above the sphaleron decoupling, $T_{\rm dec}^{\rm sph}$ to be effective in generating baryons.

No particle in the minimal SM satisfies this condition within the standard cosmological model, not even neutrinos, that decouple much below sphaleron decoupling. On the other hand, the SM predicts the existence of a phase transition from the broken phase at low temperatures to a symmetric phase above, i.e. the EW phase transition. The critical temperature, $T_{\rm EW}$, is closely related to the sphaleron decoupling temperature. The EW transition has been shown to be a crossover transition and therefore with insufficient departure from thermal equilibrium [129].

In the Type I seesaw extension at low scales however, some of the states are more weakly interacting than neutrinos and therefore can fulfil the requirement $\Gamma_{N_i}(T) \leq H_u(T)$, for $T \geq T_{\rm dec}^{\rm sph}$.

In the high scale scenario $M_i \gg v$, the non-equilibrium condition is met at freeze out of the heavy neutrino states. These are thermally produced and freeze out at temperatures similar to their masses [128]. A net lepton asymmetry can be produced if the decay rate is slower than the expansion of the Universe at $T \sim M_i$, as shown in Fig. 55.

In contrast, in the low-scale scenario, for $M_i < v$, the out-of-equilibrium condition is met at freeze-in [104, 130, 131], that is some of the states never reach thermal equilibrium above $T_{\rm dec}^{\rm sph}$. A non-vanishing lepton and baryon asymmetry can survive and, if this is the case, sphaleron transitions can no longer wash it out. It turns out that these conditions can be met naturally in type I seesaw models for masses in the range $[0.1, 100]$ GeV. The relevant CP asymmetries arise in the production of the heavy seesaw states via the interference of CP-odd phases from the Yukawa couplings with CP-even phases from propagation and oscillations, see Fig. 56. A quantum treatment of the corresponding kinetic equations is mandatory in this case and quite complex.

A perturbative solution to the kinetic equations [132] allows to extract the analytical solution for Y_B in terms of the CP invariants, and using the Casas–Ibarra parametrization can then be expressed in terms of the neutrino masses and mixings, CP phases and HNL parameters.

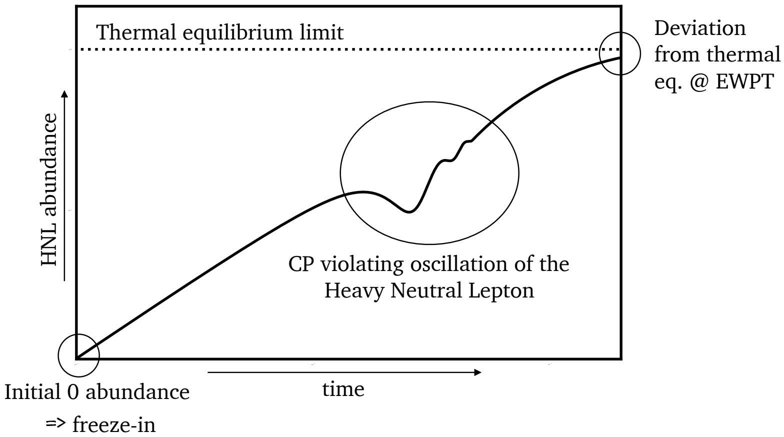


Fig. 56: Low-scale seesaw: abundance of the heavy Majorana singlets at T_{EW}.

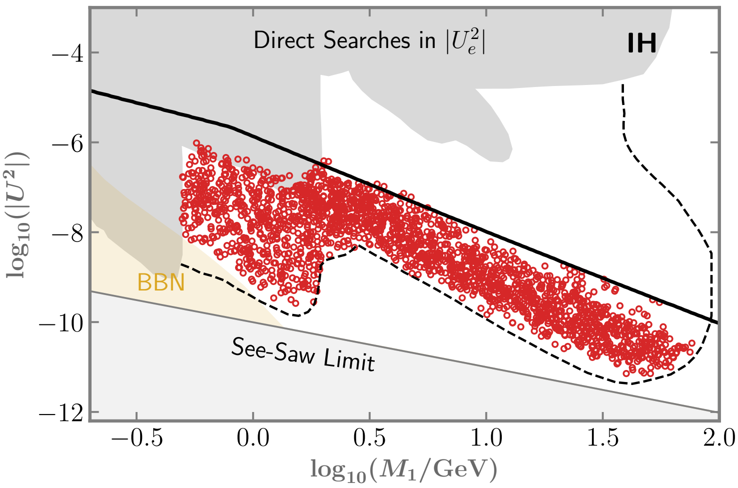


Fig. 57: Numerical scan of points on the plane of mixing versus mass of the HNL where the baryon asymmetry can be explained and within the sensitivity region of SHiP and/or FCCee (dashed line) in the minimal Type I seesaw ($n_R = 2$). The line is obtained analytically from a perturbative solution of the kinetic equations that can be expressed in terms of CP flavour invariants and maximized over unknown parameters. From Ref. [132].

In Fig. 57 we show the region on the plane U^2 v.s. M_i, where the baryon asymmetry and neutrino masses can be accounted for within the range of sensitivity of the future SHiP and FCCee projects. The solid line is the analytical upper bound to explain the baryon asymmetry based on the analytical solutions, and maximizing the asymmetry over the unknown parameters. This demonstrates the discovery potential of the future projects.

Other interesting correlations between Y_B and other observables are shown in Fig. 58. On the left, we show the HNL flavoured mixings for masses in the range accessible to FCC when neutrino masses are explained for both hierarchies. On the right plot the constrain of generating the correct baryon asymmetry is added. The correct baryon asymmetry therefore restricts the flavour of the HNL mixings as well as the PMNS CP phases (only two for $n_R = 2$), as shown in Fig. 59 for both hierarchies.

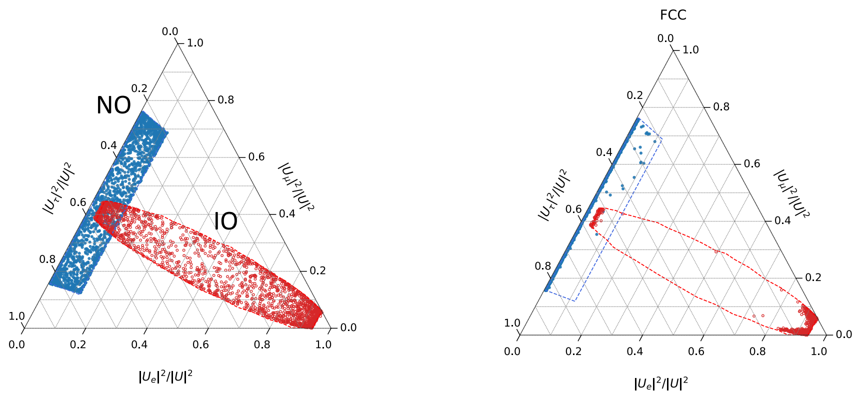


Fig. 58: Normalized mixings to e, μ, τ of HNLs with masses in the range of FCCee and mild degeneracy. Only the constrain from neutrino masses is imposed on the left plot, while also the Y_B is imposed on the right plot. The two regions correspond to neutrino orderings. From Ref. [132].

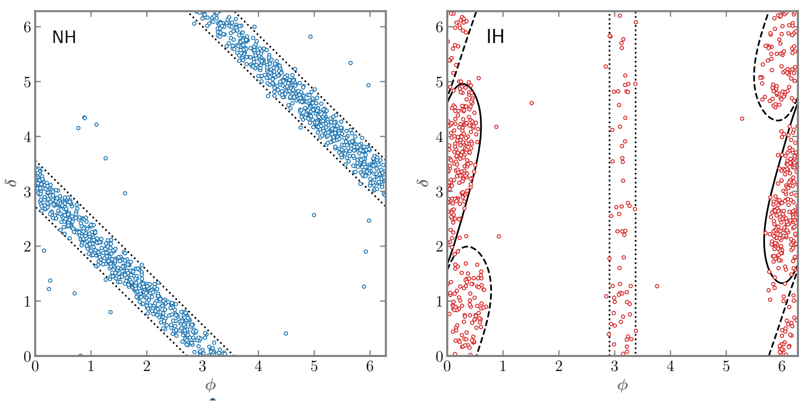


Fig. 59: Numerical scan of points that explain Y_B on the plane of the Dirac CP violating phase, δ, and the Majorana phase, ϕ in the minimal Type I seesaw ($n_R = 2$). From Ref. [132].

An interesting question is whether the baryon asymmetry can be predicted quantitatively from the measurements of CP violation in neutrino oscillations or from the CP violation in the neutrino mass matrix. Unfortunately this is not the case generically, because the asymmetry depends on more parameters than those in the light neutrino mass matrix. However, if the model is sufficiently constrained very strong correlations can occur.

For example, in the minimal Type I seesaw model, $n_R = 2$, and in the assumption that the two eigenvalues of the matrix M_N are degenerate, there are only two physical CP violating phases, that can then be parametrized by the two in the light neutrino mass matrix. They determine both Y_B and CP violation in neutrino oscillations. In this case, the measurement of the HNL mixings to electrons, muons and τ's can pin down the CP phase in neutrino oscillations and Y_B up to discrete degeneracies, as shown in Fig. 60. If the phase δ is also measured, a prediction of Y_B is possible from laboratory measurements.

This simple example demonstrates the interplay between Y_B and other observables in neutrino physics.

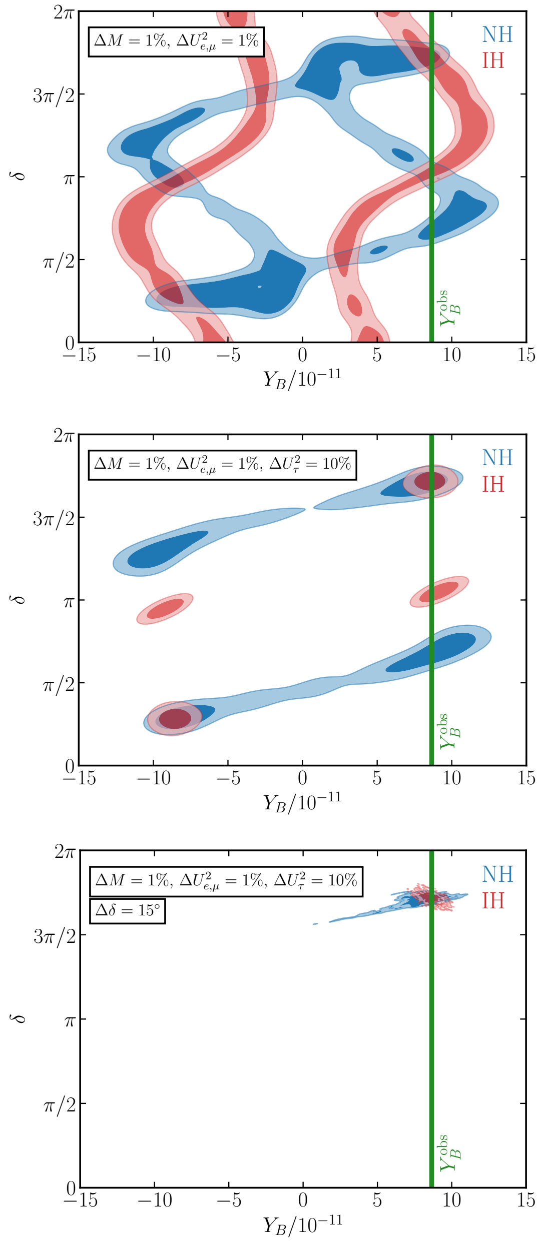


Fig. 60: Assuming the minimal Type I seesaw model with $n_R = 2$, and degenerate singlets within the FCCee range, with parameters that can explain neutrino masses and Y_B. Upper Plot: determination of δ and Y_B from a putative measurement of HNL mixings to electrons and muons and masses with accuracies as indicated, and for NO (blue) and IO(blue). Middle Plot: adding also a measurement of the HNL mixing to τ's. Bottom Plot: adding also a measurement of the phase δ from future neutrino oscillation experiments. From Ref. [133].

13 Conclusions

The results of many beautiful experiments in the last decade have demonstrated that neutrinos are massive and mix. The standard 3ν scenario can explain all available data, except that of the unconfirmed signal of LSND. The lepton flavour sector of the Standard Model is expected to be at least as complex as the quark one, even though we know it only partially.

The structure of the neutrino spectrum and mixing is quite different from the one that has been observed for the quarks: there are large leptonic mixing angles and the neutrino masses are much smaller than those of the remaining leptons. These peculiar features of the lepton sector strongly suggest that leptons and quarks constitute two complementary approaches to understanding the origin of flavour in the Standard Model. In fact, the smallness of neutrino masses can be naturally understood if there is new physics beyond the electroweak scale.

Many fundamental questions remain to be answered in future neutrino experiments, and these can have very important implications for our understanding of the Standard Model and of what lies beyond: Are neutrinos Majorana particles? Are neutrino masses the result of a new physics scale? Is CP violated in the lepton sector? Could neutrinos be the seed of the matter–antimatter asymmetry in the Universe?

A rich experimental programme lies ahead where fundamental physics discoveries are very likely (almost warranted). We can only hope that neutrinos will keep up with their old tradition and provide a window to what lies beyond the Standard Model.

References

[1] E. Fermi. Trends to a theory of β radiation. (In Italian). *Nuovo Cim.* **11** (1934) 1–19, doi:10.1007/BF02959820.

[2] H. Bethe and R. Peierls. The "neutrino". *Nature*, **133** (1934) 532, doi:10.1038/133532a0.

[3] B. Pontecorvo. Inverse β process, reprinted in K. Winter, *Neutrino physics* 2nd ed. (Cambridge Univ. Press, Cambridge, 2008), pp. 23–28, version from Centro Pontecorvo.

[4] F. Reines and C. L. Cowan. The neutrino. *Nature* **178** (1956) 446–449, doi:10.1038/178446a0.

[5] C. L. Cowan *et al.*, Detection of the free neutrino: A confirmation, *Science* **124** (1956) 103–104, doi:10.1126/science.124.3212.103.

[6] G. Danby *et al.*, Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos, *Phys. Rev. Lett.* **9** (1962) 36–44, doi:10.1103/PhysRevLett.9.36.

[7] M. Aker *et al.*, Direct neutrino-mass measurement with sub-electronvolt sensitivity, *Nature Phys.* **18**(2) (2022) 160–166, doi:10.1038/s41567-021-01463-1.

[8] K. Assamagan *et al.*, Upper limit of the muon-neutrino mass and charged pion mass from momentum analysis of a surface muon beam, *Phys. Rev.* **D53** (1996) 6065–6077, doi:10.1103/PhysRevD.53.6065.

[9] R. Barate *et al.*, An upper limit on the tau-neutrino mass from three-prong and five-prong tau decays, *Eur. Phys. J.* **C2** (1998) 395–406, doi:10.1007/s100529800850.

[10] S. Schael *et al.*, Precision electroweak measurements on the Z resonance, *Phys. Rept.* **427** (2006) 257–454, doi:10.1016/j.physrep.2005.12.006.

[11] P. Janot and S. Jadach, Improved Bhabha cross section at LEP and the number of light neutrino species, *Phys. Lett.* **B803** (2020) 135319, doi:10.1016/j.physletb.2020.135319.

[12] G. Voutsinas *et al.*, Beam-beam effects on the luminosity measurement at LEP and the number of light neutrino species, *Phys. Lett.* **B800** (2020) 135068, doi:10.1016/j.physletb.2019.135068.

[13] S. Weinberg, Baryon and lepton nonconserving processes, *Phys. Rev. Lett.* **43** (1979) 1566–1570, doi:10.1103/PhysRevLett.43.1566.

[14] S. Weinberg, Phenomenological Lagrangians, *Physica* **A96** (1979) 327–340, doi:10.1016/0378-4371(79)90223-1.

[15] W. Buchmuller and D. Wyler, Effective Lagrangian analysis of new interactions and flavor conservation, *Nucl. Phys.* **B268** (1986) 621–653, doi:10.1016/0550-3213(86)90262-2.

[16] Z. Maki, M. Nakagawa, and S. Sakata, Remarks on the unified model of elementary particles, *Prog. Theor. Phys.* **28** (1962) 870–880, doi:10.1143/PTP.28.870.

[17] B. Pontecorvo, Neutrino experiments and the problem of conservation of leptonic charge, *Sov. Phys. JETP* **26** (1968) 984–988 [*Zh. Eksp. Teor. Fiz.* **53** (1967) 1717], Inspire.

[18] P.A. Zyla *et al.*, Review of particle physics, *PTEP* **2020**(8) (2020) 083C01, doi:10.1093/ptep/ptaa104.

[19] B. Pontecorvo, Mesonium and anti-mesonium, *Sov. Phys. JETP* **6** (1957) 429 [*Zh. Eksp. Teor. Fiz.* **33** (1957) 549], Inspire.

[20] E.K. Akhmedov and A.Yu. Smirnov, Paradoxes of neutrino oscillations, *Phys. Atom. Nucl.* **72** (2009) 1363–1381, doi:10.1134/S1063778809080122, arXiv:0905.1903.

[21] E.K. Akhmedov and J. Kopp, Neutrino oscillations: quantum mechanics vs. quantum field theory, *JHEP* **04** (2010) 008, doi:10.1007/JHEP04(2010)008, [Erratum: *JHEP* **10** (2013) 052, doi:10.1007/JHEP10(2013)052].

[22] M. Cerdá, Neutrino oscillations in quantum field theory. Master thesis, University of Valencia, 2011.

[23] L. Wolfenstein, Neutrino oscillations in matter, *Phys. Rev.* **D17** (1978) 2369–2374, doi:10.1103/PhysRevD.17.2369.

[24] S.P. Mikheev and A.Yu. Smirnov, Resonance amplification of oscillations in matter and spectroscopy of solar neutrinos, *Sov. J. Nucl. Phys.* **42** (1985) 913–917 [Yad. Fiz. **42**, 1441 (1985)].

[25] H.A. Bethe, Energy production in stars, *Phys. Rev.* **55** (1939) 434–456, doi:10.1103/PhysRev.55.434.

[26] J.N. Bahcall, M.H. Pinsonneault, and S. Basu, Solar models: current epoch and time dependences, neutrinos, and helioseismological properties, *Astrophys. J.* **555** (2001) 990–1012, doi:10.1086/321493.

[27] B.T. Cleveland *et al.*, Measurement of the solar electron neutrino flux with the Homestake chlorine detector, *Astrophys. J.* **496** (1998) 505–526, doi:10.1086/305343.

[28] W. Hampel *et al.*, GALLEX solar neutrino observations: results for GALLEX IV, *Phys. Lett.* **B447** (1999) 127–133, doi:10.1016/S0370-2693(98)01579-2.

[29] J.N. Abdurashitov *et al.*, Solar neutrino flux measurements by the Soviet-American Gallium Experiment (SAGE) for half the 22 year solar cycle, *J. Exp. Theor. Phys.* **95** (2002) 181–193 [*Zh. Eksp. Teor. Fiz.* **122** (2002) 211], doi:10.1134/1.1506424, arXiv:astro-ph/0204245.

[30] Y. Fukuda *et al.*, Solar neutrino data covering solar cycle 22, *Phys. Rev. Lett.* **77** (1996) 1683–1686, doi:10.1103/PhysRevLett.77.1683.

[31] Y. Fukuda *et al.*, Measurement of the solar neutrino energy spectrum using neutrino electron scattering, *Phys. Rev. Lett.* **82** (1999) 2430–2434, doi:10.1103/PhysRevLett.82.2430, arXiv:hep-ex/9812011.

[32] J. Hosaka *et al.*, Solar neutrino measurements in Super-Kamiokande-I, *Phys. Rev.* **D73** (2006) 112001, doi:10.1103/PhysRevD.73.112001, arXiv:hep-ex/0508053.

[33] Q.R. Ahmad *et al.*, Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory, *Phys. Rev. Lett.* **87** (2001) 071301, doi:10.1103/PhysRevLett.87.071301, arXiv:nucl-ex/0106015.

[34] Q.R. Ahmad *et al.*, Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory, *Phys. Rev. Lett.* **89** (2002) 011301, doi:10.1103/PhysRevLett.89.011301, arXiv:nucl-ex/0204008.

[35] K. Eguchi *et al.*, First results from KamLAND: evidence for reactor anti-neutrino disappearance, *Phys. Rev. Lett.* **90** (2003) 021802, doi:10.1103/PhysRevLett.90.021802, arXiv:hep-ex/0212021.

[36] S. Abe *et al.*, Precision measurement of neutrino oscillation parameters with KamLAND, *Phys. Rev. Lett.* **100** (2008) 221803, doi:10.1103/PhysRevLett.100.221803, arXiv:0801.4589.

[37] G. Bellini *et al.*, Final results of Borexino Phase-I on low energy solar neutrino spectroscopy, *Phys. Rev.* **D89** (2014) 112007, doi:10.1103/PhysRevD.89.112007, arXiv:1308.0443.

[38] M. Honda *et al.*, A new calculation of the atmospheric neutrino flux in a 3-dimensional scheme, *Phys. Rev. D* **70** (2004) 043008, doi:10.1103/PhysRevD.70.043008, arXiv:astro-ph/0404457.

[39] Y. Fukuda *et al.*, Evidence for oscillation of atmospheric neutrinos, *Phys. Rev. Lett.* **81** (1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.

[40] Y. Ashie *et al.*, Measurement of atmospheric neutrino oscillation parameters by Super-Kamiokande I, *Phys. Rev.* **D71** (2005) 112005, doi:10.1103/PhysRevD.71.112005, arXiv:hep-ex/0501064.

[41] Y. Ashie *et al.*, Evidence for an oscillatory signature in atmospheric neutrino oscillation, *Phys. Rev. Lett.* **93** (2004) 101801, doi:10.1103/PhysRevLett.93.101801, arXiv:hep-ex/0404034.

[42] M.H. Ahn *et al.*, Measurement of neutrino oscillation by the K2K experiment, *Phys. Rev.* **D74** (2006) 072003, doi:10.1103/PhysRevD.74.072003, arXiv:hep-ex/0606032.

[43] D.G. Michael *et al.*, Observation of muon neutrino disappearance with the MINOS detectors and the NuMI neutrino beam, *Phys. Rev. Lett.* **97** (2006) 191801, doi:10.1103/PhysRevLett.97.191801, arXiv:hep-ex/0607088.

[44] N. Agafonova *et al.*, Discovery of τ neutrino appearance in the CNGS neutrino beam with the OPERA experiment, *Phys. Rev. Lett.* **115** (2015) 121802, doi:10.1103/PhysRevLett.115.121802, arXiv:1507.01417.

[45] A. Holin, Results from the MINOS experiment and new MINOS+ data, *PoS* **NUFACT2014** (2014) 028, doi:10.22323/1.226.0028, arXiv:1507.08564.

[46] F.P. An *et al.*, Observation of electron-antineutrino disappearance at Daya Bay, *Phys. Rev. Lett.* **108** (2012) 171803, doi:10.1103/PhysRevLett.108.171803, arXiv:1203.1669.

[47] J.K. Ahn *et al.*, Observation of reactor electron antineutrino disappearance in the RENO experiment, *Phys. Rev. Lett.* **108** (2012) 191802, doi:10.1103/PhysRevLett.108.191802.

[48] Y. Abe *et al.*, Indication for the disappearance of reactor electron antineutrinos in the Double Chooz experiment, *Phys. Rev. Lett.* **108** (2012) 131801, doi:10.1103/PhysRevLett.108.131801.

[49] K. Abe *et al.*, Observation of electron neutrino appearance in a muon neutrino beam, *Phys. Rev. Lett.* **112** (2014) 061802, doi:10.1103/PhysRevLett.112.061802.

[50] P. Adamson *et al.*, First measurement of electron neutrino appearance in NOvA, *Phys. Rev. Lett.* **116** (2016) 151806, doi:10.1103/PhysRevLett.116.151806.

[51] I. Esteban *et al.*, The fate of hints: updated global analysis of three-flavor neutrino oscillations, *JHEP* **09** (2020) 178, doi:10.1007/JHEP09(2020)178.

[52] P.F. de Salas *et al.*, 2020 global reassessment of the neutrino oscillation picture, *JHEP* **02** (2021) 071, doi:10.1007/JHEP02(2021)071.

[53] F. Capozzi *et al.*, Global constraints on absolute neutrino masses and their ordering, *Phys. Rev. D* **95** (2017) 096014, doi:10.1103/PhysRevD.95.096014. [Addendum: *Phys. Rev.* **D101** (2020) 116013, doi:10.1103/PhysRevD.101.116013].

[54] S.K. Agarwalla and P. Hernandez, Probing the neutrino mass hierarchy with Super-Kamiokande, *JHEP* **10** (2012) 086, doi:10.1007/JHEP10(2012)086.

[55] S.T. Petcov and M. Piai, The LMA MSW solution of the solar neutrino problem, inverted neutrino mass hierarchy and reactor neutrino experiments, *Phys. Lett.* **B533** (2002) 94–106, doi:10.1016/S0370-2693(02)01591-5.

[56] S. Choubey, S.T. Petcov, and M. Piai, Precision neutrino oscillation physics with an intermediate baseline reactor neutrino experiment, *Phys. Rev.* **D68** (2003) 113006, doi:10.1103/PhysRevD.68.113006.

[57] F. An *et al.*, Neutrino physics with JUNO, *J. Phys.* **G43** (2016) 030401, doi:10.1088/0954-3899/43/3/030401.

[58] A. Cervera *et al.*, Golden measurements at a neutrino factory, *Nucl. Phys.* **B579** (2000) 17–55, doi:10.1016/S0550-3213(00)00221-2. [Erratum: *Nucl. Phys.* **B593** (2001) 731, doi:10.1016/S0550-3213(00)00606-4].

[59] J. Bian *et al.*, Hyper-Kamiokande experiment: A Snowmass white paper, in *Snowmass 2021*, March 2022, doi:10.48550/arXiv.2203.02029.

[60] B. Abi *et al.*, Long-baseline neutrino oscillation physics potential of the DUNE experiment, *Eur. Phys. J.* **C80** (2020) 978, doi:10.1140/epjc/s10052-020-08456-z.

[61] N. Aghanim *et al.*, Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641** (2020) A6, doi:10.1051/0004-6361/201833910. [Erratum: *Astron. Astrophys.* **652** (2021) C4, doi:10.1051/0004-6361/201833910e].

[62] A. Aguilar-Arevalo *et al.*, Evidence for neutrino oscillations from the observation of $\bar{\nu}_e$ appearance in a $\bar{\nu}_\mu$ beam, *Phys. Rev.* **D64** (2001) 112007, doi:10.1103/PhysRevD.64.112007.

[63] B. Armbruster *et al.*, Upper limits for neutrino oscillations $\bar{\nu}_\mu \to \bar{\nu}_e$ from muon decay at rest, *Phys. Rev.* **D65** (2002) 112001, doi:10.1103/PhysRevD.65.112001.

[64] A.A. Aguilar-Arevalo *et al.*, Improved search for $\bar{\nu}_\mu \to \bar{\nu}_e$ oscillations in the MiniBooNE experiment, *Phys. Rev. Lett.* **110** (2013) 161801, doi:10.1103/PhysRevLett.110.161801.

[65] A.A. Aguilar-Arevalo *et al.*, Significant excess of electron-like events in the MiniBooNE short-baseline neutrino experiment, *Phys. Rev. Lett.* **121** (2018) 221801, doi:10.1103/PhysRevLett.121.221801.

[66] P. Abratenko *et al.*, Search for an excess of electron neutrino interactions in MicroBooNE using multiple final-state topologies, *Phys. Rev. Lett.* **128** (2022) 241801, doi:10.1103/PhysRevLett.128.241801.

[67] T.A. Mueller *et al.*, Improved predictions of reactor antineutrino spectra, *Phys. Rev.* **C83** (2011) 054615, doi:10.1103/PhysRevC.83.054615.

[68] G. Mention *et al.*, The reactor antineutrino anomaly, *Phys. Rev.* **D83** (2011) 073006, doi:10.1103/PhysRevD.83.073006.

[69] P. Huber, Determination of antineutrino spectra from nuclear reactors, *Phys. Rev.* **C84** (2011) 024617, doi:10.1103/PhysRevC.84.024617. [Erratum: *Phys. Rev.* **C85** (2012) 029901, doi:10.1103/PhysRevC.85.029901].

[70] F.P. An *et al.*, Measurement of the reactor antineutrino flux and spectrum at Daya Bay, *Phys. Rev. Lett.* **116** (2016) 061801, doi:10.1103/PhysRevLett.116.061801.

[71] P. Adamson *et al.*, Improved constraints on sterile neutrino mixing from disappearance searches in the MINOS, MINOS+, Daya Bay, and Bugey-3 experiments, *Phys. Rev. Lett.* **125** (2020) 071801, doi:10.1103/PhysRevLett.125.071801.

[72] C. Giunti *et al.*, Gallium anomaly: critical view from the global picture of ν_e and $\bar{\nu}_e$ disappearance, *JHEP* **10** (2022) 164, doi:10.1007/JHEP10(2022)164.

[73] R.L. Workman *et al.*, Review of particle physics, *PTEP* **2022** (2022) 083C01, doi:10.1093/ptep/ptac097.

[74] P.F. Harrison, D.H. Perkins, and W.G. Scott, Tri-bimaximal mixing and the neutrino oscillation data, *Phys. Lett.* **B530** (2002) 167, doi:10.1016/S0370-2693(02)01336-9.

[75] F. Vissani, Do experiments suggest a hierarchy problem? *Phys. Rev.* **D57** (1998) 7027–7030, doi:10.1103/PhysRevD.57.7027.

[76] J.A. Casas, J.R. Espinosa, and I. Hidalgo, Implications for new physics from fine-tuning arguments. 1. Application to SUSY and seesaw cases, *JHEP* **11** (2004) 057, doi:10.1088/1126-6708/2004/11/057.

[77] G. Aad *et al.*, Combined measurement of the Higgs boson mass in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS experiments, *Phys. Rev. Lett.* **114** (2015) 191803, doi:10.1103/PhysRevLett.114.191803.

[78] G. Degrassi *et al.*, Higgs mass and vacuum stability in the Standard Model at NNLO, *JHEP* **08** (2012) 098, doi:10.1007/JHEP08(2012)098.

[79] P. Minkowski, $\mu \to e\gamma$ at a rate of one out of 10^9 muon decays?, *Phys. Lett.* **B67** (1977) 421–428, doi:10.1016/0370-2693(77)90435-X.

[80] M. Gell-Mann, P. Ramond, and R. Slansky. Complex spinors and unified theories, *Conf. Proc.* **C790927** (1979) 315–321, doi:10.48550/arXiv.1306.4669.

[81] T. Yanagida. Horizontal gauge symmetry and masses of neutrinos, *Conf. Proc.* **C7902131** (1979) 95–99, Inspire.

[82] R.N. Mohapatra and G. Senjanovic, Neutrino mass and spontaneous parity violation, *Phys. Rev. Lett.* **44** (1980) 912, doi:10.1103/PhysRevLett.44.912.

[83] M. Magg and C. Wetterich, Neutrino mass problem and gauge hierarchy, *Phys. Lett.* **B94** (1980) 61, doi:10.1016/0370-2693(80)90825-4.

[84] J. Schechter and J.W.F. Valle, Neutrino masses in SU(2) $\otimes$ U(1) theories, *Phys. Rev.* **D22** (1980) 2227, doi:10.1103/PhysRevD.22.2227.

[85] C. Wetterich, Neutrino masses and the scale of $B - L$ violation, *Nucl. Phys.* **B187** (1981) 343, doi:10.1016/0550-3213(81)90279-0.

[86] G. Lazarides, Q. Shafi, and C. Wetterich, Proton lifetime and fermion masses in an SO(10) model, *Nucl. Phys.* **B181** (1981) 287–300, doi:10.1016/0550-3213(81)90354-0.

[87] R.N. Mohapatra and G. Senjanovic, Neutrino masses and mixings in gauge models with spontaneous parity violation, *Phys. Rev.* **D23** (1981) 165, doi:10.1103/PhysRevD.23.165.

[88] R. Foot *et al.*, Seesaw neutrino masses induced by a triplet of leptons, *Z. Phys.* **C44** (1989) 441, doi:10.1007/BF01415558.

[89] E. Ma, Pathways to naturally small neutrino masses, *Phys. Rev. Lett.* **81** (1998) 1171–1174, doi:10.1103/PhysRevLett.81.1171.

[90] A. Zee, A theory of lepton number violation, neutrino Majorana mass, and oscillation, *Phys. Lett.* **B93** (1980) 389, doi:10.1016/0370-2693(80)90349-4. [Erratum: *Phys. Lett.* **B95** (1980) 461, doi:10.1016/0370-2693(80)90193-8].

[91] A. Zee, Charged scalar field and quantum number violations, *Phys. Lett.* **B161** (1985) 141, doi:10.1016/0370-2693(85)90625-2.

[92] K.S. Babu, Model of "calculable" Majorana neutrino masses, *Phys. Lett.* **B203** (1988) 132, doi:10.1016/0370-2693(88)91584-5.

[93] J.A. Casas and A. Ibarra, Oscillating neutrinos and $\mu \to e\gamma$, *Nucl. Phys.* **B618** (2001) 171–204, doi:10.1016/S0550-3213(01)00475-8.

[94] A. Donini *et al.*, The minimal $3+2$ neutrino model versus oscillation anomalies, *JHEP* **07** (2012) 161, doi:10.1007/JHEP07(2012)161.

[95] M. Blennow *et al.*, Neutrinoless double beta decay in seesaw models, *JHEP* **07** (2010) 096, doi:10.1007/JHEP07(2010)096.

[96] A. Ibarra, E. Molinaro, and S.T. Petcov, TeV scale see-saw mechanisms of neutrino mass generation, the Majorana nature of the heavy singlet neutrinos, and $(\beta\beta)_{0\nu}$-decay, *JHEP* **09** (2010) 108, doi:10.1007/JHEP09(2010)108.

[97] M. Mitra, G. Senjanovic, and F. Vissani, Neutrinoless double beta decay and heavy sterile neutrinos, *Nucl. Phys.* **B856** (2012) 26–73, doi:10.1016/j.nuclphysb.2011.10.035.

[98] J. Lopez-Pavon, E. Molinaro, and S.T. Petcov, Radiative corrections to light neutrino masses in low-scale type I seesaw scenarios and neutrinoless double beta decay, *JHEP* **11** (2015) 030, doi:10.1007/JHEP11(2015)030.

[99] R. Barbieri and A. Dolgov, Bounds on sterile-neutrinos from nucleosynthesis, *Phys. Lett.* **B237** (1990) 440–445, doi:10.1016/0370-2693(90)91203-N.

[100] K. Kainulainen, Light singlet neutrinos and the primordial nucleosynthesis, *Phys. Lett.* **B244** (1990) 191–195, doi:10.1016/0370-2693(90)90054-A.

[101] P. Hernandez, M. Kekic, and J. Lopez-Pavon, Low-scale seesaw models versus $N_{\rm eff}$, *Phys. Rev.* **D89** (2014) 073009, doi:10.1103/PhysRevD.89.073009.

[102] P. Hernandez, M. Kekic, and J. Lopez-Pavon, $N_{\rm eff}$ in low-scale seesaw models versus the lightest neutrino mass, *Phys. Rev.* **D90** (2014) 065033, doi:10.1103/PhysRevD.90.065033.

[103] S. Dodelson and L.M. Widrow, Sterile-neutrinos as dark matter, *Phys. Rev. Lett.* **72** (1994) 17–20, doi:10.1103/PhysRevLett.72.17.

[104] T. Asaka and M. Shaposhnikov, The νMSM, dark matter and baryon asymmetry of the universe, *Phys. Lett.* **B620** (2005) 17–26, doi:10.1016/j.physletb.2005.06.020.

[105] E. Bulbul *et al.*, Detection of an unidentified emission line in the stacked X-ray spectrum of galaxy clusters, *Astrophys. J.* **789** (2014) 13, doi:10.1088/0004-637X/789/1/13.

[106] A. Boyarsky *et al.*, Unidentified line in X-ray spectra of the Andromeda galaxy and Perseus galaxy cluster, *Phys. Rev. Lett.* **113** (2014) 251301, doi:10.1103/PhysRevLett.113.251301.

[107] X.-D. Shi and G.M. Fuller, A new dark matter candidate: Nonthermal sterile neutrinos, *Phys. Rev. Lett.* **82** (1999) 2832–2835, doi:10.1103/PhysRevLett.82.2832.

[108] A. Atre *et al.*, The search for heavy Majorana neutrinos, *JHEP* **05** (2009) 030, doi:10.1088/1126-6708/2009/05/030.

[109] C. Ahdida *et al.*, Sensitivity of the SHiP experiment to heavy neutral leptons, *JHEP* **04** (2019) 077, doi:10.1007/JHEP04(2019)077.

[110] A. Blondel *et al.*, Search for heavy right-handed neutrinos at the FCC-ee, *Nucl. Part. Phys. Proc.* **273-275** (2016) 1883–1890, doi:10.1016/j.nuclphysbps.2015.09.323.

[111] P. Abreu *et al.*, Search for neutral heavy leptons produced in Z decays, *Z. Phys.* **C74** (1997) 57–71, doi:10.1007/s002880050459. [Erratum: *Z. Phys.* **C75**, 580 (1997)].

[112] G. Aad *et al.*, Search for heavy neutral leptons in decays of W bosons using a dilepton displaced vertex in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector, *Phys. Rev. Lett.* **131** (2023) 061803 doi:10.1103/PhysRevLett.131.061803.

[113] A. Tumasyan *et al.*, Search for long-lived heavy neutral leptons with displaced vertices in proton-proton collisions at $\sqrt{s} = 13$ TeV, *JHEP* **07** (2022) 081, doi:10.1007/JHEP07(2022)081.

[114] J.C. Helo, M. Hirsch, and S. Kovalenko, Heavy neutrino searches at the LHC with displaced vertices, *Phys. Rev.* **D89** (2014) 073005, doi:10.1103/PhysRevD.89.073005.

[115] E. Izaguirre and B. Shuve, Multilepton and lepton jet probes of sub-weak-scale right-handed neutrinos, *Phys. Rev.* **D91** (2015) 093010, doi:10.1103/PhysRevD.91.093010.

[116] A.M. Gago *et al.*, Probing the Type I seesaw mechanism with displaced vertices at the LHC, *Eur. Phys. J.* **C75** (2015) 470, doi:10.1140/epjc/s10052-015-3693-1.

[117] F. del Aguila and J. A. Aguilar-Saavedra, Distinguishing seesaw models at LHC with multi-lepton signals, *Nucl. Phys.* **B813** (2009) 22–90, doi:10.1016/j.nuclphysb.2008.12.029.

[118] G. Aad *et al.*, Search for heavy neutral leptons in decays of W bosons produced in 13 TeV pp collisions using prompt and displaced signatures with the ATLAS detector, *JHEP* **10** (2019) 265, doi:10.1007/JHEP10(2019)265.

[119] A.M. Sirunyan *et al.*, Search for heavy neutral leptons in events with three charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV, *Phys. Rev. Lett.* **120** (2018) 221801, doi:10.1103/PhysRevLett.120.221801.

[120] A.D. Sakharov, Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32–35, [*Usp. Fiz. Nauk* **161**(1991) 61–64, *Sov. Phys. Usp.* **34** (1991) 392–393, doi:10.1070/PU1991v034n05ABEH002497].

[121] G. 't Hooft, Symmetry breaking through Bell–Jackiw anomalies, *Phys. Rev. Lett.* **37** (1976) 8–11, doi:10.1103/PhysRevLett.37.8.

[122] V.A. Kuzmin, V.A. Rubakov, and M.E. Shaposhnikov, On the anomalous electroweak baryon number nonconservation in the early universe, *Phys. Lett.* **B155** (1985) 36, doi:10.1016/0370-2693(85)91028-7.

[123] M. D'Onofrio, K. Rummukainen, and A. Tranberg, Sphaleron rate in the minimal Standard Model, *Phys. Rev. Lett.* **113** (2014) 141602, doi:10.1103/PhysRevLett.113.141602.

[124] C. Jarlskog, Commutator of the quark mass matrices in the Standard Electroweak Model and a measure of maximal CP violation, *Phys. Rev. Lett.* **55** (1985) 1039, doi:10.1103/PhysRevLett.55.1039.

[125] M.B. Gavela *et al.*, Standard Model CP violation and baryon asymmetry. Part 2: Finite temperature, *Nucl. Phys.* **B430** (1994) 382–426, doi:10.1016/0550-3213(94)00410-2.

[126] G.C. Branco *et al.*, A Bridge between CP violation at low energies and leptogenesis, *Nucl. Phys.* **B617** (2001) 475–492, doi:10.1016/S0550-3213(01)00425-4.

[127] E.E. Jenkins and A.V. Manohar, Algebraic structure of lepton and quark flavor invariants and CP violation, *JHEP* **10** (2009) 094, doi:10.1088/1126-6708/2009/10/094.

[128] M. Fukugita and T. Yanagida, Baryogenesis without grand unification, *Phys. Lett.* **B174** (1986) 45, doi:10.1016/0370-2693(86)91126-3.

[129] K. Kajantie *et al.*, Is there a hot electroweak phase transition at $m_H \gtrsim m_W$?, *Phys. Rev. Lett.* **77** (1996) 2887–2890, doi:10.1103/PhysRevLett.77.2887.

[130] E.K. Akhmedov, V. Rubakov, and A.Y. Smirnov, Baryogenesis via neutrino oscillations, *Phys. Rev. Lett.* **81** (1998) 1359–1362, doi:10.1103/PhysRevLett.81.1359.

[131] L. Canetti, M. Drewes, T. Frossard, and M. Shaposhnikov. Dark matter, baryogenesis and neutrino oscillations from right handed neutrinos. *Phys. Rev.*, D87:093006, 2013, 1208.4607.

[132] P. Hernandez *et al.*, Bounds on right-handed neutrino parameters from observable leptogenesis. *JHEP*, 12 (2022) 012, doi:10.1007/JHEP11(2023)153.

[133] S. Sandner *et al.*, Predicting the baryon asymmetry with degenerate right-handed neutrinos, *JHEP* **11** (2023) 153, 10.1007/JHEP11(2023)153

Flavour physics

Yossi Nir
Weizmann, Institute of Science, Rehovot 7610001, Israel.
E-mail: yosef.nir@weizmann.ac.il

We explain the reasons for the interest in flavor physics. We describe flavor physics and the related CP violation within the Standard Model, with emphasis on the predictions of the model related to features such as flavor universality and flavor diagonality. We describe the flavor structure of flavor changing charged current interactions, and how they are used to extract the CKM parameters. We describe the structure of flavor changing neutral current interactions, and explain why they are highly suppressed in the Standard Model. We explain how the B-factories proved that the CKM (KM) mechanism dominates the flavor changing (CP violating) processes that have been observed in meson decays. We explain the implications of flavor physics for new physics, with emphasis on the "new physics flavor puzzle", and present the idea of minimal flavor violation as a possible solution. We explain the "Standard Model flavor puzzle", and present the Froggatt–Nielsen mechanism as a possible solution. We show that measurements of the Higgs boson decays may provide new opportunities for making progress on the various flavor puzzles. We briefly discuss two sets of measurements and some of their possible theoretical implications: $R(K^{(*)})$ and $R(D^{(*)})$.

Based in part on the book Y. Grossman and Y. Nir, *The Standard Model: From fundamental symmetries to experimental tests*, (Princeton University Press, Princeton, NJ, 2023).
This article should be cited as: Flavour physics, Yossi Nir, DOI: 10.23730/CYRSP-2025-001.209, 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. 209.

1 Introduction

1.1 What is flavor?

The term "**flavors**" is used, in the jargon of particle physics, to describe several mass eigenstates of the same gauge representation, namely several fields that are assigned the same quantum charges under the unbroken symmetries. Within the Standard Model (SM), when thinking of its unbroken $SU(3)_{\mathrm{C}} \times U(1)_{\mathrm{EM}}$ gauge group, there are four different types of fermions, each coming in three flavors:

– Up-type quarks in the $(3)_{+2/3}$ representation: u, c, t;
– Down-type quarks in the $(3)_{-1/3}$ representation: d, s, b;
– Charged leptons in the $(1)_{-1}$ representation: e, μ, τ;
– Neutrinos in the $(1)_0$ representation: ν_1, ν_2, ν_3.

The term "**flavor physics**" refers to interactions that distinguish between flavors. By definition, gauge interactions, namely interactions that are related to unbroken symmetries and mediated therefore by massless gauge bosons, do not distinguish among the flavors and do not constitute part of flavor physics. Within the Standard Model, flavor-physics refers to the weak and Yukawa interactions.

The term "**flavor parameters**" refers to parameters that carry flavor indices. Within the Standard Model, these are the nine masses of the charged fermions and the four "mixing parameters" (three angles and one phase) that describe the interactions of the charged weak-force carriers ($W^\pm$) with quark–anti-quark pairs. If one augments the Standard Model with Majorana mass terms for the neutrinos, one should add to the list three neutrino masses and six mixing parameters (three angles and three phases) for the $W^\pm$ interactions with lepton–anti-lepton pairs.

The term "**flavor universal**" refers to interactions with couplings (or to parameters) that are proportional to the unit matrix in flavor space. Thus, the strong and electromagnetic interactions are flavor-universal. An alternative term for "flavor-universal" is "**flavor-blind**".

The term "**flavor diagonal**" refers to interactions with couplings (or to parameters) that are diagonal, but not necessarily universal, in the flavor space. Within the Standard Model, the Yukawa interactions of the Higgs boson are flavor diagonal.

The term "**flavor changing**" refers to processes where the initial and final flavor-numbers (that is, the number of particles of a certain flavor minus the number of anti-particles of the same flavor) are different. In "**flavor changing charged current**" (FCCC) processes, both up-type and down-type flavors, and/or both charged lepton and neutrino flavors are involved. Examples are (i) $\mu \to e\bar{\nu}_e\nu_\mu$, (ii) $K^- \to \mu^-\bar{\nu}_\mu$ (which corresponds, at the quark level, to $s\bar{u} \to \mu^-\bar{\nu}_\mu$), and (iii) $B \to \psi K$ ($b \to c\bar{c}s$). Within the Standard Model, these processes are mediated by the W-bosons and occur at tree level. In "**flavor changing neutral current**" (FCNC) processes, either up-type or down-type flavors but not both, and/or either charged lepton or neutrino flavors but not both, are involved. Example are (i) $\mu \to e\gamma$, (ii) $K_L \to \mu^+\mu^-$ (which corresponds, at the quark level, to $s\bar{d} \to \mu^+\mu^-$), and (iii) $B \to \phi K$ ($b \to s\bar{s}s$). Within the Standard Model, these processes do not occur at tree level, and are often highly suppressed.

Another useful term is "**flavor violation**". We will explain it later in these lectures.

1.2 Why is flavor physics interesting?

Flavor physics is interesting, on one hand, as a tool for discovery and, on the other hand, because of intrinsic puzzling features:

- Flavor physics can discover new physics or probe it before it is directly observed in experiments. Here are some examples from the past:
  - The smallness of $\frac{\Gamma(K_L\to\mu^+\mu^-)}{\Gamma(K^+\to\mu^+\nu)}$ led to predicting a fourth (the charm) quark;
  - The size of Δm_K led to a successful prediction of the charm mass;
  - The size of Δm_B led to a successful prediction of the top mass;
  - The measurement of ε_K led to predicting the third generation;
  - The measurement of neutrino flavor transitions led to the discovery of neutrino masses.
- CP violation is closely related to flavor physics. Within the Standard Model, there is a single CP violating parameter, the Kobayashi–Maskawa phase $\delta_{\rm KM}$. Baryogenesis tells us, however, that there must exist new sources of CP violation. Measurements of CP violation in flavor changing processes might provide evidence for such sources.
- The fine-tuning problem of the Higgs mass, and the puzzle of the dark matter suggest that there may exist new physics at, or below, the TeV scale. If such new physics had a generic flavor structure, it would contribute to flavor changing neutral current (FCNC) processes orders of magnitude above the observed rates. The question of why this does not happen constitutes the *new physics flavor puzzle*.
- Most of the charged fermion flavor parameters are small and hierarchical. The Standard Model does not provide any explanation of these features. This is the *Standard Model flavor puzzle*. The puzzle became even deeper after neutrino masses and lepton mixing were measured because, so far, neither smallness nor hierarchy in these parameters have been established.

2 The Standard Model

A model of elementary particles and their interactions is defined by the following ingredients: (i) The symmetries of the Lagrangian and the pattern of spontaneous symmetry breaking (SSB); (ii) The representations of fermions and scalars. The Standard Model (SM) is defined as follows:

- The symmetry is a local
$$G_{\rm SM}=SU(3)_C\times SU(2)_L\times U(1)_Y. \tag{2.1}$$
- It is spontaneously broken,
$$G_{\rm SM}\to SU(3)_C\times U(1)_{\rm EM}\quad (Q_{\rm EM}=T_3+Y), \tag{2.2}$$
by the VEV of a single scalar field,
$$\phi(1,2)_{+1/2},\quad (\langle\phi^0\rangle=v/\sqrt{2}). \tag{2.3}$$

– There are three fermion generations, each consisting of five representations of $G_{\rm SM}$:

$$Q_{Li}(3,2)_{+1/6},\;\; U_{Ri}(3,1)_{+2/3},\;\; D_{Ri}(3,1)_{-1/3},\;\; L_{Li}(1,2)_{-1/2},\;\; E_{Ri}(1,1)_{-1}. \tag{2.4}$$

The SM scalar field is called the Higgs field. The $SU(3)_C$-triplet fermions fields are called quark fields, and the $SU(3)_C$-singlet fermions fields are called lepton fields.

2.1 The Lagrangian

The most general renormalizable Lagrangian with scalar and fermion fields can be decomposed into

$$\mathcal{L} = \mathcal{L}_{\rm kin} + \mathcal{L}_\psi + \mathcal{L}_{\rm Yuk} + \mathcal{L}_\phi. \tag{2.5}$$

Here $\mathcal{L}_{\rm kin}$ describes free propagation in spacetime, as well as gauge interactions, $\mathcal{L}_\psi$ gives fermion mass terms, $\mathcal{L}_{\rm Yuk}$ describes the Yukawa interactions, and $\mathcal{L}_\phi$ gives the scalar potential. We now find the specific form of the Lagrangian made of the fermion fields Q_{Li}, U_{Ri}, D_{Ri}, L_{Li} and E_{Ri} (2.4), and the scalar field (2.3), subject to the gauge symmetry (2.1) and leading to the SSB of Eq. (2.2).

2.1.1 $\mathcal{L}_{\rm kin}$

The local symmetry requires the following gauge boson degrees of freedom:

$$G_a^\mu(8,1)_0,\quad W_a^\mu(1,3)_0,\quad B^\mu(1,1)_0. \tag{2.6}$$

The corresponding field strengths are given by

$$\begin{aligned} G_a^{\mu\nu} &= \partial^\mu G_a^\nu - \partial^\nu G_a^\mu - g_s f_{abc} G_b^\mu G_c^\nu, \\ W_a^{\mu\nu} &= \partial^\mu W_a^\nu - \partial^\nu W_a^\mu - g\epsilon_{abc} W_b^\mu W_c^\nu, \\ B^{\mu\nu} &= \partial^\mu B^\nu - \partial^\nu B^\mu. \end{aligned} \tag{2.7}$$

The covariant derivative is

$$D^\mu = \partial^\mu + ig_s G_a^\mu L_a + igW_b^\mu T_b + ig' B^\mu Y, \tag{2.8}$$

where the L_a's are $SU(3)_C$ generators (the 3×3 Gell-Mann matrices $\frac{1}{2}\lambda_a$ for triplets, 0 for singlets), the T_b's are $SU(2)_L$ generators (the 2×2 Pauli matrices $\frac{1}{2}\tau_b$ for doublets, 0 for singlets), and the Y's are the $U(1)_Y$ charges. Explicitly, the covariant derivatives acting on the various scalar and fermion fields are given by

$$\begin{aligned} D^\mu\phi &= \left(\partial^\mu + \frac{i}{2}gW_b^\mu\tau_b + \frac{i}{2}g'B^\mu\right)\phi, \\ D^\mu Q_{Li} &= \left(\partial^\mu + \frac{i}{2}g_s G_a^\mu\lambda_a + \frac{i}{2}gW_b^\mu\tau_b + \frac{i}{6}g'B^\mu\right)Q_{Li}, \\ D^\mu U_{Ri} &= \left(\partial^\mu + \frac{i}{2}g_s G_a^\mu\lambda_a + \frac{2i}{3}g'B^\mu\right)U_{Ri}, \end{aligned}$$

$$
\begin{aligned}
D^\mu D_{Ri} &= \left(\partial^\mu + \frac{i}{2}g_s G_a^\mu \lambda_a - \frac{i}{3}g' B^\mu\right) D_{Ri}, \\
D^\mu L_{Li} &= \left(\partial^\mu + \frac{i}{2}g W_b^\mu \tau_b - \frac{i}{2}g' B^\mu\right) L_{Li}, \\
D^\mu E_{Ri} &= \left(\partial^\mu - ig' B^\mu\right) E_{Ri}.
\end{aligned} \tag{2.9}
$$

$\mathcal{L}_{\rm kin}$ is given by

$$
\begin{aligned}
\mathcal{L}_{\rm kin}^{\rm SM} &= -\frac{1}{4}G_a^{\mu\nu}G_{a\mu\nu} - \frac{1}{4}W_b^{\mu\nu}W_{b\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \\
&\quad + i\overline{Q_{Li}}\not{D}Q_{Li} + i\overline{U_{Ri}}\not{D}U_{Ri} + i\overline{D_{Ri}}\not{D}D_{Ri} + i\overline{L_{Li}}\not{D}L_{Li} + i\overline{E_{Ri}}\not{D}E_{Ri} \\
&\quad + (D^\mu\phi)^\dagger(D_\mu\phi).
\end{aligned} \tag{2.10}
$$

This part of the interaction Lagrangian is generation-universal. In addition, it conserves CP.

2.1.2 $\mathcal{L}_\psi$

There are no mass terms for the fermions in the SM. We cannot write Dirac mass terms for the fermions because they are assigned to chiral representations of the gauge symmetry. We cannot write Majorana mass terms for the fermions because they all have $Y \neq 0$. Thus,

$$\mathcal{L}_\psi^{\rm SM} = 0. \tag{2.11}$$

2.1.3 $\mathcal{L}_{\rm Yuk}$

The Yukawa part of the Lagrangian is given by

$$-\mathcal{L}_{\rm Y}^{\rm SM} = Y_{ij}^d \overline{Q_{Li}}\phi D_{Rj} + Y_{ij}^u \overline{Q_{Li}}\tilde{\phi} U_{Rj} + Y_{ij}^e \overline{L_{Li}}\phi E_{Rj} + {\rm h.c.}, \tag{2.12}$$

where $\tilde{\phi}_a = \epsilon_{ab}\phi_b^*$ (a, b are the $SU(2)$-indices). The Y^f's are general complex 3×3 matrices of dimensionless couplings. This part of the Lagrangian is, in general, generation-dependent (that is, $Y^f \not\propto \mathbf{1}$) and CP violating.

We now present three special interaction bases. Without loss of generality, we can use a bi-unitary transformation,

$$Y^e \to \hat{Y}_e = U_{eL} Y^e U_{eR}^\dagger, \tag{2.13}$$

to change the basis to one where Y^e is diagonal and real:

$$\hat{Y}^e = {\rm diag}(y_e, y_\mu, y_\tau). \tag{2.14}$$

In the basis defined in Eq. (2.14), we denote the components of the lepton $SU(2)$-doublets, and the three lepton $SU(2)$-singlets, as follows:

$$\begin{pmatrix}\nu_{eL}\\ e_L\end{pmatrix}, \qquad \begin{pmatrix}\nu_{\mu L}\\ \mu_L\end{pmatrix}, \qquad \begin{pmatrix}\nu_{\tau L}\\ \tau_L\end{pmatrix}; \qquad e_R, \qquad \mu_R, \qquad \tau_R, \tag{2.15}$$

where e, μ, τ are ordered by the size of $y_{e,\mu,\tau}$ (from smallest to largest).

Similarly, without loss of generality, we can use a bi-unitary transformation,

$$Y^u \to \hat{Y}^u = V_{uL} Y^u V_{uR}^\dagger, \tag{2.16}$$

to change the basis to one where $\hat{Y}^u$ is diagonal and real:

$$\hat{Y}^u = \mathrm{diag}(y_u, y_c, y_t). \tag{2.17}$$

In the basis defined in Eq. (2.17), we denote the components of the quark $SU(2)$-doublets, and the quark up $SU(2)$-singlets, as follows:

$$\begin{pmatrix} u_L \\ d_{uL} \end{pmatrix}, \quad \begin{pmatrix} c_L \\ d_{cL} \end{pmatrix}, \quad \begin{pmatrix} t_L \\ d_{tL} \end{pmatrix}; \quad u_R, \quad c_R, \quad t_R, \tag{2.18}$$

where u, c, t are ordered by the size of $y_{u,c,t}$ (from smallest to largest).

We can use yet another bi-unitary transformation,

$$Y^d \to \hat{Y}^d = V_{dL} Y^d V_{dR}^\dagger, \tag{2.19}$$

to change the basis to one where $\hat{Y}^d$ is diagonal and real:

$$\hat{Y}^d = \mathrm{diag}(y_d, y_s, y_b). \tag{2.20}$$

In the basis defined in Eq. (2.20), we denote the components of the quark $SU(2)$-doublets, and the quark down $SU(2)$-singlets, as follows:

$$\begin{pmatrix} u_{dL} \\ d_L \end{pmatrix}, \quad \begin{pmatrix} u_{sL} \\ s_L \end{pmatrix}, \quad \begin{pmatrix} u_{bL} \\ b_L \end{pmatrix}; \quad d_R, \quad s_R, \quad b_R, \tag{2.21}$$

where d, s, b are ordered by the size of $y_{d,s,b}$ (from smallest to largest).

2.1.4 $\mathcal{L}_\phi$

The scalar potential is given by

$$\mathcal{L}_\phi^{\rm SM} = -\mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2. \tag{2.22}$$

Choosing $\mu^2 < 0$ and $\lambda > 0$ leads to the required spontaneous symmetry breaking. This part of the Lagrangian is also CP conserving.

2.2 The spectrum

The fermion masses arise from the Yukawa couplings as a result of the spontaneous symmetry breaking. The mass matrices are given by

$$M_f = (v/\sqrt{2}) Y^f \quad (f = e, u, d). \tag{2.23}$$

Table 1: The SM particles.

particle	spin	color	$Q_{\rm EM}$	mass $[v]$
$W^\pm$	1	(1)	± 1	$\frac{1}{2}g$
Z^0	1	(1)	0	$\frac{1}{2}\sqrt{g^2+g'^2}$
A^0	1	(1)	0	0
g	1	(8)	0	0
h	0	(1)	0	$\sqrt{2\lambda}$
e,μ,τ	1/2	(1)	-1	$y_{e,\mu,\tau}/\sqrt{2}$
ν_e,ν_μ,ν_τ	1/2	(1)	0	0
u,c,t	1/2	(3)	$+2/3$	$y_{u,c,t}/\sqrt{2}$
d,s,b	1/2	(3)	$-1/3$	$y_{d,s,b}/\sqrt{2}$

It is clear then that the bases of diagonal Yukawa matrices—the $\hat{Y}^e$ basis of Eq. (2.14), the $\hat{Y}^u$ basis of Eq. (2.17), and the $\hat{Y}^d$ basis of Eq. (2.20)—are mass bases for, respectively, the charged leptons, the up quarks and the down quarks, with $m_f = (v/\sqrt{2})y_f$. The spectrum of the Standard Model is presented in Table 1.

All masses are proportional to the VEV of the scalar field, v. For the three massive gauge bosons, and for the fermions, this is expected: In the absence of spontaneous symmetry breaking, the former would be protected by the gauge symmetry and the latter by their chiral nature. For the Higgs boson, the situation is different, as a mass-squared term does not violate any symmetry.

For the charged fermions, the spontaneous symmetry breaking allows their masses because they are in vector-like representations of the $SU(3)_C\times U(1)_{\rm EM}$ group: The LH and RH charged lepton fields, e, μ and τ, are in the $(1)_{-1}$ representation; The LH and RH up-type quark fields, u, c and t, are in the $(3)_{+2/3}$ representation; The LH and RH down-type quark fields, d, s and b, are in the $(3)_{-1/3}$ representation. On the other hand, the neutrinos remain massless in spite of the fact that they are in the $(1)_0$ representation of $SU(3)_C\times U(1)_{\rm EM}$, which allows for Majorana masses. Such masses require a VEV carried by a scalar field in the $(1,3)_{+1}$ representation of the $SU(3)_C\times SU(2)_L\times U(1)_Y$ symmetry, but there is no such field in the SM.

The experimental values of the charged fermion masses are [1][1]

$$\begin{aligned}
&m_e = 0.510998946(3)\ \text{MeV},\ \ m_\mu = 105.6583745(24)\ \text{MeV},\ \ m_\tau = 1776.86(12)\ \text{MeV},\\
&m_u = 2.2^{+0.5}_{-0.3}\ \text{MeV},\ \ m_c = 1.27\pm 0.02\ \text{GeV},\ \ m_t = 172.9\pm 0.4\ \text{GeV},\\
&m_d = 4.7^{+0.5}_{-0.2}\ \text{MeV},\ \ m_s = 93^{+11}_{-5}\ \text{MeV},\ \ m_b = 4.18^{+0.03}_{-0.02}\ \text{GeV}.
\end{aligned} \tag{2.24}$$

2.2.1 The CKM matrix

In the derivation above, there is an important difference between the analysis of the quark spectrum and the analysis of the lepton spectrum. For the leptons, there exists a basis that is simultaneously an

[1]See Ref. [1] for detailed explanations of the quoted quark masses. For $q = u,d,s,c,b$, m_q are the running quark masses in the $\overline{\rm MS}$ scheme, with $m_{u,d,s} = m_{u,d,s}(\mu = 2\ \text{GeV})$ and $m_{c,b} = m_{c,b}(\mu = m_{c,b})$.

Table 2: The SM fermion interactions.

interaction	fermions	force carrier	coupling	flavor
Electromagnetic	u, d, ℓ	A^0	eQ	universal
Strong	u, d	g	g_s	universal
NC weak	u, d, e, ν	Z^0	$\frac{e(T_3 - s_W^2 Q)}{s_W c_W}$	universal
CC weak (q)	$\bar{u}d$	$W^\pm$	gV	non-universal
CC weak (l)	$\bar{\ell}\nu$	$W^\pm$	g	universal
Yukawa	u, d, ℓ	h	y_q	diagonal

interaction basis and a mass basis for both the charged leptons and the neutrinos, that is the $\hat{Y}_e$ basis. In contrast, for the quarks, in general there is no interaction basis that is also a mass basis for both up-type and down-type quarks. To see that, we denote $u^i = (u, c, t)$ and $d^i = (d, s, b)$, and write the relation of these mass eigenstates to the interaction eigenstates:

$$u_L^i = (V_{uL})_{ij} U_L^j, \qquad u_R^i = (V_{uR})_{ij} U_R^j, \qquad d_L^i = (V_{dL})_{ij} D_L^j, \qquad d_R^i = (V_{dR})_{ij} D_R^j. \tag{2.25}$$

If $V_{uL} \neq V_{dL}$, as is the general case, then the interaction basis defined by Eq. (2.17) is different from the interaction basis defined by Eq. (2.20). In the former, Y^d can be written as a unitary matrix times a diagonal one,

$$Y^u = \hat{Y}^u, \qquad Y^d = V\hat{Y}^d. \tag{2.26}$$

In the latter, Y^u can be written as a unitary matrix times a diagonal one,

$$Y^d = \hat{Y}^d, \qquad Y^u = V^\dagger \hat{Y}^u. \tag{2.27}$$

In either case, the unitary matrix V is given by

$$V = V_{uL} V_{dL}^\dagger, \tag{2.28}$$

where V_{uL} and V_{dL} are defined in Eqs. (2.16) and (2.19), respectively. Note that each of V_{uL}, V_{uR}, V_{dL} and V_{dR} depends on the basis from which we start the diagonalization. The combination $V = V_{uL}V_{dL}^\dagger$, however, does not. This is a hint that V is physical. The matrix V is called the Cabibbo–Kobayashi–Maskawa (CKM) matrix [2, 3]. Its physical significance becomes clear in Section 2.3.3.

2.3 The interactions

Within the SM, the fermions have five types of interactions. These interactions are summarized in Table 2. In the next few subsections, we explain the entries of this table.

2.3.1 EM and strong interactions

By construction, a local $SU(3)_C \times U(1)_{\rm EM}$ symmetry survives the SSB. The SM has thus the photon and gluon massless gauge fields. All charged fermions interact with the photon:

$$\mathcal{L}_{{\rm QED},\psi} = -\frac{2e}{3}\overline{u_i}\not{A}u_i + \frac{e}{3}\overline{d_i}\not{A}d_i + e\overline{\ell_i}\not{A}\ell_i, \tag{2.29}$$

where $u_{1,2,3} = u, c, t$, $d_{1,2,3} = d, s, b$ and $\ell_{1,2,3} = e, \mu, \tau$. We emphasize the following points:

1. The photon couplings are *vector-like*.
2. The EM interactions are *P, C and T conserving*.
3. *Diagonality:* The photon couples to e^+e^-, $\mu^+\mu^-$ and $\tau^+\tau^-$, but not to $e^\pm\mu^\mp$, $e^\pm\tau^\mp$ or $\mu^\pm\tau^\mp$ pairs, and similarly in the up and down sectors.
4. *Universality:* The couplings of the photon to different generations are universal.

All colored fermions (namely, quarks) interact with the gluon:

$$\mathcal{L}_{{\rm QCD},\psi} = -\frac{g_s}{2}\overline{q}\lambda_a\not{G}_a q, \tag{2.30}$$

where $q = u, c, t, d, s, b$. We emphasize the following points:

1. The gluon couplings are *vector-like*.
2. The strong interactions are *P, C and T conserving*.
3. *Diagonality:* The gluon couples to $\bar{t}t$, $\bar{c}c$, *etc.*, but not to $\bar{t}c$ or any other flavor changing pair.
4. *Universality:* The couplings of the gluon to different quark generations are universal.

The universality of the photon and gluon couplings is a result of the $SU(3)_C \times U(1)_{\rm EM}$ gauge invariance, and thus holds in any model, and not just within the SM.

2.3.2 Neutral current weak interactions

All SM fermions couple to the Z-boson:

$$\begin{aligned}\mathcal{L}_{Z,\psi} \;=\; & \frac{e}{s_W c_W}\left[-\left(\frac{1}{2}-s_W^2\right)\overline{e_{Li}}\not{Z}e_{Li} + s_W^2\,\overline{e_{Ri}}\not{Z}e_{Ri} + \frac{1}{2}\,\overline{\nu_{L\alpha}}\not{Z}\nu_{L\alpha}\right. \\ & \left. + \left(\frac{1}{2}-\frac{2}{3}s_W^2\right)\overline{u_{Li}}\not{Z}u_{Li} - \frac{2}{3}s_W^2\,\overline{u_{Ri}}\not{Z}u_{Ri} - \left(\frac{1}{2}-\frac{1}{3}s_W^2\right)\overline{d_{Li}}\not{Z}d_{Li} + \frac{1}{3}s_W^2\,\overline{d_{Ri}}\not{Z}d_{Ri}\right].\end{aligned} \tag{2.31}$$

where $\nu_\alpha = \nu_e, \nu_\mu, \nu_\tau$. We emphasize the following points:

1. The Z-boson couplings are *chiral* and *parity violating*.
2. *Diagonality:* The Z-boson couples diagonally. For example, in the lepton sector, the Z-boson couples to e^+e^- and to $\mu^+\mu^-$ but not to $e^\pm\mu^\mp$ pairs. The diagonality in the lepton sector holds to all orders in perturbation theory, due to an accidental $[U(1)]^3$ symmetry of the SM (see below).
3. *Universality:* The couplings of the Z-boson in each of the seven sectors $(\nu_L, \ell_L, \ell_R, d_L, d_R, u_L, u_R)$ are universal. This is a result of a special feature of the SM:

all fermions of given chirality, EM charge and $SU(3)_C$ representation come from the same $SU(2)_L \times U(1)_Y$ representation (see below).

As an example to experimental tests of diagonality and universality, we can take the leptonic sector. The branching ratios of the Z-boson into charged lepton pairs [1],

$$\begin{aligned} \mathrm{BR}(Z \to e^+e^-) &= (3.363 \pm 0.004)\%\,, \\ \mathrm{BR}(Z \to \mu^+\mu^-) &= (3.366 \pm 0.007)\%\,, \\ \mathrm{BR}(Z \to \tau^+\tau^-) &= (3.370 \pm 0.008)\%\,, \end{aligned} \tag{2.32}$$

beautifully confirms universality:

$$\begin{aligned} \Gamma(\mu^+\mu^-)/\Gamma(e^+e^-) &= 1.0001 \pm 0.0024, \\ \Gamma(\tau^+\tau^-)/\Gamma(e^+e^-) &= 1.002 \pm 0.003. \end{aligned}$$

Diagonality is also tested by the following experimental searches:

$$\begin{aligned} \mathrm{BR}(Z \to e^+\mu^-) &< 7.5 \times 10^{-7}, \\ \mathrm{BR}(Z \to e^+\tau^-) &< 5.0 \times 10^{-6}, \\ \mathrm{BR}(Z \to \mu^+\tau^-) &< 6.5 \times 10^{-6}. \end{aligned} \tag{2.33}$$

Thus, for example,

$$\begin{aligned} \Gamma(e^+\mu^-)/\Gamma(\ell^+\ell^-) &< 2.2 \times 10^{-5}, \\ \Gamma(\mu^+\tau^-)/\Gamma(\tau^+\tau^-) &< 1.9 \times 10^{-4}. \end{aligned} \tag{2.34}$$

2.3.3 Charged current weak interactions

We now study the couplings of the charged vector bosons, $W^\pm$, to fermion pairs. For the lepton mass eigenstates, things are simple, because there exists an interaction basis that is also a mass basis. Thus,

$$\mathcal{L}_{W,\ell} = -\frac{g}{\sqrt{2}} \left(\overline{\nu_{eL}}\, \not{W}^+ e_L^- + \overline{\nu_{\mu L}}\, \not{W}^+ \mu_L^- + \overline{\nu_{\tau L}}\, \not{W}^+ \tau_L^- + \mathrm{h.c.} \right). \tag{2.35}$$

Eq. (2.35) reveals some important features of the model:

1. *Parity violation*: The W-boson couplings are chiral. More specifically, only left-handed particles take part in charged-current interactions. Consequently, parity is violated.
2. *Universality*: the couplings of the W-boson to $\tau\bar{\nu}_\tau$, to $\mu\bar{\nu}_\mu$ and to $e\bar{\nu}_e$ are equal. This is a result of the local nature of the imposed $SU(2)$: a global symmetry would have allowed an independent coupling to each lepton pair.

All of these predictions have been experimentally tested. As an example of how well universality works, consider the decay rates of the W-bosons to the three lepton pairs [1]:

$$\mathrm{BR}(W^+ \to e^+\nu_e) = (10.71 \pm 0.16) \times 10^{-2},$$

$$\begin{aligned} \mathrm{BR}(W^+ \to \mu^+\nu_\mu) &= (10.63 \pm 0.15) \times 10^{-2}, \\ \mathrm{BR}(W^+ \to \tau^+\nu_\tau) &= (11.38 \pm 0.21) \times 10^{-2}. \end{aligned} \tag{2.36}$$

These results confirm universality:

$$\begin{aligned} \Gamma(\mu^+\nu)/\Gamma(e^+\nu) &= 0.996 \pm 0.008, \\ \Gamma(\tau^+\nu)/\Gamma(\mu^+\nu) &= 1.043 \pm 0.024. \end{aligned} \tag{2.37}$$

As concerns quarks, things are more complicated, since there is no interaction basis that is also a mass basis. In the interaction basis, the W interactions have the following form:

$$\mathcal{L}_{W,q} = -\frac{g}{\sqrt{2}}\overline{U_L^i}\, \not{W}^+ D_L^i + \mathrm{h.c.}. \tag{2.38}$$

Using Eq. (2.25) to write $U_L^i = (V_{uL}^\dagger)_{ij} u_L^j$ and $D_L^i = (V_{dL}^\dagger)_{ij} d_L^j$, we can rewrite $\mathcal{L}_{W,q}$ in terms of the mass eigenstates:

$$\mathcal{L}_{W,q} = -\frac{g}{\sqrt{2}}\, \overline{u_L^k} (V_{uL})_{ki}\, \not{W}^+ (V_{dL}^\dagger)_{il} d_L^l + \mathrm{h.c.} = -\frac{g}{\sqrt{2}}\, \overline{u_L^k} V_{kl}\, \not{W}^+ d_L^l + \mathrm{h.c.}, \tag{2.39}$$

where V is the CKM matrix defined in Eq. (2.28).

Eq. (2.39) reveals some important features of the model:

1. Only left-handed particles take part in charged-current interactions. Consequently, parity is violated by these interactions.
2. The W couplings to the quark mass eigenstates are not universal. The universality of gauge interactions is hidden in the unitarity of the CKM matrix, V.
3. The W couplings are not diagonal. This is a manifestation of the fact that no pair of an up-type and a down-type mass eigenstates fits into an $SU(2)_L$ doublet. For example, the d and u mass eigenstates are not members of a single $SU(2)_L$ doublet.

The matrix V is called the CKM matrix [2,3]. The (hidden) universality within the quark sector is tested by the prediction

$$\Gamma(W \to uX) = \Gamma(W \to cX) = \frac{1}{2}\Gamma(W \to \mathrm{hadrons}). \tag{2.40}$$

Experimentally,

$$\Gamma(W \to cX)/\Gamma(W \to \mathrm{hadrons}) = 0.49 \pm 0.04. \tag{2.41}$$

2.3.4 Yukawa interactions

The Yukawa interactions are given by

$$\begin{aligned} \mathcal{L}_{\mathrm{Yuk}} = \; - \; & \frac{h}{v}\big(m_e\, \overline{e_L}\, e_R + m_\mu\, \overline{\mu_L}\, \mu_R + m_\tau\, \overline{\tau_L}\, \tau_R \\ & + m_u\, \overline{u_L}\, u_R + m_c\, \overline{c_L}\, c_R + m_t\, \overline{t_L}\, t_R + m_d\, \overline{d_L}\, d_R + m_s\, \overline{s_L}\, s_R + m_b\, \overline{b_L}\, b_R + \mathrm{h.c.}\big)\,. \end{aligned}$$

To see that the Higgs boson couples diagonally to the fermion mass eigenstates, let us take the example of the down quarks, and start from an arbitrary interaction basis:

$$\begin{aligned} h\overline{D_L}Y^d D_R &= h\overline{D_L}(V_{dL}^\dagger V_{dL})Y^d(V_{dR}^\dagger V_{dR})D_R \\ &= h(\overline{D_L}V_{dL}^\dagger)(V_{dL}Y^d V_{dR}^\dagger)(V_{dR}D_R) \\ &= h(\overline{d_L}\,\overline{s_L}\,\overline{b_L})\hat{Y}^d(d_R\,s_R\,b_R)^T. \end{aligned} \tag{2.42}$$

We conclude that the Higgs couplings to the fermion mass eigenstates have the following features:

1. *Diagonality.*
2. *Non-universality.*
3. *Proportionality* to the fermion masses: the heavier the fermion, the stronger the coupling. The factor of proportionality is m_f/v.
4. *CP conservation.*

Thus, the Higgs boson decay is dominated by the heaviest particle which can be pair-produced in the decay. For $m_h \sim 125$ GeV, this is the bottom quark. Indeed, the SM predicts the following branching ratios for the leading decay modes:

$$\mathrm{BR}_{\bar{b}b} : \mathrm{BR}_{WW^*} : \mathrm{BR}_{gg} : \mathrm{BR}_{\tau^+\tau^-} : \mathrm{BR}_{ZZ^*} : \mathrm{BR}_{c\bar{c}} = 0.58 : 0.21 : 0.09 : 0.06 : 0.03 : 0.03. \tag{2.43}$$

The following comments are in order with regard to Eq. (2.43):

1. From the six branching ratios, three (b, τ, c) stand for two-body tree-level decays. Thus, at tree level, the respective branching ratios obey $\mathrm{BR}_{\bar{b}b} : \mathrm{BR}_{\tau^+\tau^-} : \mathrm{BR}_{c\bar{c}} = 3m_b^2 : m_\tau^2 : 3m_c^2$. QCD radiative corrections somewhat suppress the two modes with the quark final states (b, c) compared to one with the lepton final state (τ).
2. The WW^* and ZZ^* modes stand for the three-body tree-level decays ($W\overline{f}f'$ and $Z\overline{f}f$, respectively), where one of the vector bosons is on-shell and the other off-shell.
3. The Higgs boson does not have a tree-level coupling to gluons since it carries no color (and the gluons have no mass). The decay into final gluons proceeds via loop diagrams. The dominant contribution comes from the top-quark loop.
4. Similarly, the Higgs decays into final two photons via loop diagrams with small ($\mathrm{BR}_{\gamma\gamma} \sim 0.002$), but observable, rate. The dominant contributions come from the W and the top-quark loops which interfere destructively.

Experimentally, the decays into final ZZ^*, WW^*, $\gamma\gamma$, $\bar{b}b$ and $\tau^+\tau^-$ have been established with rates that are consistent with the SM predictions.

2.4 Global symmetries

In the absence of the Yukawa matrices, $\mathcal{L}_{\text{Yuk}} = 0$, the SM has a $[U(3)]^5$ global symmetry:

$$G_{\text{global}}^{\text{SM}}(Y^{u,d,e} = 0) = SU(3)_q^3 \times SU(3)_\ell^2 \times U(1)^5, \tag{2.44}$$

where

$$\begin{aligned} SU(3)^3_q &= SU(3)_Q \times SU(3)_U \times SU(3)_D, \\ SU(3)^2_\ell &= SU(3)_L \times SU(3)_E, \\ U(1)^5 &= U(1)_Q \times U(1)_U \times U(1)_D \times U(1)_L \times U(1)_E. \end{aligned} \tag{2.45}$$

The point that is important for our purposes is that $\mathcal{L}_{\rm kin}$ respects the non-Abelian flavor symmetry $SU(3)^3_q \times SU(3)^2_\ell$, under which

$$Q_L \to V_Q Q_L, \quad U_R \to V_U U_R, \quad D_R \to V_D D_R, \quad L_L \to V_L L_L, \quad E_R \to V_E E_R, \tag{2.46}$$

where the V_i are unitary matrices. The Yukawa interactions (2.12) break the global symmetry,

$$G^{\rm SM}_{\rm global}(Y^{u,d,e} \neq 0) = U(1)_B \times U(1)_e \times U(1)_\mu \times U(1)_\tau. \tag{2.47}$$

Thus, the transformations of Eq. (2.46) are not a symmetry of $\mathcal{L}_{\rm SM}$. Instead, they correspond to a change of the interaction basis. These observations also offer an alternative way of defining flavor physics: it refers to interactions that break the $[SU(3)]^5$ symmetry (2.46). Thus, the term "**flavor violation**" is often used to describe processes or parameters that break the symmetry.

One can think of the quark Yukawa couplings as spurions that break the global $SU(3)^3_q$ symmetry (but are neutral under $U(1)_B$),

$$Y^u \sim (3, \bar{3}, 1)_{SU(3)^3_q}, \quad Y^d \sim (3, 1, \bar{3})_{SU(3)^3_q}, \tag{2.48}$$

and of the lepton Yukawa couplings as spurions that break the global $SU(3)^2_\ell$ symmetry (but are neutral under $U(1)_e \times U(1)_\mu \times U(1)_\tau$),

$$Y^e \sim (3, \bar{3})_{SU(3)^2_\ell}. \tag{2.49}$$

The spurion formalism is convenient for several purposes: parameter counting (see below), identification of flavor suppression factors (see Section 7), and the idea of minimal flavor violation (see Section 7.3).

2.5 Counting parameters

How many independent parameters are there in $\mathcal{L}^q_{\rm Yuk}$? The two Yukawa matrices, Y^u and Y^d, are 3×3 and complex. Consequently, there are 18 real and 18 imaginary parameters in these matrices. Not all of them are, however, physical. The pattern of $G_{\rm global}$ breaking means that there is freedom to remove 9 real and 17 imaginary parameters (the number of parameters in three 3×3 unitary matrices minus the phase related to $U(1)_B$). For example, we can use the unitary transformations $Q_L \to V_Q Q_L$, $U_R \to V_U U_R$ and $D_R \to V_D D_R$, to lead to the following interaction basis:

$$Y^d = \hat{Y}^d, \quad Y^u = V^\dagger \hat{Y}^u, \tag{2.50}$$

where $\hat{Y}^{d,u}$ are diagonal,

$$\hat{Y}^d = \mathrm{diag}(y_d, y_s, y_b), \quad \hat{Y}^u = \mathrm{diag}(y_u, y_c, y_t), \tag{2.51}$$

while V is a unitary matrix that depends on three real angles and one complex phase. We conclude that there are 10 quark flavor parameters: 9 real ones and a single phase. In the mass basis, we identify the nine real parameters as six quark masses and three mixing angles, while the single phase is $\delta_{\rm KM}$.

How many independent parameters are there in $\mathcal{L}^{\ell}_{\rm Yuk}$? The Yukawa matrix Y^e is 3×3 and complex. Consequently, there are 9 real and 9 imaginary parameters in this matrix. There is, however, freedom to remove 6 real and 9 imaginary parameters (the number of parameters in two 3×3 unitary matrices minus the phases related to $U(1)^3$). For example, we can use the unitary transformations $L_L \to V_L L_L$ and $E_R \to V_E E_R$, to lead to the following interaction basis:

$$Y^e = \hat{Y}^e = \mathrm{diag}(y_e, y_\mu, y_\tau). \tag{2.52}$$

We conclude that there are 3 real lepton flavor parameters. In the mass basis, we identify these parameters as the three charged lepton masses. We must, however, modify the model when we take into account the evidence for neutrino masses.

3 Flavor changing charged current (FCCC) processes

3.1 The CKM matrix

Among the SM interactions, the W-mediated interactions are the only ones that are not diagonal in the mass basis. Consequently, all flavor changing processes depend on the CKM parameters. The fact that there are only four independent CKM parameters, while the number of measured flavor changing processes is much larger, allows for stringent tests of the CKM mechanism for flavor changing processes.

3.1.1 The standard parametrization

The CKM matrix is defined in Eq. (2.28). Its explicit form is not unique. First, there is freedom in defining V in that we can permute between the various generations. This freedom is fixed by ordering the up quarks and the down quarks by their masses, i.e. $(u_1, u_2, u_3) \to (u, c, t)$ and $(d_1, d_2, d_3) \to (d, s, b)$. We then write the W interaction of Eq. (2.39) as

$$\mathcal{L}_{W,q} = -\frac{g}{\sqrt{2}} \begin{pmatrix} \overline{u_L} & \overline{c_L} & \overline{t_L} \end{pmatrix} V \not{W}^+ \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} + \text{h.c.}. \tag{3.1}$$

The elements of V are therefore written as follows:

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{3.2}$$

Second, we can redefine the phases of the quark fields in such a way that the masses remain real but the phase structure of the CKM matrix changes. This freedom can be used to choose an explicit parametrization that depends on three real and one imaginary parameters. For example, the standard parametrization [4, 5], used by the PDG, is given by

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{3.3}$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$. The three θ_{ij} are the three mixing angles while δ is the Kobayashi–Maskawa phase. With the fixed mass ordering explained above, we have $\theta_{ij} \in \{0, \pi/2\}$ and $\delta \in \{0, 2\pi\}$. The mixing angles θ_{ij} are often referred to as the real parameters, and δ as the imaginary one, or the CP violating one.

The fitted values of the four parameters are given by

$$\begin{aligned} \sin\theta_{12} &= 0.2250 \pm 0.0007, \\ \sin\theta_{23} &= 0.0418 \pm 0.0008, \\ \sin\theta_{13} &= 0.0037 \pm 0.0001, \\ \delta &= 1.20 \pm 0.04. \end{aligned} \tag{3.4}$$

This translates into the following ranges for the magnitude of the CKM elements:

$$|V| = \begin{pmatrix} 0.97401 \pm 0.00011 & 0.22650 \pm 0.00048 & 0.00361 \pm 0.00010 \\ 0.22636 \pm 0.00048 & 0.97320 \pm 0.00011 & 0.04053^{+0.00083}_{-0.00061} \\ 0.00854^{+0.00023}_{-0.00016} & 0.03978^{+0.00082}_{-0.00060} & 0.999172^{+0.000024}_{-0.000035} \end{pmatrix}. \tag{3.5}$$

We discuss some of the ways in which these entries are determined below.

3.1.2 The Wolfenstein parametrization

Equation (3.5) implies that the CKM matrix is numerically close to a unit matrix, with small off-diagonal terms that obey the following hierarchy:

$$|V_{ub}|, |V_{td}| \ll |V_{cb}|, |V_{ts}| \ll |V_{us}|, |V_{cd}|. \tag{3.6}$$

This situation inspires an approximate parametrization, known as the Wolfenstein parametrization. The Wolfenstein parameters consist of the three real parameters λ, A and ρ, and the imaginary (CP violating) parameter $i\eta$. The expansion is in the small parameter,

$$\lambda = |V_{us}| \approx 0.23. \tag{3.7}$$

The order of magnitude of each element can be read from the power of λ. To $\mathcal{O}(\lambda^3)$, the CKM matrix is written in terms of the Wolfenstein parameters as follows [6, 7]:

$$V = \begin{pmatrix} 1-\frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix}. \tag{3.8}$$

The relations between the standard parameters and the Wolfenstein parameters are given by

$$\lambda = s_{12}, \qquad A\lambda^2 = s_{23}, \qquad A\lambda^3(\rho - i\eta) = s_{13}e^{-i\delta}. \tag{3.9}$$

The fitted values of the four parameters can be read from Eq. (3.4)

$$\begin{aligned} \rho &= 0.16 \pm 0.01, \\ \eta &= 0.35 \pm 0.01, \\ A &= 0.83 \pm 0.02, \\ \lambda &= 0.2250 \pm 0.0007. \end{aligned} \tag{3.10}$$

The experimental fact that the CKM matrix is close to a unit matrix is one of the ingredients of *the SM* that are far from a generic SM. The hierarchy in the quark masses constitutes another such ingredient.

3.1.3 CP violation

Various parameterizations differ in the way that the freedom of phase rotation is used to leave a single phase in V. One can define, however, a CP violating quantity in V that is independent of the parametrization. This quantity, the Jarlskog invariant [8, 9], $J_{\rm CKM}$, is defined through

$$\mathcal{I}m(V_{ij}V_{kl}V_{il}^*V_{kj}^*) = J_{\rm CKM}\sum_{m,n=1}^{3}\epsilon_{ikm}\epsilon_{jln}, \qquad (i,j,k,l = 1,2,3). \tag{3.11}$$

(There is no sum over the i, j, k, l indices.) In terms of the explicit parameterizations given in Eqs. (3.3) and (3.8), the Jarlskog invariant is given by

$$J_{\rm CKM} = c_{12}c_{23}c_{13}^2 s_{12}s_{23}s_{13}\sin\delta \approx \lambda^6 A^2 \eta. \tag{3.12}$$

Note that $|J_{\rm CKM}|$ is bounded from above,

$$|J_{\rm CKM}| \leq \frac{1}{6\sqrt{3}} \sim 0.1. \tag{3.13}$$

The current best fit for $J_{\rm CKM}$ is given by

$$J_{\rm CKM} = (3.00^{+0.15}_{-0.09}) \times 10^{-5}, \tag{3.14}$$

which is much smaller than the upper bound of Eq. (3.13). More significantly, the experimental value is much smaller than the value it would have if all relevant parameters were $O(1)$. This is one more demonstration that, within the flavor sector, the SM has non-generic features.

While a generic SM violates CP, specific realizations of it could still conserve CP. In order that *the SM* violates CP, the following necessary and sufficient condition must be fulfilled:

$$X_{CP} \equiv \Delta m^2_{tc}\Delta m^2_{tu}\Delta m^2_{cu}\Delta m^2_{bs}\Delta m^2_{bd}\Delta m^2_{sd} J_{\rm CKM} \neq 0, \tag{3.15}$$

where $\Delta m^2_{ij} \equiv m^2_i - m^2_j$. Equation (3.15) puts the following requirements on the SM in order that CP is violated:

1. Within each quark sector, there should be no mass degeneracy;
2. The Jarlskog invariant does not vanish.

These conditions can also be written as a single requirement on the quark mass matrices in any interaction basis [8, 9]:

$$X_{CP} = \mathcal{I}m\left\{\det\left[M_d M_d^\dagger, M_u M_u^\dagger\right]\right\} \neq 0 \;\Leftrightarrow\; CP \text{ violation}. \tag{3.16}$$

This is a convention independent condition.

3.1.4 SM2: CP conserving

Consider a two generation Standard Model, SM2. This model is similar to the one defined in Section 2, which in this section will be referred to as SM3, except that there are two, rather than three fermion generations. Many features of SM2 are similar to SM3, but there is one important difference: CP is a good symmetry of SM2, but not of SM3. To see how this difference comes about, let us examine the accidental symmetries of SM2. We follow here the line of analysis of SM3 in Section 2.5.

If we set the Yukawa couplings to zero, $\mathcal{L}^{\rm SM2}_{\rm Yuk} = 0$, SM2 gains an accidental global symmetry:

$$G^{\rm global}_{\rm SM2}(Y^{u,d,e} = 0) = U(2)_Q \times U(2)_U \times U(2)_D \times U(2)_L \times U(2)_E, \tag{3.17}$$

where the two generations of each gauge representation are a doublet of the corresponding $U(2)$. The Yukawa couplings break this symmetry into the subgroup

$$G^{\rm global}_{\rm SM2} = U(1)_B \times U(1)_e \times U(1)_\mu. \tag{3.18}$$

A-priori, the Yukawa terms depend on three 2×2 complex matrices, namely $12_R + 12_I$ parameters. The global symmetry breaking, $[U(2)]^5 \to [U(1)]^3$, implies that we can remove $5 \times (1_R + 3_I) - 3_I = 5_R + 12_I$ parameters. Thus the number of physical flavor parameters is 7 real parameters and no imaginary parameter. The real parameters can be identified as two charged lepton masses, four quark masses, and the single real mixing angle, $\sin\theta_c = |V_{us}|$.

The important conclusion for our purposes is that all imaginary couplings can be removed from SM2, and CP is an accidental symmetry of the model.

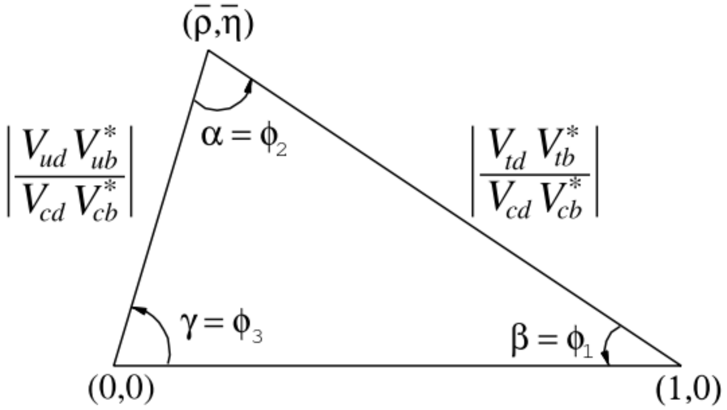


Fig. 1: The rescaled unitarity triangle.

3.1.5 Unitarity triangles

A very useful concept with regard to CP violation is that of the unitarity triangles. The unitarity of the CKM matrix leads to various relations among its elements. Of particular interest are the six relations:

$$\begin{aligned} \sum_{i=u,c,t} V_{iq}V_{iq'}^* &= 0 \qquad (qq' = ds, db, sb), \\ \sum_{i=d,s,b} V_{qi}V_{q'i}^* &= 0 \qquad (qq' = uc, ut, ct). \end{aligned} \tag{3.19}$$

Each of these relations requires the sum of three complex quantities to vanish. Therefore, they can be geometrically represented in the complex plane as triangles and are called "unitarity triangles". It is a feature of the CKM matrix that all six unitarity triangles have equal areas. Moreover, the area of each unitarity triangle equals $|J_{\mathrm{CKM}}|/2$ while the sign of J_{CKM} gives the direction of the complex vectors around the triangles.

The triangle which corresponds to the relation

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0, \tag{3.20}$$

has its three sides of roughly the same length, of $\mathcal{O}(\lambda^3)$—see Eq. (3.8). Furthermore, both the lengths of its sides and its angles are experimentally accessible. For these reasons, the term "the unitarity triangle" is reserved for Eq. (3.20).

We further define the rescaled unitarity triangle. It is derived from Eq. (3.20) by choosing a phase convention such that $(V_{cd}V_{cb}^*)$ is real and dividing the lengths of all sides by $|V_{cd}V_{cb}^*|$. The rescaled unitarity triangle is similar to the unitarity triangle. Two vertices of the rescaled unitarity triangle are fixed at (0,0) and (1,0). The coordinates of the remaining vertex correspond to the Wolfenstein parameters (ρ, η). The rescaled unitarity triangle is shown in Fig. 1. The lengths of the two complex sides are

$$R_u \equiv \left|\frac{V_{ud}V_{ub}}{V_{cd}V_{cb}}\right| = \sqrt{\rho^2+\eta^2}, \qquad R_t \equiv \left|\frac{V_{td}V_{tb}}{V_{cd}V_{cb}}\right| = \sqrt{(1-\rho)^2+\eta^2}. \tag{3.21}$$

The three angles of the unitarity triangle are defined as follows:

$$\alpha \equiv \arg\left[-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right], \qquad \beta \equiv \arg\left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right], \qquad \gamma \equiv \arg\left[-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right]. \tag{3.22}$$

They are physical quantities and can be independently measured, as we discuss below. Another commonly used notation is $\phi_1 = \beta$, $\phi_2 = \alpha$, and $\phi_3 = \gamma$. Note that in the standard parametrization $\gamma = \delta$.

3.2 Tree level determination of the CKM parameters

The charged current weak interactions allow the determination of CKM parameters from tree level processes. There is an inherent difficulty in determining the CKM parameters: While the SM Lagrangian has the quarks as its degrees of freedom, in Nature they appear only within hadrons. There are various tools to overcome this difficulty, particularly for semileptonic decays, such as isospin symmetry and heavy quark symmetry.

At tree level, the W-mediated interactions lead to only FCCC processes. These suffice, however, to over-constrain the CKM parameters. The most useful processes are semileptonic ones. Here we give a short summary of the results:

- Processes related to $d \to u\ell^-\bar{\nu}$ transitions give $|V_{ud}| = 0.97370 \pm 0.00014$.
- Processes related to $s \to u\ell^-\bar{\nu}$ transitions give $|V_{us}| = 0.2245 \pm 0.0008$.
- Processes related to $c \to d\ell^+\nu$ or to $\nu_\mu + d \to c + \mu^-$ transitions give $|V_{cd}| = 0.221 \pm 0.004$.
- Processes related to $c \to s\ell^+\nu$ or to $c\bar{s} \to \ell^+\nu$ transitions give $|V_{cs}| = 0.987 \pm 0.011$.
- Processes related to $b \to c\ell^-\bar{\nu}$ transitions give $|V_{cb}| = 0.0410 \pm 0.0014$.
- Processes related to $b \to u\ell^-\bar{\nu}$ transitions give $|V_{ub}| = 0.00382 \pm 0.00024$.

There are two additional classes of tree level processes that depend on the CKM parameters:

- Processes related to single top production in hadron colliders give $|V_{tb}| = 1.013 \pm 0.030$.
- Processes related to $b \to sc\bar{u}$ and $b \to su\bar{c}$ transitions give $\gamma = (72 \pm 5)^o$.

These eight distinct classes of processes depend on only four CKM parameters. The system is thus over-constrained and tests the SM.

The values of λ and A can be straightforwardly extracted from the measurements of $|V_{us}|$ and $|V_{cb}|$, respectively:

$$\lambda = 0.2250 \pm 0.0007, \qquad A = 0.83 \pm 0.02. \tag{3.23}$$

The values of ρ and η are extracted mainly from combining the measurements of $|V_{ub}|$ and γ, as shown in Fig. 2:

$$\rho = 0.13 \pm 0.03, \qquad \eta = 0.38 \pm 0.02. \tag{3.24}$$

The fact that the ranges of the four parameters in Eqs. (3.23) and (3.24) are consistent with all the measurements means that the SM passes the test successfully.

Note that the error bars on the determination here, Eqs. (3.23) and (3.24), is larger than the one in Eq. (3.10). The reason is that here we only consider tree level processes.

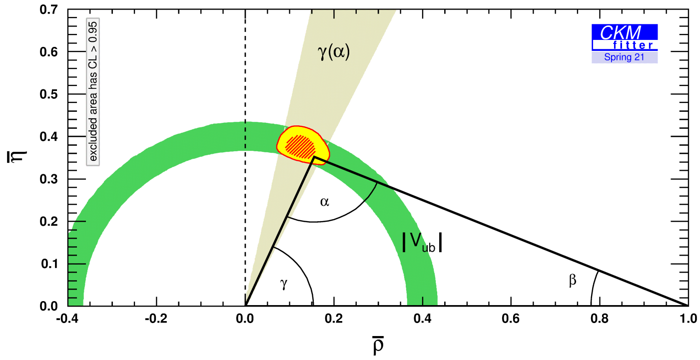


Fig. 2: Allowed region in the (ρ, η) plane from SM tree level processes (taken from Ref. [10]).

4 Flavor changing neutral current (FCNC) processes

4.1 No FCNC at tree level

Historically, the strong suppression of FCNC played a very important role in constructing the SM. At present it continues to play a significant role in testing the SM and in searching for new physics. In this subsection we explain why, within the SM, there are no tree level contributions to FCNC processes. Since there is no symmetry that forbids FCNC in the quark sector, there are loop contributions to these processes. These are discussed in the following subsections.

The W-boson cannot mediate FCNC processes at tree level, since it couples to up–down pairs, or to neutrino–charged-lepton pairs. Only neutral bosons could mediate FCNC at tree level. The SM has four neutral bosons: the gluon, the photon, the Z-boson and the Higgs-boson. As derived explicitly in Section 2, within the SM all of them couple diagonally in the mass basis, and therefore cannot mediate FCNC at tree level. Here we explain the qualitative features of the SM that lead to this situation.

4.1.1 Photon- and gluon-mediated FCNC

As concerns the massless gauge bosons, the gluon and the photon, their couplings are flavor-universal and, in particular, flavor-diagonal. This is guaranteed by gauge invariance. The universality of the kinetic terms in the canonical basis requires universality of the gauge couplings related to the unbroken symmetries. Hence neither the gluon nor the photon can mediate flavor changing processes at tree level. Since we require that extensions of the SM respect the local $SU(3)_C \times U(1)_{\rm EM}$ symmetry, this result holds in all such extensions.

4.1.2 Z-mediated FCNC

The Z-boson, similarly to the W-boson, corresponds to a broken gauge symmetry (as manifest in the fact that it is massive). Hence, there is no fundamental symmetry principle that forbids flavor changing Z couplings. Yet, as we explicitly find in Section 2.3.2, in the SM the Z couplings are universal and diagonal.

The key point is the following. The Z couplings are proportional to $T_3 - Q\sin^2\theta_W$. A sector of mass eigenstates is characterized by spin, $SU(3)_C$ representation and $U(1)_{\rm EM}$ charge. While Q must be the same for all the flavors in a given sector, there are two possibilities regarding T_3:

1. All mass eigenstates in this sector originate from interaction eigenstates in the same $SU(2)_L \times U(1)_Y$ representation, and thus have the same T_3 and Y.
2. The mass eigenstates in this sector mix interaction eigenstates with the same $Q = T_3 + Y$ but different $SU(2)_L \times U(1)_Y$ representations and, more specifically, different T_3 and Y.

Let us examine the Z couplings in the interaction and mass bases for several flavors of (hypothetical) fermions in the same $SU(3)_C \times U(1)_{\rm EM}$ representation:

1. In the first class, the Z couplings in the fermion interaction basis are universal, namely they are proportional to the unit matrix (times $T_3 - Q\sin^2\theta_W$ of the relevant interaction eigenstates). The rotation to the mass basis maintains the universality:

$$V_{fM} \times \mathbf{1} \times V_{fM}^\dagger = \mathbf{1}, \qquad (f = u, d, e; M = L, R). \tag{4.1}$$

2. In the second class, the Z couplings in the fermion interaction basis are diagonal but not universal. Each diagonal entry is proportional to the relevant $T_3 - Q\sin^2\theta_W$. Generally in this case, the rotation to the mass basis does not maintain the diagonality:

$$V_{fM} \times \hat{G}_{\rm diagonal} \times V_{fM}^\dagger = G_{\rm non-diagonal}, \qquad (f = u, d, e; M = L, R). \tag{4.2}$$

The SM fermions belong to the first class: All fermion mass eigenstates with a given chirality and in a given $SU(3)_C \times U(1)_{\rm EM}$ representation come from the same $SU(3)_C \times SU(2)_L \times U(1)_Y$ representation. For example, all the left-handed up quark mass eigenstates, which are in the $(3)_{+2/3}$ representation, come from interaction eigenstates in the $(3,2)_{+1/6}$ representation. This is the reason that the SM predicts universal Z couplings to fermions. If, for example, Nature had also left-handed quarks in the $(3,1)_{+2/3}$ representation, then the Z couplings in the left-handed up sector would be non-universal and the Z-boson could mediate FCNC, such as $t \to cZ$ decay, at tree level.

4.1.3 Higgs-mediated FCNC

The Yukawa couplings of the Higgs boson are not universal. In fact, in the interaction basis, they are given by completely general 3×3 matrices. Yet, as explained in Section 2.3.4, in the fermion mass basis they are diagonal. The reason is that the fermion mass matrix is proportional to the corresponding Yukawa matrix and, consequently, the mass matrix and the Yukawa matrix are simultaneously diagonalized. The general condition for the absence of Higgs-mediated FCNC at tree level is that the only source of masses for any fermion type is a single Higgs field.

The relevant features of the SM are the following:

1. All the SM fermions are chiral and charged (under $SU(2)_L \times U(1)_Y$), and therefore there are no bare mass terms.
2. The scalar sector has a single Higgs doublet.

In contrast, either of the following possible extensions would lead to flavor changing Higgs couplings:

1. There are quarks and/or leptons in vector-like representations, and thus there are bare mass terms.
2. There is more than one $SU(2)_L$-doublet scalar that couples to a specific type of fermions.

Subsection 4.1.4 provides an example of the first case. Subsection 4.1.5 provides an example of the second case.

We conclude that, within the SM, all FCNC processes in the quark sector are loop suppressed (while in the lepton sector they are forbidden). However, in extensions of the SM, FCNC can appear at the tree level, mediated by the Z-boson, the Higgs boson, or by new massive bosons.

To summarize, FCNC processes cannot be mediated at tree level in the SM. Yet, since there is no symmetry that forbids them in the quark sector, they are mediated at the loop level. Concretely, the W-mediated interactions lead to FCNC at the one-loop level. Since the W-boson couplings are charged current flavor changing, an even number of insertions of W-boson couplings are needed to generate an FCNC process. We consider two classes of FCNC based on the change in F (the charge under the global $[U(1)]^6$ flavor symmetry of the QCD Lagrangian):

– FCNC decays ($\Delta F = 1$ processes) have two insertions of W-couplings.
– Neutral meson mixings ($\Delta F = 2$ processes) have four insertions of W-couplings.

4.1.4 SM1.5: FCNC at tree level

Consider a model with the SM gauge group and pattern of SSB, but with only three quark flavors: u, d, s. Such a situation cannot fit into a model with all left-handed quarks in doublets of $SU(2)_L$. How can we incorporate the interactions of the strange quark in this picture? The solution that we now describe is wrong. Yet, it is of historical significance and, moreover, helps us to understand some of the unique properties of the SM described above. In particular, it leads to FCNC at tree level. We define the three flavor Standard Model (SM1.5) as follows (we ignore the lepton sector):

– The symmetry is a local

$$G_{\mathrm{SM}} = SU(3)_C \times SU(2)_L \times U(1)_Y. \tag{4.3}$$

– It is spontaneously broken by the VEV of a single Higgs scalar,

$$\phi(1,2)_{+1/2}, \quad (\langle \phi^0 \rangle = v/\sqrt{2}), \tag{4.4}$$

$$G_{\mathrm{SM}} \to SU(3)_C \times U(1)_{\mathrm{EM}} \quad (Q_{\mathrm{EM}} = T_3 + Y). \tag{4.5}$$

– The colored fermion representations are the following:

$$Q_L(3,2)_{+1/6}, \;\; D_L(3,1)_{-1/3}, \;\; U_R(3,1)_{+2/3}, \;\; D_{Ri}(3,1)_{-1/3} \;\; (i = 1,2). \tag{4.6}$$

We point out two important ingredients that are different from the SM:

1. There are quarks in a vector-like representation ($D_L + D_R$);
2. Not all $(3)_{-1/3}$ quarks come from the same type of $SU(2)_L \times U(1)_Y$ representations.

We first note that D_L does not couple to the W-bosons:

$$\mathcal{L}_W = \frac{g}{2}\overline{Q_L}\not{W}_b\tau_b Q_L. \tag{4.7}$$

The Yukawa interactions are given by

$$\mathcal{L}_{\rm Yuk} = -y_u\overline{Q_L}\tilde{\phi}U_R - Y_i^d\overline{Q_L}\phi D_{Ri} + {\rm h.c.}. \tag{4.8}$$

Unlike the SM, we now have bare mass terms for fermions:

$$\mathcal{L}_q = -m_{di}\overline{D_L}D_{Ri} + {\rm h.c.}. \tag{4.9}$$

Given that there is a single up generation, the interaction basis is also the up mass basis. Explicitly, we identify the up-component of Q_L with u_L (and denote the down component of the doublet as d_{uL}), and U_R with u_R. With the SSB, we have the following mass terms:

$$-\mathcal{L}_{\rm mass} = (\overline{d_{uL}}\ \overline{D_L})\begin{pmatrix} Y_{d1}\frac{v}{\sqrt{2}} & Y_{d2}\frac{v}{\sqrt{2}} \\ m_{d1} & m_{d2}\end{pmatrix}\begin{pmatrix} D_{R1} \\ D_{R2}\end{pmatrix} + y_u\frac{v}{\sqrt{2}}\overline{u_L}u_R + {\rm h.c.}. \tag{4.10}$$

We now rotate to the down mass basis:

$$V_{dL}\begin{pmatrix} Y_{d1}\frac{v}{\sqrt{2}} & Y_{d2}\frac{v}{\sqrt{2}} \\ m_{d1} & m_{d2}\end{pmatrix}V_{dR}^\dagger = \begin{pmatrix} m_d & \\ & m_s\end{pmatrix}. \tag{4.11}$$

The resulting mixing matrix for the charged current interactions is a 1×2 matrix:

$$-\mathcal{L}_{W,q} = \frac{g}{\sqrt{2}}\overline{u_L}\not{W}^+(\cos\theta_C\ \ \sin\theta_C)\begin{pmatrix} d_L \\ s_L\end{pmatrix} + {\rm h.c.}, \tag{4.12}$$

where θ_C is the rotation angle of V_{dL}. The neutral current interactions in the left-handed down sector are neither universal nor diagonal:

$$\begin{aligned}\mathcal{L}_{Z,q} \;=\; & \frac{g}{c_W}\left[\left(\frac{1}{2}-\frac{2}{3}s_W^2\right)\overline{u_L}\not{Z}u_L - \frac{2}{3}s_W^2\,\overline{u_R}\not{Z}u_R + \frac{1}{3}s_W^2(\overline{d_L}\not{Z}d_L + \overline{s_L}\not{Z}s_L + \overline{d_R}\not{Z}d_R + \overline{s_R}\not{Z}s_R)\right] \\ & -\frac{g}{2c_W}(\overline{d_L}\ \overline{s_L})\not{Z}\begin{pmatrix}\cos^2\theta_C & \cos\theta_C\sin\theta_C \\ \cos\theta_C\sin\theta_C & \sin^2\theta_C\end{pmatrix}\begin{pmatrix} d_L \\ s_L\end{pmatrix}.\end{aligned} \tag{4.13}$$

The Higgs interactions in the down sector are neither proportional to the mass matrix nor diagonal:

$$\mathcal{L}^q_{\rm Yuk} = y_u h\overline{u_L}u_R + h(\overline{d_L}\ \overline{s_L})\left[V_{dL}\begin{pmatrix} Y_{d1} & Y_{d2} \\ 0 & 0\end{pmatrix}V_{dR}^\dagger\right]\begin{pmatrix} d_R \\ s_R\end{pmatrix} + {\rm h.c.}. \tag{4.14}$$

Thus, in this model, both the Z-boson and the h-boson mediate FCNC at tree level. For example, $K_L\to\mu^+\mu^-$ and $K^0-\overline{K}^0$ mixing get Z- and h-mediated tree-level contributions.

4.1.5 2HDM: FCNC at tree level

Consider a model with two Higgs doublets. The symmetry structure, the pattern of spontaneous symmetry breaking, and the fermion content are the same as in the SM. However, the scalar content is extended:

– The scalar representations are

$$\phi_i(1,2)_{+1/2}, \quad i = 1,2. \tag{4.15}$$

We are particularly interested in the modification of the Yukawa terms:

$$\mathcal{L}_{\rm Yuk} = (Y_k^u)_{ij}\overline{Q_{Li}}U_{Rj}\,\tilde{\phi}_k + (Y_k^d)_{ij}\overline{Q_{Li}}D_{Rj}\,\phi_k + (Y_k^e)_{ij}\overline{L_{Li}}E_{Rj}\,\phi_k + \text{h.c.}. \tag{4.16}$$

Without loss of generality, we can work in a basis (commonly called "the Higgs basis") (ϕ_A, ϕ_M), where one the Higgs doublets carries the VEV, $\langle\phi_M\rangle = v$, while the other has zero VEV, $\langle\phi_A\rangle = 0$. In this basis, Y_M^f is known and related to the fermions masses in the same way as the Yukawa matrices of the SM:

$$Y_M^f = \sqrt{2}M_f/v. \tag{4.17}$$

The entries in the Yukawa matrices Y_A^f are, however, free parameters and, in general, unrelated to the fermion masses. The rotation angle from the Higgs basis to the basis of neutral CP-even Higgs states, (ϕ_h, ϕ_H), is denoted by $(\alpha - \beta)$. The Yukawa matrix of the light Higgs field h is given by

$$Y_h^f = c_{\alpha-\beta}Y_A^f - s_{\alpha-\beta}Y_M^f. \tag{4.18}$$

Given the arbitrary structure of Y_A^f, the Higgs boson can have couplings that are neither proportional to the mass matrix nor diagonal.

It is interesting to note, however, that not all multi Higgs doublet models lead to flavor changing Higgs couplings. If all the fermions of a given sector couple to one and the same doublet, then the Higgs couplings in that sector would still be diagonal. For example, in a model with two Higgs doublets, ϕ_1 and ϕ_2, and Yukawa terms of the form

$$\mathcal{L}_{\rm Yuk} = Y_{ij}^u\overline{Q_{Li}}U_{Rj}\,\tilde{\phi}_2 + Y_{ij}^d\overline{Q_{Li}}D_{Rj}\,\phi_1 + Y_{ij}^e\overline{L_{Li}}E_{Rj}\,\phi_1 + \text{h.c.}, \tag{4.19}$$

the Higgs couplings are flavor diagonal:

$$Y_h^u = (c_\alpha/s_\beta)Y_M^u, \quad Y_h^d = -(s_\alpha/c_\beta)Y_M^d, \quad Y_h^e = -(s_\alpha/c_\beta)Y_M^e, \tag{4.20}$$

where β [α] is the rotation angle from the (ϕ_1, ϕ_2) basis to the (ϕ_A, ϕ_M) [(ϕ_h, ϕ_H)] basis. In the physics jargon, we say that such models have *natural flavor conservation* (NFC) [11–13].

4.2 CKM and GIM suppressions in FCNC decays

In this section, we discuss FCNC meson decays, which are $\Delta F = 1$ processes. To demonstrate the generic features of one-loop FCNC, we consider the example of $s \to d$ transitions. Since the change of flavor QNs is $\Delta s = -\Delta d = 1$, this transition belongs to the class of $\Delta F = 1$ processes. The FCNC part

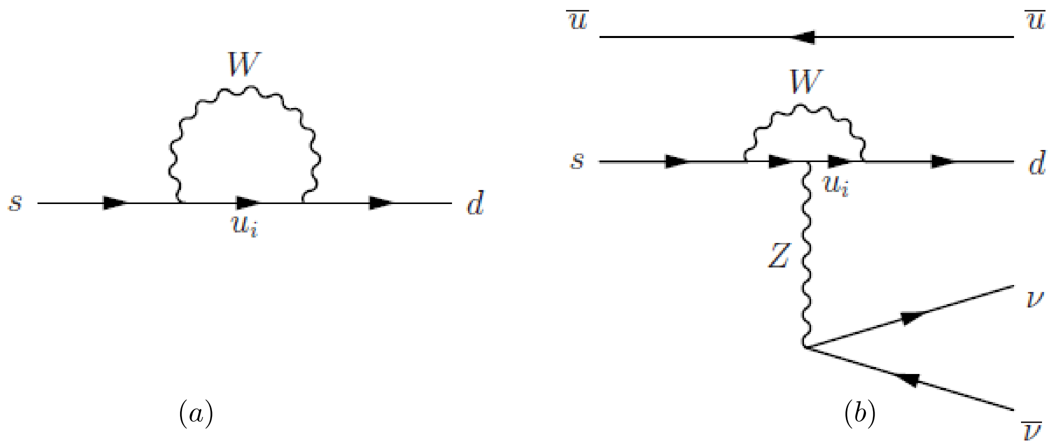


Fig. 3: (a) One loop diagrams for $\Delta F = 1$ $s \to d$ FCNC. (b) A one loop diagram that contributes to $K^- \to \pi^- \nu\bar{\nu}$.

of any process that involves $s \to d$ transition is plotted in Fig. 3(a). For example, in Fig. 3(b) we show a full diagram that contributes to the decay $K \to \pi\nu\bar{\nu}$, and which includes Fig. 3(a) as a sub-diagram (diagrams with such a topology are usually called penguin diagrams).

By inspecting the diagram in Fig. 3(a), we learn that its flavor structure is given by

$$\mathcal{A}_{s\to d} \sim \sum_{i=u,c,t} (V_{is}V_{id}^*)\; f(x_i), \qquad x_i = \frac{m_i^2}{m_W^2}, \tag{4.21}$$

where $f(x_i)$ depends on the specific decay. CKM unitarity implies

$$V_{ud}^* V_{us} + V_{cd}^* V_{cs} + V_{td}^* V_{ts} = 0. \tag{4.22}$$

We can then use this unitarity condition to eliminate one of the three CKM terms in the sum in Eq. (4.21). We choose to eliminate the u-term and write

$$\mathcal{A}_{s\to d} \sim \sum_{i=c,t} \left[f\left(x_i\right) - f\left(x_u\right)\right] V_{is}V_{id}^*. \tag{4.23}$$

We draw the following lessons:

- The contribution of the m_i-independent terms in $f(x_i)$ to $\mathcal{A}_{s\to d}$ vanishes when summed over all internal quarks.
- $\mathcal{A}_{s\to d}$ would vanish if the up-type quarks were all degenerate and, therefore, it must depend on the mass-splittings among the up-type quarks.

The explicit dependence on the mass-splittings among the quarks depends on the process. In many cases, for small x_i we have

$$f(x_i) \sim x_i. \tag{4.24}$$

Using this crude approximation we can write

$$\mathcal{A}_{s\to d} \sim \left[(x_t - x_u)V_{ts}V_{td}^* + (x_c - x_u)V_{cs}V_{cd}^*\right]. \tag{4.25}$$

Inspecting Eq. (4.25), we identify two suppression factors:

(i) *CKM suppression:* The amplitude is proportional to at least one off-diagonal CKM matrix element. Given the specific structure of the CKM matrix, off-diagonal elements are small. Specifically, $|V_{ts}V_{td}^*| \sim \lambda^5$ and $|V_{cs}V_{cd}^*| \sim \lambda$ where λ is the Wolfenstein parameter defined in Eq. (3.7).

(ii) *GIM suppression:* The amplitude is proportional to mass-squared differences between the up-type quarks. In particular, $(x_c - x_u) \sim (m_c/m_W)^2$. The suppression by factors of small quark masses is called the Glashow–Iliopoulos–Maiani (GIM) mechanism [14].

While we derive the results based on one specific example, the CKM and GIM suppressions play a role in all FCNC processes. For the other FCNC in the down sector, $b \to q$ with $q = d, s$, we have

$$\mathcal{A}_{b\to q} \sim (x_t - x_u)V_{tb}V_{tq}^* + (x_c - x_u)V_{cb}V_{cq}^*. \tag{4.26}$$

With regard to FCNC in the up sector, for $c \to u$, we have

$$\mathcal{A}_{c\to u} \sim (x_b - x_d)V_{cb}^*V_{ub} + (x_s - x_d)V_{cs}^*V_{us}, \tag{4.27}$$

and, for $t \to q$ with $q = u, c$, we have

$$\mathcal{A}_{t\to q} \sim (x_b - x_d)V_{tb}^*V_{qb} + (x_s - x_d)V_{ts}^*V_{qs}. \tag{4.28}$$

The CKM suppression applies to FCNC decay rates. It does not necessarily apply, however, to the corresponding branching ratios. The reason is that branching ratios depend on the ratio between the FCNC decay rate and the full decay width which, in the down sector, is CKM-suppressed, and thus the ratio of CKM factors is not necessarily small. In particular, the leading (FCCC) K decay rate is suppressed by $|V_{us}| \sim |V_{cs}V_{cd}|$ and the leading (FCCC) B decay rate is suppressed by $|V_{cb}| \sim |V_{tb}V_{ts}|$.

Several remarks are in order:

1. While the $f(x_i) \sim x_i$ approximation is not valid for the top quark, it gives a reasonable order of magnitude estimate, and we use it for the purpose of demonstration. For example, while $x_t/x_c \sim 10^4$, we have for $f(x)$ defined in Eq. (4.31), $f(x_t)/f(x_c) \approx 10^3$.
2. The exact form of the dependence on the mass splitting is process dependent, but in all cases the amplitude vanishes when the internal quarks are degenerate. We refer to quadratic dependence $[x_i - x_j]$ as hard GIM, and to logarithmic dependence $[\log(x_i/x_j)]$ as soft GIM.
3. The size of FCNC amplitudes increases with the mass of the internal quark. The reason that this does not violate the decoupling theorem is that the mass comes from SSB, so larger masses correspond to larger Yukawa couplings.

4.2.1 Examples: $K \to \pi\nu\bar{\nu}$ and $B \to \pi\nu\bar{\nu}$

As examples of $\Delta F = 1$ processes, we consider the semileptonic decays

$$K^+ \to \pi^+\nu\bar{\nu}, \qquad B^+ \to \pi^+\nu\bar{\nu}, \tag{4.29}$$

which proceed via the $\bar{s} \to \bar{d}\nu\bar{\nu}$ and $\bar{b} \to \bar{d}\nu\bar{\nu}$ transitions.

Consider the following ratios of FCNC-to-FCCC semileptonic decay rates:

$$\begin{aligned} R_{K\pi} &= \frac{\Gamma(K^+ \to \pi^+\nu\bar{\nu})}{\Gamma(K^+ \to \pi^0 e^+\nu)} = \frac{3}{2}\frac{g^4}{16\pi^2}\left|\frac{V_{ts}^* V_{td}[f(x_t) - f(x_u)] + V_{cs}^* V_{cd}[f(x_c) - f(x_u)]}{V_{us}}\right|^2, \\ R_{B\pi} &= \frac{\Gamma(B^+ \to \pi^+\nu\bar{\nu})}{\Gamma(B^+ \to \pi^0 e^+\nu)} = \frac{3}{2}\frac{g^4}{16\pi^2}\left|\frac{V_{tb}^* V_{td}[f(x_t) - f(x_u)] + V_{cb}^* V_{cd}[f(x_c) - f(x_u)]}{V_{ub}}\right|^2, \end{aligned} \tag{4.30}$$

where

$$f(x) = \frac{x}{8}\left[\frac{2+x}{1-x} - \frac{3x-6}{(1-x)^2}\log x\right]. \tag{4.31}$$

For small x, we have

$$f(x \ll 1) \approx \frac{x(3\log x + 1)}{4}. \tag{4.32}$$

Since $f(x_u) \ll 1$, we have to a very good approximation $f(x_i) - f(x_u) \approx f(x_i)$ for $i = c, t$.

A few comments are in order with regard to Eq. (4.30):

1. The factor of 3 comes from summing over the neutrino flavors, while the factor of $1/2$ is an isospin factor between the $P^+ \to \pi^+$ and $P^+ \to \pi^0$ ($P = K, B$) transitions.
2. The $g^4/16\pi^2$ is the loop suppression factor.
3. For $R_{K\pi}$, the t-term is CKM-suppressed, $|V_{ts}V_{td}/V_{us}| \sim \lambda^4$, but not GIM-suppressed, $f(x_t) = \mathcal{O}(1)$. The c-term is GIM-suppressed, $f(x_c) = \mathcal{O}(m_c^2/m_W^2)$, but not CKM-suppressed, $|V_{cs}V_{cd}/V_{us}| \simeq 1$. The two terms contribute comparably.
4. For $R_{B\pi}$, the t-term is neither CKM-suppressed, $|V_{tb}V_{td}/V_{ub}| = \mathcal{O}(1)$, nor GIM-suppressed, $f(x_t) = \mathcal{O}(1)$. The c-term is not CKM-suppressed, $|V_{cb}V_{cd}/V_{ub}| = \mathcal{O}(1)$, but it is GIM-suppressed, $f(x_c) = \mathcal{O}(m_c^2/m_W^2)$. Thus, the contribution of the c-term is negligible.
5. As a result of the different CKM and GIM suppression factors, we obtain numerically very different predictions:

$$R_{K\pi} \sim 10^{-9}, \qquad R_{B\pi} \sim 10^{-4}. \tag{4.33}$$

These predictions have not been fully tested yet, as we have only experimental upper bounds, $R_{K\pi} \lesssim 10^{-8}$ and $R_{B\pi} \lesssim 0.18$.

The comparison of $R_{K\pi}$ and $R_{B\pi}$ demonstrates how the CKM and GIM suppression factors depend crucially on the specific quarks involved, and how they come into play in determining the various FCNC rates. While for the two specific examples presented here there exist only experimental upper bounds, many FCNC decays have been observed and their rates measured. To date, all measured FCNC decay rates in the quark sector agree with the SM predictions.

Table 3: Measurements related to neutral meson mixing

Sector	CP-conserving	CP-violating
sd	$\Delta m_K/m_K = 7.0 \times 10^{-15}$	$\epsilon_K = 2.3 \times 10^{-3}$
cu	$\Delta m_D/m_D = 8.7 \times 10^{-15}$	$A_\Gamma/y_{\rm CP} \lesssim 0.05$
bd	$\Delta m_B/m_B = 6.3 \times 10^{-14}$	$S_{\psi K} = +0.699 \pm 0.017$
bs	$\Delta m_{B_s}/m_{B_s} = 2.1 \times 10^{-12}$	$S_{\psi\phi} = +0.046 \pm 0.020$

4.3 CKM and GIM suppressions in neutral meson mixing

A very useful class of FCNC is that of neutral meson mixing. Nature provides us with four pairs of neutral mesons: $K^0 - \overline{K}^0$, $B^0 - \overline{B}^0$, $B_s^0 - \overline{B}_s^0$, and $D^0 - \overline{D}^0$. Mixing in this context refers to a transition such as $K^0 \to \overline{K}^0$ ($\bar{s}d \to \bar{d}s$).[2] The experimental results for CP conserving and CP violating observables related to neutral meson mixing (mass splittings and CP asymmetries in tree level decays, respectively) are given in Table 3.

Neutral meson mixing is a $\Delta F = 2$ process. This phenomenon is observed and measured via meson oscillations, as discussed in Appendix B. In Appendix A we present the explicit SM calculation. In this section we show that the general lessons learned from $\Delta F = 1$ processes about the loop, CKM, and GIM suppression factors of FCNC, mostly carry over to $\Delta F = 2$ processes.

To demonstrate the features of $\Delta F = 2$ FCNC processes, we consider the example of $K^0 - \overline{K}^0$ mixing. It is generated by the $s\bar{d} \to d\bar{s}$ transition which is a $\Delta s = -\Delta d = 2$ process. The leading diagram for this transition is plotted in Fig. 4. By inspecting this diagram, we learn that its flavor structure is given by

$$\mathcal{A}_{s\bar{d}\to d\bar{s}} \sim \sum_{i,j=u,c,t} (V_{is}V_{id}^* V_{js}V_{jd}^*)\, S(x_i, x_j), \tag{4.34}$$

where $S(x_j, x_i)$ [$x_i \equiv m_i^2/m_W^2$] is given explicitly in Eq. (A.6). We draw the following lessons:

– The contribution of the m_i-independent terms in $S(x_i, x_j)$ to $\mathcal{A}_{s\bar{d}\to d\bar{s}}$ vanishes when summed over all internal quarks.

– $\mathcal{A}_{s\bar{d}\to d\bar{s}}$ would vanish if the up-type quarks were all degenerate and, therefore, it must depend on the mass-splittings among the up-type quarks.

To proceed, we use the unitarity condition of Eq. (4.22) and approximate $x_u = 0$ to eliminate the u-terms in the sum. We obtain:

$$\mathcal{A}_{s\bar{d}\to d\bar{s}} \sim (V_{cs}V_{cd}^*)^2 S(x_c, x_c) + 2V_{cs}V_{cd}^* V_{ts}V_{td}^* S(x_c, x_t) + (V_{ts}V_{td}^*)^2 S(x_t, x_t). \tag{4.35}$$

We conclude that $\Delta F = 2$ amplitudes have, in addition to the loop suppression factor, also the following suppression factors:

[2]These transitions involve four-quark operators. When calculating the matrix elements of these operators between meson-antimeson states, approximate symmetries of QCD are of no help. Instead, one uses lattice calculations to relate, for example, the $B^0 \to \overline{B}^0$ transition to the corresponding quark process, $\bar{b}d \to \bar{d}b$.

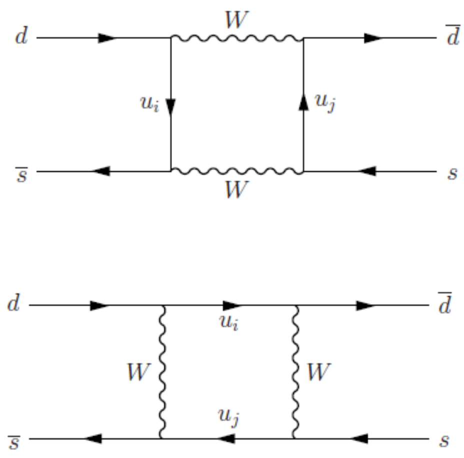


Fig. 4: The one loop diagrams for $\Delta F = 2$ FCNC.

(i) *CKM suppression:* The amplitude is proportional to a least two off-diagonal CKM matrix elements.

(ii) *GIM suppression:* The amplitude depends on the mass-squared differences between the up-type quarks.

While we derive the results based on one example, the CKM and GIM suppressions play a role in the mixing of all neutral mesons. In fact, there are three more $\Delta F = 2$ amplitudes that we should consider:

$$\begin{aligned}
\mathcal{A}_{b\bar{d}\to d\bar{b}} &\sim (V_{cb}V_{cd}^*)^2 S(x_c, x_c) + 2V_{cb}V_{cd}^*V_{tb}V_{td}^* S(x_c, x_t) + (V_{tb}V_{td}^*)^2 S(x_t, x_t), \\
\mathcal{A}_{b\bar{s}\to s\bar{b}} &\sim (V_{cb}V_{cs}^*)^2 S(x_c, x_c) + 2V_{cb}V_{cs}^*V_{tb}V_{ts}^* S(x_c, x_t) + (V_{tb}V_{ts}^*)^2 S(x_t, x_t), \\
\mathcal{A}_{c\bar{u}\to u\bar{c}} &\sim (V_{cs}^*V_{us})^2 S(x_s, x_s) + 2V_{cs}^*V_{us}V_{cb}^*V_{ub} S(x_s, x_b) + (V_{cb}^*V_{ub})^2 S(x_b, x_b),
\end{aligned} \tag{4.36}$$

which correspond to $B^0 - \overline{B}^0$, $B_s^0 - \overline{B_s}^0$, and $D^0 - \overline{D}^0$ mixing, respectively.

4.3.1 *Examples: Δm_K, Δm_B and Δm_{B_s}*

The hadronic process of $K^0 - \overline{K}^0$ mixing proceeds via the $s\bar{d} \to d\bar{s}$ quark transition, and leads to the mass splitting Δm_K between the two neutral kaon mass eigenstates. The SM calculation gives [see Eq. (A.10)]

$$\frac{|\Delta m_K|}{m_K} = \frac{g^4}{96\pi^2}\frac{m_K^2}{m_W^2}\frac{B_K f_K^2}{m_K^2}\left|(V_{cs}^*V_{cd})^2 S(x_c, x_c) + 2V_{cs}^*V_{cd}V_{ts}^*V_{td} S(x_c, x_t) + (V_{ts}^*V_{td})^2 S(x_t, x_t)\right|. \tag{4.37}$$

To estimate the relative size of the three terms, we note that

$$\frac{|V_{ts}^*V_{td}|}{|V_{cs}^*V_{cd}|} \sim 10^{-3}, \qquad \frac{S(x_c, x_t)}{S(x_c, x_c)} \sim 10, \qquad \frac{S(x_t, x_t)}{S(x_c, x_c)} \sim 10^4. \tag{4.38}$$

We conclude that the contributions of the terms proportional to $S(x_t, x_t)$ are smaller by a factor of $\mathcal{O}(100)$ than the contribution of the $S(x_c, x_c)$ term and can thus be neglected:

$$\frac{\Delta m_K}{m_K} \approx \frac{B_K f_K^2}{m_K^2} \times \frac{g^4}{96\pi^2} \times \frac{m_K^2}{m_W^2} \times |V_{cs}V_{cd}|^2 \times \frac{m_c^2}{m_W^2}. \tag{4.39}$$

The $B_K f_K^2/m_K^2 \sim O(1)$ factor encodes the QCD hadronic matrix element. The m_K^2/m_W^2 factor is related to the fact that the flavor changing processes are W-mediated, so we get the scale suppression. This factor is also present in FCCC tree level processes. The other three factors are the following:

- The $g^4/(96\pi^2)$ factor represents the one-loop suppression.
- The $|V_{cs}V_{cd}|^2$ factor represents the CKM suppression.
- The m_c^2/m_W^2 factor represents the GIM suppression.

The $B^0 - \overline{B}^0$ and $B_s - \overline{B}_s$ mixing amplitudes are given in Eqs. (A.7) and (A.8), respectively. In both cases the $S(x_t, x_t)$ is the largest of the S-functions while the CKM factors are of the same order in all three terms. We thus have

$$\frac{\Delta m_B}{m_B} \propto \frac{g^4}{96\pi^2}\left(\frac{m_t^2}{m_W^2}\right)|V_{tb}V_{td}|^2, \qquad \frac{\Delta m_{B_s}}{m_{B_s}} \propto \frac{g^4}{96\pi^2}\left(\frac{m_t^2}{m_W^2}\right)|V_{tb}V_{ts}|^2. \tag{4.40}$$

The GIM- and CKM-suppression factors are thus different among the various neutral meson systems of the down sector:

- $B_s^0 - \overline{B}_s^0$ mixing: CKM suppression by $|V_{tb}V_{ts}|^2 \sim 2 \times 10^{-3}$, and no GIM suppression;
- $B^0 - \overline{B}^0$ mixing: CKM suppression by $|V_{tb}V_{td}|^2 \sim 10^{-4}$, and no GIM suppression;
- $K^0 - \overline{K}^0$ mixing: CKM and GIM suppression by $|V_{cs}V_{cd}|^2(m_c^2/m_W^2) \sim 10^{-5}$.

We learn that the SM predicts hierarchy among the $\Delta F = 2$ processes:

$$\frac{\Delta m_K}{m_K} \ll \frac{\Delta m_B}{m_B} \ll \frac{\Delta m_{B_s}}{m_{B_s}}. \tag{4.41}$$

The experimental results,

$$\frac{\Delta m_K}{m_K} = 7.0 \times 10^{-15}, \qquad \frac{\Delta m_B}{m_B} = 6.3 \times 10^{-14}, \qquad \frac{\Delta m_{B_s}}{m_{B_s}} = 2.1 \times 10^{-12}, \tag{4.42}$$

show that this pattern is indeed realized in Nature.

4.3.2 CPV suppression

In some cases, CP violating (CPV) observables are CKM suppressed beyond their CP conserving (CPC) counterparts. Whether this is the case can be understood by examining the relevant unitarity triangle: The CPV observables depend on the area of it, while CPC observables depend on the length-squared of one side. Thus, in cases where the unitarity triangle is squashed (such as the sd and bs triangles), we can have a situation where the area of the triangle, $|J_{\rm CKM}|/2 \sim \lambda^6$, is much smaller than the length-squared of one of its sides, resulting in an extra suppression for CPV observables. Explicitly, for FCNC in the down sector, we have

$$\begin{aligned}
sd: \ J_{\rm CKM}/|V_{us}V_{ud}|^2 &= \mathcal{O}(\lambda^4),\\
bs: \ J_{\rm CKM}/|V_{tb}V_{ts}|^2 &= \mathcal{O}(\lambda^2),\\
bd: \ J_{\rm CKM}/|V_{tb}V_{td}|^2 &= \mathcal{O}(1).
\end{aligned} \tag{4.43}$$

CP asymmetries measure the ratios between the CPV difference between two CP-conjugate rates and the CPC sum of these rates:

- CPV in $K^0-\overline{K}^0$ mixing is the source of δ_L, the CP asymmetry in $K_L \to \pi\ell\nu$ defined in Eq. (C.13);
- CPV in the interference of $B_s - \overline{B_s}$ mixing with $b \to c\bar{c}s$ decay is the source of $\mathcal{I}m(\lambda_{\psi\phi})$, the CP asymmetry in $B_s \to \psi\phi$ defined similarly to Eq. (C.15);
- CPV in the interference of $B^0 - \overline{B}^0$ mixing with $b \to c\bar{c}d$ decay is the source of $\mathcal{I}m(\lambda_{D^+D^-})$, the CP asymmetry in $B \to D^+D^-$ defined in Eq. (C.15).

The pattern of a possible significant CP suppression in the sd sector, possible intermediate CP suppression in the bs sector, and no CP suppression in the bd sector, is manifest in the SM predictions:

$$\begin{aligned}
\delta_L &\propto J_{\rm CKM}/|V_{us}V_{ud}|^2 \sim 10^{-3},\\
\mathcal{I}m(\lambda_{\psi\phi}) &\propto J_{\rm CKM}/|V_{tb}V_{ts}|^2 \sim 10^{-2},\\
\mathcal{I}m(\lambda_{D^+D^-}) &\propto J_{\rm CKM}/|V_{tb}V_{td}|^2 \sim 1.
\end{aligned} \tag{4.44}$$

Experiments confirm this pattern:

$$\begin{aligned}
\delta_L &= (3.34 \pm 0.07) \times 10^{-3},\\
\mathcal{I}m(\lambda_{\psi\phi}) &= (5.0 \pm 2.0) \times 10^{-2},\\
\mathcal{I}m(\lambda_{D^+D^-}) &= -0.76^{+0.15}_{-0.13}.
\end{aligned} \tag{4.45}$$

4.4 Summary

Within the SM, we identify four possible suppression factors of FCNC processes relative to FCCC ones:

1. Loop suppression.
2. CKM suppression.
3. GIM suppression in processes that are not dominated by the top quark contribution.
4. CPV suppression in some of the processes related to squashed unitarity triangles.

5 Testing the CKM sector

Within the SM, the CKM matrix is the only source of flavor changing processes and of CP violation. In Section 3.2 we use only tree level processes to extract the values of CKM parameters. Here we add

FCNC to the set of CKM measurements to form a global test of the SM. The primary question is whether the long list of measurements can be fitted by the four CKM parameters.

5.1 $S_{\psi K_S}$

The CP asymmetry in $B \to \psi K_S$ decays plays a major role in testing the KM mechanism. Before we explain the test itself, we should understand why the theoretical interpretation of the asymmetry is exceptionally clean, and what are the theoretical parameters on which it depends, within and beyond the Standard Model.

The CP asymmetry in neutral B meson decays into final CP eigenstates f_{CP} is defined as follows:

$$\mathcal{A}_{f_{CP}}(t) \equiv \frac{d\Gamma/dt[\overline{B}^0_{\rm phys}(t) \to f_{CP}] - d\Gamma/dt[B^0{}_{\rm phys}(t) \to f_{CP}]}{d\Gamma/dt[\overline{B}^0_{\rm phys}(t) \to f_{CP}] + d\Gamma/dt[B^0{}_{\rm phys}(t) \to f_{CP}]} . \tag{5.1}$$

A detailed evaluation of this asymmetry is given in Appendix B. It leads to the following form:

$$\begin{aligned} \mathcal{A}_{f_{CP}}(t) &= S_{f_{CP}} \sin(\Delta m_B t) - C_{f_{CP}} \cos(\Delta m_B t), \\ S_{f_{CP}} &\equiv \frac{2\,\mathcal{I}m(\lambda_{f_{CP}})}{1+|\lambda_{f_{CP}}|^2}, \quad C_{f_{CP}} \equiv \frac{1-|\lambda_{f_{CP}}|^2}{1+|\lambda_{f_{CP}}|^2}, \end{aligned} \tag{5.2}$$

where

$$\lambda_{f_{CP}} = e^{-i\phi_B}(\overline{A}_{f_{CP}}/A_{f_{CP}}) . \tag{5.3}$$

Here ϕ_B refers to the phase of $M_{B\bar{B}}$ [see Eq. (C.3)]. Within the Standard Model, the corresponding phase factor is given by

$$e^{-i\phi_B} = (V_{tb}^* V_{td})/(V_{tb} V_{td}^*) . \tag{5.4}$$

The decay amplitudes A_f and $\overline{A}_f$ are defined in Eq. (B.1).

The $B^0 \to J/\psi K^0$ decay [15, 16] proceeds via the quark transition $\bar{b} \to \bar{c}c\bar{s}$. There are contributions from both tree (t) and penguin (p^{q_u}, where $q_u = u, c, t$ is the quark in the loop) diagrams (see Fig. 5) which carry different weak phases:

$$A_f = (V_{cb}^* V_{cs})\, t_f + \sum_{q_u = u,c,t} (V_{q_u b}^* V_{q_u s})\, p_f^{q_u} \tag{5.5}$$

(the distinction between tree and penguin contributions is a heuristic one, the separation by the operator that enters is more precise. For a detailed discussion of the more complete operator product approach, which also includes higher order QCD corrections, see, for example, Ref. [17]). Using CKM unitarity, these decay amplitudes can always be written in terms of just two CKM combinations:

$$A_{\psi K} = (V_{cb}^* V_{cs})\, T_{\psi K} + (V_{ub}^* V_{us})\, P_{\psi K}^u, \tag{5.6}$$

where $T_{\psi K} = t_{\psi K} + p_{\psi K}^c - p_{\psi K}^t$ and $P_{\psi K}^u = p_{\psi K}^u - p_{\psi K}^t$. A subtlety arises in this decay that is related to the fact that $B^0 \to J/\psi K^0$ and $\overline{B}^0 \to J/\psi \overline{K}^0$. A common final state, e.g. $J/\psi K_S$, can be reached via $K^0 - \overline{K}^0$ mixing. Consequently, the phase factor corresponding to neutral K mixing,

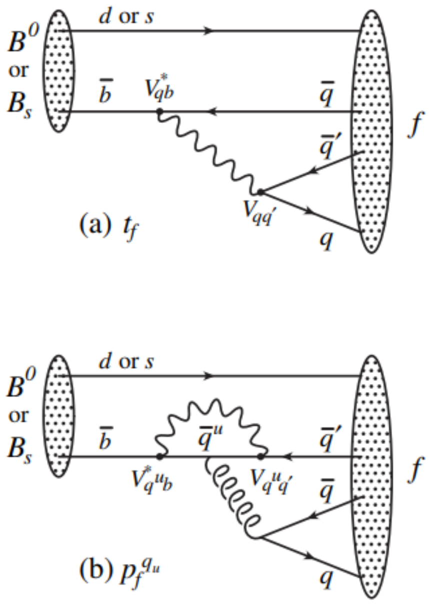


Fig. 5: Feynman diagrams for (a) tree and (b) penguin amplitudes contributing to $B^0 \to f$ or $B_s \to f$ via a $\bar{b} \to \bar{q}q\bar{q}'$ quark-level process.

$e^{-i\phi_K} = (V_{cd}^* V_{cs})/(V_{cd} V_{cs}^*)$, plays a role:

$$\frac{\overline{A}_{\psi K_S}}{A_{\psi K_S}} = -\frac{(V_{cb} V_{cs}^*)\, T_{\psi K} + (V_{ub} V_{us}^*)\, P_{\psi K}^u}{(V_{cb}^* V_{cs})\, T_{\psi K} + (V_{ub}^* V_{us})\, P_{\psi K}^u} \times \frac{V_{cd}^* V_{cs}}{V_{cd} V_{cs}^*}. \tag{5.7}$$

The crucial point is that, for $B \to J/\psi K_S$ and other $\bar{b} \to \bar{c}c\bar{s}$ processes, we can neglect the P^u contribution to $A_{\psi K}$, in the SM, to an approximation that is better than one percent:

$$|P_{\psi K}^u / T_{\psi K}| \times |V_{ub}/V_{cb}| \times |V_{us}/V_{cs}| \sim (\text{loop factor}) \times 0.1 \times 0.23 \lesssim 0.005. \tag{5.8}$$

Thus, to an accuracy better than one percent,

$$\lambda_{\psi K_S} = \left(\frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}\right) \left(\frac{V_{cb} V_{cd}^*}{V_{cb}^* V_{cd}}\right) = -e^{-2i\beta}, \tag{5.9}$$

where β is defined in Eq. (3.22), and consequently

$$S_{\psi K_S} = \sin 2\beta, \quad C_{\psi K_S} = 0 \tag{5.10}$$

(below the percent level, several effects modify this equation [18–21]).

Exercise 1: *Show that, if the $B \to \pi\pi$ decays were dominated by tree diagrams, then $S_{\pi\pi} = \sin 2\alpha$.*

Exercise 2: *Estimate the accuracy of the predictions $S_{\phi K_S} = \sin 2\beta$ and $C_{\phi K_S} = 0$.*

The experimental measurements give the following ranges [22]:

$$S_{\psi K_S} = +0.70 \pm 0.02, \quad C_{\psi K_S} = -0.005 \pm 0.015 \ . \tag{5.11}$$

5.2 Is the CKM picture self-consistent?

The present status of our knowledge of the absolute values of the various entries in the CKM matrix is given in Eq. (3.5). The values there take into account all the relevant tree-level and loop processes. Yet, as explained above, the test of the SM is stronger when we reduce the above to the four CKM parameters. Indeed, the following ranges for the four Wolfenstein parameters are consistent with all measurements:

$$\lambda = 0.2265 \pm 0.0005, \qquad A = 0.790 \pm 0.015, \qquad \rho = 0.14 \pm 0.02, \qquad \eta = 0.36 \pm 0.01. \tag{5.12}$$

For the purpose of demonstration, it is useful to project the individual constraints onto the (ρ, η) plane:

– Charmless semileptonic B decays can be used to extract

$$\left|\frac{V_{ub}}{V_{cb}}\right|^2 = \lambda^2(\rho^2 + \eta^2). \tag{5.13}$$

– $B \to DK$ decays can be used to extract

$$\tan\gamma = \left(\frac{\eta}{\rho}\right). \tag{5.14}$$

– $S_{\psi K_S}$, the CP asymmetry in $B \to \psi K_S$, is used to extract

$$\sin 2\beta = \frac{2\eta(1-\rho)}{(1-\rho)^2 + \eta^2}. \tag{5.15}$$

– The CP asymmetries of various $B \to \pi\pi$, $B \to \rho\pi$, and $B \to \rho\rho$ decays depend on the phase

$$\alpha = \pi - \beta - \gamma. \tag{5.16}$$

– The ratio between the mass splittings in the B and B_s systems depends on

$$\left|\frac{V_{td}}{V_{ts}}\right|^2 = \lambda^2[(1-\rho)^2 + \eta^2] \tag{5.17}$$

– The CP violation in $K \to \pi\pi$ decays, ϵ_K, depends in a complicated way on ρ and η.

The resulting constraints are shown in Fig. 6. The consistency of the various constraints is impressive. This is a triumph of the SM in that such a variety of measurements, with different sources of uncertainties, all agree to a high precision. We conclude that the flavor structure of the SM passes a highly non-trivial test.

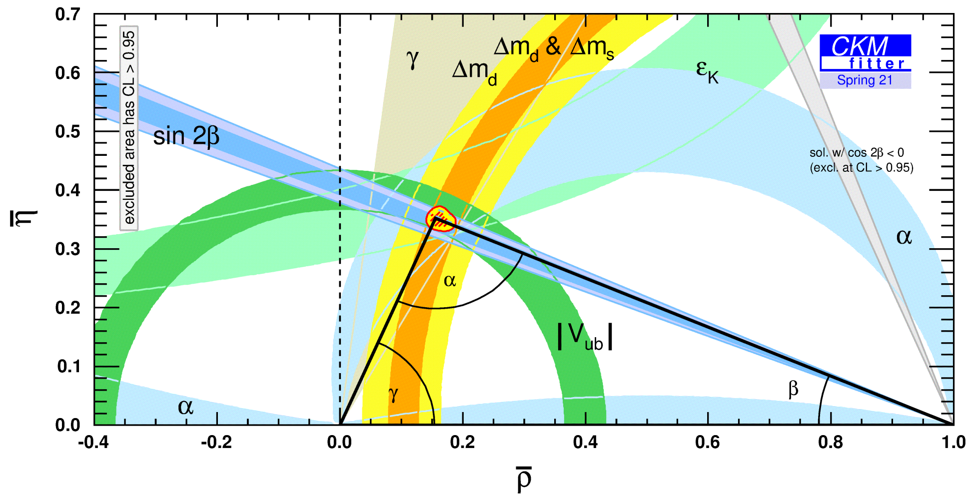


Fig. 6: Allowed region in the ρ, η plane. Superimposed are the individual constraints from charmless semileptonic B decays ($|V_{ub}|$), mass differences in the B^0 (Δm_d) and B_s (Δm_s) neutral meson systems, and CP violation in $K \to \pi\pi$ (ε_K), $B \to \psi K$ ($\sin 2\beta$), $B \to \pi\pi, \rho\pi, \rho\rho$ (α), and $B \to DK$ (γ). Taken from Ref. [10].

6 Probing BSM

In spite of the enormous experimental success of the SM, we know that the SM is not a complete theory of Nature. In this section, we explain this statement and discuss the formalism and the experimental probes to be used in case that the physics that extends the SM takes place at a high energy scale.

One obvious reason that we know that the SM is not the full theory of Nature is that it does not include gravity. There are, however, reasons to think that, beyond gravity and the SM list of elementary particles and fundamental interactions, there must exist degrees of freedom that are yet unknown to us. These reasons can be roughly divided into four classes:

1. Experiments: measurements that are inconsistent with the SM predictions.
2. Cosmology and astrophysics: observations that cannot be explained by the SM.
3. Fine-tuning: parameters whose values can be explained in the SM only with accidental fine-tuned cancellations among several contributions.
4. Clues: various non-generic features that are just parameterized in the SM, but not explained.

We elaborate on this list with specific examples in Sections 6.1 and 6.2.

Models that extend the SM by adding degrees of freedom (DoF), and often by imposing larger symmetries, come under the general name of “Beyond the SM,” or BSM for short. The fact that experiments have not observed any particles related to such hypothetical new fields tells us that either these new particles are very heavy, or that their couplings to the SM particles are very weak. In light BSM scenarios, where the new DoF are at or below the weak scale, the SM is not a good low energy effective theory. Each such feebly coupled BSM scenario requires a specific discussion of how to probe it. We do not discuss such theories any further.

The situation is different for heavy BSM scenarios, where the new DoF are much above the weak scale. There is a unified framework that allows one to understand the possible probes of heavy BSM scenarios, while remaining agnostic about the details of the new degrees of freedom. We present this framework in Section 6.4 and employ it in our discussion of BSM flavor physics.

Direct searches for BSM physics aim to produce the new particles on shell and study their properties. Numerous such searches have been conducted but, as of now, no BSM particle has been discovered. Instead, these searches have set combination of lower bounds on the masses and upper bounds on the couplings of such states to SM states. Roughly speaking, the lower bounds on the masses of particles with order one couplings to the SM particles are of order 1 TeV. What sets this scale is the center of mass energy of the most powerful accelerator in action (the LHC). Indirect searches for BSM physics aim to observe virtual effects of the new states at low energies. We discuss this method below.

6.1 Experimental and observational problems

There are several experimental results and observational data that cannot be explained within the SM. They provide the most direct evidence that we need to extend the SM. Here we present the three that are the most robust.

Neutrino masses. There are several, related, pieces of experimental evidence for BSM physics from the neutrino sector. All of these measurements prove that the neutrinos are massive, in contrast to the SM prediction that they are massless. First, measurements of the ratio of ν_μ to ν_e fluxes of atmospheric neutrinos and the directional dependence of the ν_μ flux are different from the SM predictions. Both facts are beautifully explained by neutrino masses and mixing which lead to $\nu_\mu - \nu_\tau$ oscillations. Second, measurements of the solar neutrino flux find that, while the Sun produces only electron-neutrinos, their flux on Earth is significantly smaller than the total flux of neutrinos. This puzzle is beautifully explained by $\nu_e - \nu_{\mu,\tau}$ mixing. Both the atmospheric neutrino result and the solar neutrino result are now confirmed by terrestrial accelerator and reactor neutrino experiments.

The baryon asymmetry of the universe (BAU). There exists observational evidence for BSM physics from cosmology. The features of the Cosmic Microwave Background (CMB) radiation imply a certain baryon asymmetry of the Universe. Similarly, the standard Big Bang Nucleosynthesis (BBN) scenario is consistent with the observed abundances of light elements only for a certain range of the baryon asymmetry, consistent with the CMB constraint. Baryogenesis, the dynamical generation of a baryon asymmetry, requires CP violation. The CP violation in the SM generates baryon asymmetry that is smaller by at least ten orders of magnitude than the observed asymmetry. This implies that there must exist new sources of CP violation, beyond the SM. Furthermore, baryogenesis requires a departure from thermal equilibrium at a very early time after the Big Bang, and the one provided by the SM is not of the right kind.

Dark matter (DM). The evidence for dark matter—particles that are EM neutral and do not carry the color charge of the strong interactions—comes from several observations: Rotation curves in galaxies, gravitational lensing, the CMB, and the large scale structure of the Universe. The neutrinos of the SM do constitute dark matter, but their abundance is too small to be all the dark matter abundance. Thus, there must exist DoF beyond those of the SM.

6.2 Theoretical considerations

Some of the SM parameters are small. We distinguish between two classes of small parameters. "Technically natural" small parameters are those that, if set to zero, the theory gains an extra symmetry. The small parameters that are not technically natural are those where the symmetry of the theory is not enlarged when setting them to zero. An equivalent way to distinguish the two classes is based on their renormalization properties: For technically natural parameters the renormalization is multiplicative, while for non-technically natural parameters it is additive. For a technically natural parameter, loop corrections are proportional to the parameter itself, and if the parameter is set to be small at tree level, it remains small to all orders in perturbation theory. For a non-technically natural parameter, the radiative corrections are not proportional to the tree level parameter, and in cases where the radiative corrections are much larger than the measured value of the parameter, the smallness of the parameter can only be maintained by fine-tuned cancellation between the tree level and loop level contributions.

The existence in the SM of small parameters that are not technically natural is suggestive of BSM frameworks, where the smallness of the parameters is protected against large radiative corrections by some symmetry. There are two parameters of this kind in the SM: m_h^2 and $\theta_{\rm QCD}$.

The Higgs fine-tuning problem. Within the SM, the mass-squared of the Higgs μ^2, gets additive, quadratically divergent, radiative corrections. Given that the SM is an effective theory, the radiative corrections must be finite and proportional to the scale above which the SM is no longer valid. The higher the cutoff scale above the weak scale, the stronger the fine-tuned cancellation between the tree-level mass-squared term and the radiative corrections must be. In particular, if there is no BSM physics below $m_{\rm Pl}$, the bare mass-squared term and the loop contributions have to cancel each other to an accuracy of about thirty four orders of magnitude. Among the theories that aim to solve the Higgs fine-tuning problem, supersymmetry and composite Higgs have been intensively studied and searched for.

The strong CP problem. The CP violating $\theta_{\rm QCD}$ parameter contributes to the electric dipole moment of the neutron d_N. The experimental upper bound on d_N puts an upper bound on $\theta_{\rm QCD}$ of $\mathcal{O}(10^{-9})$. The smallness of $\theta_{\rm QCD}$ is not technically natural. Among the theories that aim to solve the strong CP problem, the Peccei–Quinn mechanism, and its prediction of an ultra-light pseudoscalar, the axion, have been intensively studied and searched for.

Other features of the SM parameters provide hints for the existence of BSM physics because they are non-generic, but they are not related to non-technically natural small parameters. We mention two of them below.

The flavor parameters. The Yukawa couplings are small (except for y_t) and hierarchical. For example, the electron Yukawa is of $\mathcal{O}(10^{-5})$. These are technically natural small numbers, but their non-generic structure—smallness and hierarchy—is suggestive of BSM physics. Among the theories that aim to solve this puzzle, the Froggatt–Nielsen mechanism, $U(2)$ flavor models, and models of extra dimensions, have been intensively studied.

Grand unification. The three gauge couplings of the strong, weak, and electromagnetic interactions seem to converge to a unified value at a high energy scale. The SM cannot explain this fact, which is just accidental within this model. Yet, it can be explained if the gauge group of the SM is part of a larger simple group. This idea is called Grand Unified Theory, or GUT, and among the relevant unifying groups, $SU(5)$ and $SO(10)$ have been intensively studied.

6.3 The BSM scale

The SM has a single mass scale that we call the weak scale and denote by $\Lambda_{\rm EW}$. It can be represented by the masses of the weak force carriers, m_W or m_Z, or by the VEV of the Higgs field, v. As an order of magnitude estimate, we take $\Lambda_{\rm EW} \sim 10^2$ GeV.

Some of the problems of the SM presented above are suggestive of where the BSM scale lies. We present these well-motivated scales in decreasing order. Of course, there could be more than one scale for the BSM physics.

The Planck scale, $m_{\rm Pl} \sim 10^{19}$ GeV. The Planck scale constitutes a cut-off scale of all QFTs. At this scale, gravitational effects become as important as the known gauge interactions and cannot be neglected.

The GUT scale, $\Lambda_{\rm GUT} \sim 10^{16}$ GeV. The GUT scale is the one where the three gauge couplings of the SM roughly unify. It is an indication that at that scale, the GUT symmetry group is broken into the SM symmetry group. For example, in $SU(5)$ GUT models, $\Lambda_{\rm GUT}$ can be represented by the VEV of the scalar field that breaks $SU(5) \to SU(3) \times SU(2) \times U(1)$, or by the masses of the gauge bosons that correspond to the broken $SU(5)$ generators.

The seesaw scale, $\Lambda_\nu \sim 10^{15}$ GeV. The value of the neutrino masses m_ν hints that new degrees of freedom appear at or below the so-called seesaw scale, $\Lambda_\nu \sim v^2/m_\nu$. This scale is intriguingly close to the GUT scale, and thus the two might be in fact related to the same BSM physics.

The Higgs fine-tuning scale $\Lambda_{\rm FT} \sim 1$ TeV. No fine tuning is necessary to explain the smallness of m_h^2 if radiative corrections are cut-off at a scale $\Lambda_{\rm FT}$ of order $4\pi m_h/y_t \sim$ TeV.

The WIMP scale $\Lambda_{\rm DM} \sim 1$ TeV. If the DM particles are weakly interacting massive particles (WIMPs), the cross section of their annihilation that is required to explain the DM abundance is of order $1/(20\ {\rm TeV})^2$. If the relevant coupling is of order α_W, the relevant scale is of order 1 TeV.

6.4 The SMEFT

As argued above, the SM is not a full theory of Nature. If the BSM degrees of freedom are at a scale $\Lambda \gg \Lambda_{\rm EW}$, then the SM is a good low energy effective theory which is valid below Λ. In such a case, the SM Lagrangian should be extended to include all nonrenormalizable terms, suppressed by powers of Λ:

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \frac{1}{\Lambda} O_{d=5} + \frac{1}{\Lambda^2} O_{d=6} + \cdots . \tag{6.1}$$

Here $\mathcal{L}_{\rm SM}$ is the renormalizable SM Lagrangian and $O_{d=n}$ represents operators that are products of SM fields, of overall dimension n in the fields, and transforming as singlets under the SM gauge group. The SM extended to include such non-renormalizable term is called the SM effective field theory, or SMEFT for short. For physics at an energy scale E well below Λ, the effects of operators of dimension $n > 4$ are suppressed by $(E/\Lambda)^{n-4}$. Thus, in general, the higher the dimension of an operator, the smaller its effect at low energies.

Nonrenormalizable operators are generated by extensions of the SM, which introduce new degrees of freedom that are much heavier than the electroweak scale. By studying nonrenormalizable operators, we allow the most general extension of the SM and remain agnostic about its specific structure. At the same time, constraints on nonrenormalizable terms can be translated into constraints on specific BSM models.

The low energy effects of nonrenormalizable operators are small. Thus, when we study them, we have to consider also loop effects in the SM. We can classify the effects of including loop corrections and nonrenormalizable terms into three broad categories:

1. *Forbidden processes:* Various processes are forbidden by the accidental symmetries of the SM. Nonrenormalizable terms, but not loop corrections, can break these accidental symmetries and allow the forbidden processes to occur. Examples include neutrino masses, and proton decay. In particular, neutrino masses and mixing violate the lepton flavor symmetries of the SM.
2. *Rare processes:* Within the SM, various processes that are not forbidden do not occur at tree level. Here both loop corrections and nonrenormalizable terms can contribute. Examples include FCNC processes.
3. *Tree level processes:* Often tree level processes in a particular sector depend on a small subset of the SM parameters. This situation leads to relations among different processes within this sector. These relations are violated by both loop effects and nonrenormalizable terms. Here, precision measurements and precision theory calculations are needed to observe these small effects. Examples include electroweak precision measurements.

As concerns the last two types of effects, where loop corrections and nonrenormalizable terms may both contribute, their use in phenomenology can be divided into two eras. Before all the SM particles have been directly discovered and all the SM parameters measured, one could assume the validity of the renormalizable SM and indirectly measure the properties of the yet unobserved SM particles. Indeed, the masses of charm quark, the top quark, and the Higgs boson were first indirectly measured in this way. Once all the SM particles have been observed and the parameters measured directly, the loop corrections can be quantitatively determined, and the effects of nonrenormalizable terms in the SMEFT can be unambiguously probed. Thus, at present, all three classes of processes serve to search for BSM physics.

In this section, we go beyond testing the self-consistency of the CKM picture of flavor physics and CP violation. The aim is to quantify how much room is left for BSM physics in the flavor sector and to translate these constraints into lower bounds on the scale of higher-dimension flavor-violating operators in the SMEFT. We make the following working assumption:

– The contribution of new physics to FCCC processes, where the SM contributions are tree-level, can be neglected.

On the other hand, we allow BSM physics of arbitrary size and phase to contribute to FCNC processes.

6.5 New physics contributions to $B^0 - \overline{B^0}$ mixing

We consider BSM effects in the FCNC process of $B^0-\overline{B^0}$ mixing, which plays a role in the mass splitting Δm_B and in the CP asymmetry $S_{\psi K_S}$. The SM amplitude is given by Eq. (A.7). The modification of the mixing amplitude by general BSM physics can be parameterized as follows:

$$M_{B\overline{B}} = M_{B\overline{B}}^{\rm SM}\Delta_d, \tag{6.2}$$

where Δ_d is a dimensionless complex parameter. BSM physics will be signalled by $\Delta_d \neq 1$. Our aim is to find the phenomenological constraints on Δ_d.

Our first step is to use all relevant tree level processes which, under our assumption, can be used to determine the CKM parameters. This was done in Section 3.2 and the results of this fit were shown in Fig. 2. Our second step is to use $\Delta B = 2$ processes to determine Δ_d:

– The mass splitting between the two neutral B-mesons is given by

$$\Delta m_B = 2|M_{B\overline{B}}^{\rm SM}(\rho,\eta)| \times |\Delta_d|. \tag{6.3}$$

– The CP asymmetry in $B \to \psi K_S$ is given by

$$S_{\psi K_S} = \sin\left[2\text{arctan}\left(\frac{\eta}{1-\rho}\right) + \arg(\Delta_d)\right]. \tag{6.4}$$

The results of the fit are (see Fig. 7)

$$\mathcal{R}e(\Delta_d) = +0.94^{+0.18}_{-0.15}, \qquad \mathcal{I}m(\Delta_d) = -0.11^{+0.11}_{-0.05}. \tag{6.5}$$

We learn that BSM physics can contribute to the $B^0 - \overline{B}^0$ mixing amplitude up to about 20% of the SM contribution.

Analogous upper bounds can be obtained for BSM contributions to the $K^0 - \overline{K}^0$ and $B_s^0 - \overline{B}_s^0$ mixing amplitudes.

6.6 Probing the SMEFT

Assuming that new degrees of freedom that have flavor changing couplings to quarks are much heavier than the electroweak breaking scale, their effects on low energy processes, such as neutral meson mixing, can be presented as higher dimension operators. Then, bounds such as Eq. (6.5), constrain the coefficients of such operators.

Consider a simple example, where we have a single dimension-six operator that contributes to Δ_d:

$$\frac{z_{bd}}{\Lambda^2}(\overline{Q_{Ld}}\gamma_\mu Q_{Lb})(\overline{Q_{Ld}}\gamma^\mu Q_{Lb}). \tag{6.6}$$

where Q_{Ld} and Q_{Lb} are the $SU(2)$-doublet quark fields whose $T_3 = -1/2$ members are the d_L and b_L fields [see Eq. (2.21)], and where we separated the coefficient into a dimensionless complex coupling, z_{bd}, and a high energy scale, Λ. We further define $\tilde{\Lambda} = \Lambda/\sqrt{z_{bd}}$. We consider the bound that can be obtained from Δm_B. Comparing Eqs. (A.1) and (6.6), we obtain

$$|\Delta_d| - 1 = \left|\frac{1}{\tilde{\Lambda}^2 C_{\rm SM}}\right| \approx \frac{1}{|\tilde{\Lambda}^2|} \times \frac{2\pi^2}{|V_{td}^* V_{tb}|^2 S(x_t,x_t) G_F^2 m_W^2}. \tag{6.7}$$

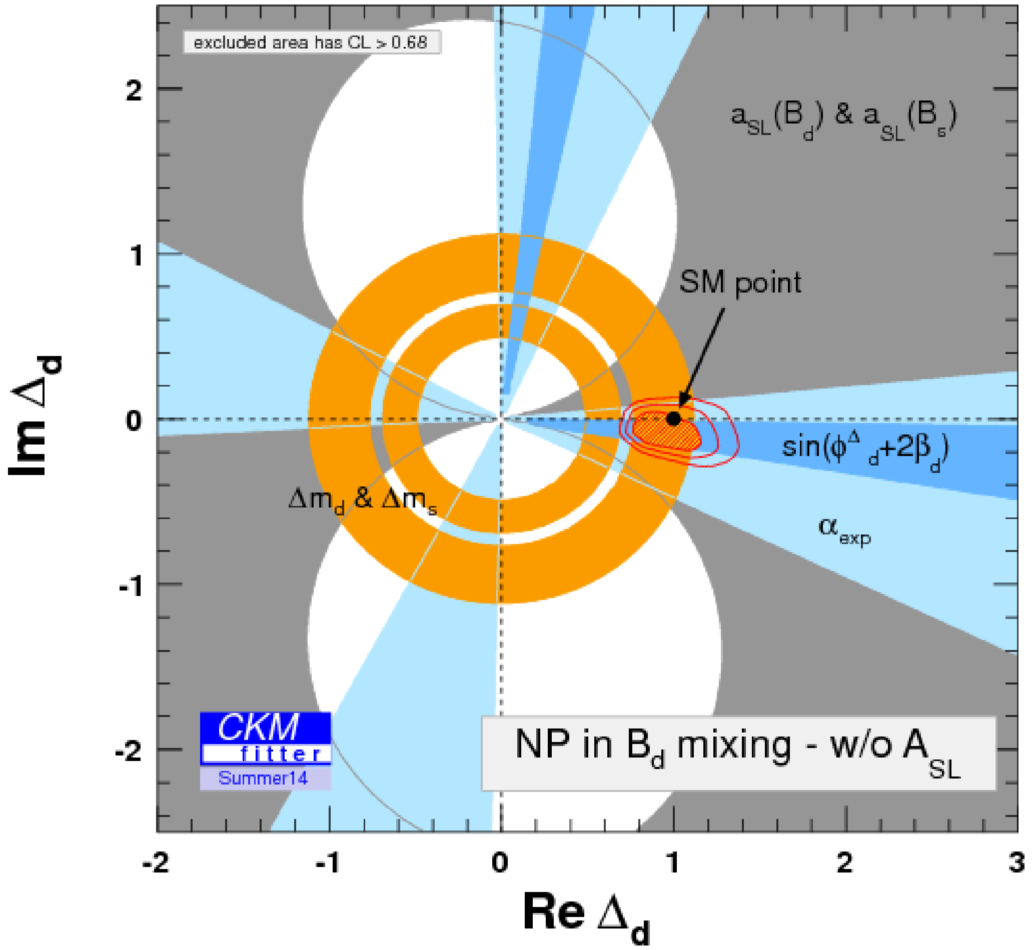


Fig. 7: Allowed region in the $(\mathcal{R}e\Delta_d, \mathcal{I}m\Delta_d)$ plane. Superimposed are the individual constraints from the mass differences in the B^0 (Δm_d), CP violation in $B \to \psi K$ $(S_{\psi K})$, and CP violation in semileptonic B^0 decay $(a_{\rm SL})$. Taken from Ref. [10].

The bound of Eq. (6.5) translates into a lower bound on $\tilde{\Lambda}$:

$$\tilde{\Lambda} \gtrsim 10^3 \text{ TeV}. \tag{6.8}$$

Using $S_{\psi K_S}$, one can obtain analogous bounds for various operators that contribute to CP violation in $B^0 - \overline{B}^0$ mixing. We can also obtain bounds for operators that affect other $\Delta F = 2$ processes. Some of these bounds are given in Table 4.

The following points are worth emphasizing:

1. BSM physics can contribute to FCNC at a level comparable to the SM contributions even if it takes place at a scale that is five orders of magnitude above the electroweak scale.
2. If the BSM physics has a generic flavor structure, that is $z_{ij} = \mathcal{O}(1)$, then its scale must be above 10^4 TeV.

Table 4: Lower bounds from CPC and CPV $\Delta F = 2$ processes on the scale of new physics $\tilde{\Lambda}$, for $(\overline{Q_{Li}}\gamma_\mu Q_{Lj})(\overline{Q_{Li}}\gamma^\mu Q_{Lj})$ operators.

i, j	$\tilde{\Lambda}$ [TeV] CPC	$\tilde{\Lambda}$ [TeV] CPV	Observables
s, d	9.8×10^2	1.6×10^4	Δm_K; ϵ_K
b, d	6.6×10^2	9.3×10^2	Δm_B; $S_{\psi K}$
b, s	1.4×10^2	2.5×10^2	Δm_{B_s}; $S_{\psi\phi}$

3. It could be that there are new particles with mass of order a TeV, but then their flavor structure must be far from generic, $|z_{ij}| \ll 1$.
4. The pattern of the bounds—those from K^0 are stronger than those from B^0 which are stronger than those from B_s—is directly related to the strength of the flavor (CKM and GIM) suppression we have in the SM, as discussed in Section 4.3.1. The reason for that is that the experimental accuracy and the QCD uncertainties are similar for the three cases.

7 The new physics flavor puzzle

7.1 A model independent discussion

Given that the SM is only an effective low energy theory, non-renormalizable terms must be added to $\mathcal{L}_{\rm SM}$. These are terms of dimension higher than four in the fields which, therefore, have couplings that are inversely proportional to the scale of new physics $\Lambda_{\rm NP}$.

The lowest dimension non-renormalizable terms are dimension-five:

$$-\mathcal{L}^{\rm dim-5}_{\rm Seesaw} = \frac{Z^\nu_{ij}}{\Lambda_{\rm NP}} L_{Li} L_{Lj} \phi\phi + {\rm h.c.}. \tag{7.1}$$

These are the seesaw terms, leading to neutrino masses.

Exercise 3: *How does the global symmetry breaking pattern in Eq. (2.47) change when Eq. (7.1) is taken into account?*

Exercise 4: *What is the number of physical lepton flavor parameters in this case? Identify these parameters in the mass basis.*

As concerns quark flavor physics, consider, for example, the following dimension-six set of operators:

$$\mathcal{L}^{\rm dim-6}_{\Delta F=2} = \sum_{i\neq j} \frac{z_{ij}}{\Lambda^2} (\overline{Q_{Li}}\gamma_\mu Q_{Lj})^2, \tag{7.2}$$

where the z_{ij} are dimensionless couplings. These terms contribute to the mass splittings between the corresponding two neutral mesons. As discussed in the previous section, the consistency of the experimental results with the SM predictions for neutral meson mixing allows us to impose the condition

$|M^{\rm NP}_{P\bar P}| < |M^{\rm SM}_{P\bar P}|$ for $P = K, B, B_s$, which implies that

$$\Lambda > \frac{3.4 \text{ TeV}}{|V_{ti}^* V_{tj}|/|z_{ij}|^{1/2}} \sim \begin{cases} 9 \times 10^3 \text{ TeV} \times |z_{sd}|^{1/2} \\ 4 \times 10^2 \text{ TeV} \times |z_{bd}|^{1/2} \\ 7 \times 10^1 \text{ TeV} \times |z_{bs}|^{1/2} \end{cases} \tag{7.3}$$

The first lesson that we draw from these bounds on Λ is that new physics can contribute to FCNC at a level comparable to the SM contributions even if it takes place at a scale that is six orders of magnitude above the electroweak scale. A second lesson is that if the new physics has a generic flavor structure, that is $z_{ij} = \mathcal{O}(1)$, then its scale must be above $10^4 - 10^5$ TeV (or, if the leading contributions involve electroweak loops, above $10^3 - 10^4$ TeV). *If indeed $\Lambda \gg$ TeV, it means that we have misinterpreted the hints from the fine-tuning problem and the dark matter puzzle.*

A different lesson can be drawn from the bounds on z_{ij}. *It could be that the scale of new physics is of order TeV, but its flavor structure is far from generic.* Specifically, if new particles at the TeV scale couple to the SM fermions, then there are two ways in which their contributions to FCNC processes, such as neutral meson mixing, can be suppressed: degeneracy and alignment. Either of these principles, or a combination of both, signifies non-generic structure.

One can use the language of effective operators also for the SM, integrating out all particles significantly heavier than the neutral mesons (that is, the top, the Higgs and the weak gauge bosons). Thus, the scale is $\Lambda_{\rm SM} \sim m_W$. Since the leading contributions to neutral meson mixings come from box diagrams, the z_{ij} coefficients are suppressed by α_2^2. To identify the relevant flavor suppression factor, one can employ the spurion formalism. For example, the flavor transition that is relevant to $B^0 - \overline{B}^0$ mixing involves $\overline{d_L} b_L$ which transforms as $(8,1,1)_{SU(3)^3_q}$. The leading contribution must then be proportional to $(Y^u Y^{u\dagger})_{13} \propto y_t^2 V_{tb} V_{td}^*$. Indeed, an explicit calculation (using VIA for the matrix element and neglecting QCD corrections) gives[3]

$$\frac{2M_{B\bar B}}{m_B} \approx -\frac{\alpha_2^2}{12}\frac{f_B^2}{m_W^2} S_0(x_t)(V_{tb}V_{td}^*)^2, \tag{7.4}$$

where $x_i = m_i^2/m_W^2$ and

$$S_0(x) = \frac{x}{(1-x)^2}\left[1 - \frac{11x}{4} + \frac{x^2}{4} - \frac{3x^2 \ln x}{2(1-x)}\right]. \tag{7.5}$$

Similar spurion analyses, or explicit calculations, allow us to extract the weak and flavor suppression factors that apply in the SM:

$$\begin{aligned} \mathcal{I}m(z_{sd}^{\rm SM}) &\sim \alpha_2^2 y_t^2 |V_{td}V_{ts}|^2 \sim 1 \times 10^{-10}, \\ z_{sd}^{\rm SM} &\sim \alpha_2^2 y_c^2 |V_{cd}V_{cs}|^2 \sim 5 \times 10^{-9}, \\ \mathcal{I}m(z_{cu}^{\rm SM}) &\sim \alpha_2^2 y_b^2 |V_{ub}V_{cb}|^2 \sim 2 \times 10^{-14}, \\ z_{bd}^{\rm SM} &\sim \alpha_2^2 y_t^2 |V_{td}V_{tb}|^2 \sim 7 \times 10^{-8}, \\ z_{bs}^{\rm SM} &\sim \alpha_2^2 y_t^2 |V_{ts}V_{tb}|^2 \sim 2 \times 10^{-6}. \end{aligned} \tag{7.6}$$

[3] A detailed derivation can be found in Appendix B of Ref. [23].

(We did not include $z_{cu}^{\rm SM}$ in the list because it requires a more detailed consideration. The naively leading short distance contribution is $\propto \alpha_2^2(y_s^4/y_c^2)|V_{cs}V_{us}|^2 \sim 5\times10^{-13}$. However, higher dimension terms can replace a y_s^2 factor with $(\Lambda/m_D)^2$ [24]. Moreover, long distance contributions are expected to dominate. In particular, peculiar phase space effects [25, 26] have been identified which are expected to enhance Δm_D to within an order of magnitude of the its measured value. The CP violating part, on the other hand, is dominated by short distance physics.)

It is clear then that contributions from new physics at $\Lambda_{\rm NP} \sim 1$ TeV should be suppressed by factors that are comparable or smaller than the SM ones. Why does that happen? This is the new physics flavor puzzle.

The fact that the flavor structure of new physics at the TeV scale must be non-generic means that flavor measurements are a good probe of the new physics. Perhaps the best-studied example is that of supersymmetry. Here, the spectrum of the superpartners and the structure of their couplings to the SM fermions will allow us to probe the mechanism of dynamical supersymmetry breaking.

7.2 Lessons for supersymmetry from neutral meson mixing

We consider, as an example, the contributions from the box diagrams involving the squark doublets of the second and third generations, $\tilde{Q}_{L2,3}$, to the $B_s - \overline{B_s}$ mixing amplitude. The contributions are proportional to $K_{3i}^{d*}K_{2i}^{d}K_{3j}^{d*}K_{2j}^{d}$, where K^d is the mixing matrix of the gluino couplings to a left-handed down quark and their supersymmetric squark partners ($\propto [(\delta_{LL}^d)_{23}]^2$ in the mass insertion approximation). We work in the mass basis for both quarks and squarks. A detailed derivation can be found in Ref. [27]. It gives:

$$M_{B_s\overline{B_s}} = \frac{\alpha_s^2 m_{B_s} f_{B_s}^2 B_{B_s}\eta_{\rm QCD}}{108 m_{\tilde{d}}^2}[11\tilde{f}_6(x)+4xf_6(x)]\frac{(\Delta\tilde{m}_{\tilde{d}}^2)^2}{\tilde{m}_{\tilde{d}}^4}(K_{32}^{d*}K_{22}^{d})^2. \tag{7.7}$$

Here $m_{\tilde{d}}$ is the average mass of the two squark generations, $\Delta m_{\tilde{d}}^2$ is the mass-squared difference, and $x = m_{\tilde{g}}^2/m_{\tilde{d}}^2$.

Equation (7.7) can be translated into our generic language:

$$\begin{aligned} \Lambda_{\rm NP} &= m_{\tilde{q}}, \\ z_1^{bs} &= \frac{11\tilde{f}_6(x)+4xf_6(x)}{18}\alpha_s^2\left(\frac{\Delta\tilde{m}_{\tilde{d}}^2}{m_{\tilde{d}}^2}\right)^2(K_{32}^{d*}K_{22}^{d})^2 \approx 10^{-4}(\delta_{23}^{LL})^2, \end{aligned} \tag{7.8}$$

where, for the last approximation, we took the example of $x=1$ [and used, correspondingly, $11\tilde{f}_6(1)+4f_6(1)=1/6$], and defined

$$\delta_{23}^{LL} = \left(\frac{\Delta\tilde{m}_{\tilde{d}}^2}{m_{\tilde{d}}^2}\right)(K_{32}^{d*}K_{22}^{d}). \tag{7.9}$$

Similar expressions can be derived for the dependence of $K^0-\overline{K^0}$ on $(\delta_{MN}^d)_{12}$, $B^0-\overline{B^0}$ on $(\delta_{MN}^d)_{13}$, and $D^0-\overline{D^0}$ on $(\delta_{MN}^u)_{12}$. Then we can use the constraints of Table 4 to put upper bounds on $(\delta_{MN}^q)_{ij}$. Some examples are given in Table 5 (see Ref. [28] for details and list of references).

We learn that, in most cases, we need $\delta_{ij}^q/m_{\tilde{q}} \ll 1/\text{TeV}$. One can immediately identify three generic ways in which supersymmetric contributions to neutral meson mixing can be suppressed:

Table 5: The phenomenological upper bounds on $(\delta^q_{LL})_{ij}$ and $\langle\delta^q_{ij}\rangle = \sqrt{(\delta^q_{LL})_{ij}(\delta^q_{RR})_{ij}}$. Here $q = u, d$ and $M = L, R$. The constraints are given for $m_{\tilde{q}} = 1$ TeV and $x = m^2_{\tilde{g}}/m^2_{\tilde{q}} = 1$. We assume that the phases could suppress the imaginary part by a factor of ~ 0.3. Taken from Ref. [28].

q	ij	$(\delta^q_{LL})_{ij}$	$\langle\delta^q_{ij}\rangle$
d	12	0.03	0.002
d	13	0.2	0.07
d	23	0.2	0.07
u	12	0.1	0.008

1. Heaviness: $m_{\tilde{q}} \gg 1$ TeV;
2. Degeneracy: $\Delta m^2_{\tilde{q}} \ll m^2_{\tilde{q}}$;
3. Alignment: $K^q_{ij} \ll 1$.

When heaviness is the only suppression mechanism, as in split supersymmetry [29], the squarks are very heavy and supersymmetry no longer solves the fine tuning problem. If we want to maintain supersymmetry as a solution to the fine tuning problem, either degeneracy or alignment or a combination of both is needed. This means that the flavor structure of supersymmetry is not generic, as argued in the previous section.

Take, for example, $(\delta^d_{LL})_{12} \leq 0.03$. Naively, one might expect the alignment to be of order $(V_{cd}V^*_{cs}) \sim 0.2$, which is far from sufficient by itself. Barring a very precise alignment ($|K^d_{12}| \ll |V_{us}|$) and accidental cancellations, we are led to conclude that the first two squark generations must be quasi-degenerate. Actually, by combining the constraints from $K^0 - \overline{K^0}$ mixing and $D^0 - \overline{D^0}$ mixing, one can show that this is the case independently of assumptions about the alignment [30–32]. Analogous conclusions can be drawn for many TeV-scale new physics scenarios: a strong level of degeneracy is required (for definitions and detailed analysis, see Ref. [33]).

Exercise 5: *Does $K^d_{31} \sim |V_{ub}|$ suffice to satisfy the Δm_B constraint with neither degeneracy nor heaviness? (Use the two generation approximation and ignore the second generation.)*

Is there a natural way to make the squarks degenerate? Degeneracy requires that the 3×3 matrix of soft supersymmetry breaking mass-squared terms $\tilde{m}^2_{Q_L} \simeq \tilde{m}^2_{\tilde{q}}\mathbf{1}$. We have mentioned already that flavor universality is a generic feature of gauge interactions. Thus, the requirement of degeneracy is perhaps a hint that supersymmetry breaking is *gauge mediated* to the MSSM fields.

7.3 Minimal flavor violation (MFV)

If supersymmetry breaking is gauge mediated, the squark mass matrices for $SU(2)_L$-doublet and $SU(2)_L$-singlet squarks have the following form at the scale of mediation m_M:

$$\begin{aligned}
\tilde{M}^2_{U_L}(m_M) &= \left(m^2_{\tilde{Q}_L} + D_{U_L}\right)\mathbf{1} + M_u M_u^\dagger, \\
\tilde{M}^2_{D_L}(m_M) &= \left(m^2_{\tilde{Q}_L} + D_{D_L}\right)\mathbf{1} + M_d M_d^\dagger,
\end{aligned}$$

$$\begin{aligned}
\tilde{M}^2_{U_R}(m_M) &= \left(m^2_{\tilde{U}_R} + D_{U_R}\right)\mathbf{1} + M_u^\dagger M_u, \\
\tilde{M}^2_{D_R}(m_M) &= \left(m^2_{\tilde{D}_R} + D_{D_R}\right)\mathbf{1} + M_d^\dagger M_d,
\end{aligned} \tag{7.10}$$

where $D_{q_A} = [(T_3)_{q_A} - (Q_{\rm EM})_{q_A} s_W^2] m_Z^2 \cos 2\beta$ are the D-term contributions. Here, the only source of the $SU(3)^3_q$ breaking are the SM Yukawa matrices.

This statement holds also when the renormalization group evolution is applied to find the form of these matrices at the weak scale. Taking the scale of the soft breaking terms $m_{\tilde{q}_A}$ to be somewhat higher than the electroweak breaking scale m_Z allows us to neglect the D_{q_A} and M_q terms in Eq. (7.10). Then we obtain

$$\begin{aligned}
\tilde{M}^2_{Q_L}(m_Z) &\sim m^2_{\tilde{Q}_L}\left(r_3 \mathbf{1} + c_u Y^u Y^{u\dagger} + c_d Y^d Y^{d\dagger}\right), \\
\tilde{M}^2_{U_R}(m_Z) &\sim m^2_{\tilde{U}_R}\left(r_3 \mathbf{1} + c_{uR} Y^{u\dagger} Y^u\right), \\
\tilde{M}^2_{D_R}(m_Z) &\sim m^2_{\tilde{D}_R}\left(r_3 \mathbf{1} + c_{dR} Y^{d\dagger} Y^d\right).
\end{aligned} \tag{7.11}$$

Here r_3 represents the universal RGE contribution that is proportional to the gluino mass ($r_3 = \mathcal{O}(6) \times (M_3(m_M)/m_{\tilde{q}}(m_M))$) and the c-coefficients depend logarithmically on m_M/m_Z and can be of $\mathcal{O}(1)$ when m_M is not far below the GUT scale.

Models of gauge mediated supersymmetry breaking (GMSB) provide a concrete example of a large class of models that obey a simple principle called *minimal flavor violation* (MFV) [34]. This principle guarantees that low energy flavor changing processes deviate only very little from the SM predictions. The basic idea can be described as follows. The gauge interactions of the SM are universal in flavor space. The only breaking of this flavor universality comes from the three Yukawa matrices, Y^u, Y^d and Y^e. If this remains true in the presence of the new physics, namely Y^u, Y^d and Y^e are the only flavor non-universal parameters, then the model belongs to the MFV class.

Let us now formulate this principle in a more formal way, using the language of spurions that we presented in section 2.4. The Standard Model with vanishing Yukawa couplings has a large global symmetry, see Eqs. (2.44, 2.45). In this section we concentrate only on the quarks. The non-Abelian part of the flavor symmetry for the quarks is $SU(3)^3_q$ of Eq. (2.45) with the three generations of quark fields transforming as follows:

$$Q_L(3,1,1), \quad U_R(1,3,1), \quad D_R(1,1,3). \tag{7.12}$$

The Yukawa interactions,

$$\mathcal{L}^q_{\rm Yuk} = \overline{Q_L} Y^d D_R H + \overline{Q_L} Y^u U_R H_c, \tag{7.13}$$

($H_c = i\tau_2 H^*$) break this symmetry. The Yukawa couplings can thus be thought of as spurions with the following transformation properties under $SU(3)^3_q$ [see Eq. (2.48)]:

$$Y^u \sim (3, \bar{3}, 1), \qquad Y^d \sim (3, 1, \bar{3}). \tag{7.14}$$

When we say "spurions", we mean that we pretend that the Yukawa matrices are fields which transform under the flavor symmetry, and then require that all the Lagrangian terms, constructed from the SM fields, Y^d and Y^u, must be (formally) invariant under the flavor group $SU(3)^3_q$. Of course, in reality, $\mathcal{L}^q_{\rm Yuk}$ breaks $SU(3)^3_q$ precisely because $Y^{d,u}$ are *not* fields and do not transform under the symmetry.

The idea of minimal flavor violation is relevant to extensions of the SM, and can be applied in two ways:

1. If we consider the SM as a low energy effective theory, then all higher-dimension operators, constructed from SM-fields and Y-spurions, are formally invariant under $G_{\rm global}$.
2. If we consider a full high-energy theory that extends the SM, then all operators, constructed from SM and the new fields, and from Y-spurions, are formally invariant under $G_{\rm global}$.

Exercise 6: *Use the spurion formalism to argue that, in MFV models, the $K_L \to \pi^0\nu\bar\nu$ decay amplitude is proportional to $y_t^2 V_{td}V_{ts}^*$.*

Exercise 7: *Find the flavor suppression factors in the z_i^{bs} coefficients, if MFV is imposed, and compare to the bounds in Table 4.*

Examples of MFV models include models of supersymmetry with gauge-mediation or with anomaly-mediation of its breaking.

8 The Standard Model flavor puzzle

The SM has thirteen flavor parameters: six quark Yukawa couplings, four CKM parameters (three angles and a phase), and three charged lepton Yukawa couplings (one can use fermions masses instead of the fermion Yukawa couplings, $y_f = \sqrt{2}m_f/v$). The orders of magnitudes of these thirteen dimensionless parameters are as follows:

$$\begin{aligned}
y_t &\sim 1, \quad y_c \sim 10^{-2}, \quad y_u \sim 10^{-5}, \\
y_b &\sim 10^{-2}, \quad y_s \sim 10^{-3}, \quad y_d \sim 10^{-4}, \\
y_\tau &\sim 10^{-2}, \quad y_\mu \sim 10^{-3}, \quad y_e \sim 10^{-6}, \\
|V_{us}| &\sim 0.2, \quad |V_{cb}| \sim 0.04, \quad |V_{ub}| \sim 0.004, \quad \delta_{\rm KM} \sim 1.
\end{aligned} \tag{8.1}$$

Only two of these parameters are clearly of $\mathcal{O}(1)$, the top-Yukawa and the KM phase. The other flavor parameters exhibit smallness and hierarchy. Their values span six orders of magnitude. It may be that this set of numerical values are just accidental. More likely, the smallness and the hierarchy have a reason. The question of why there is smallness and hierarchy in the SM flavor parameters constitutes "The Standard Model flavor puzzle."

The motivation to think that there is indeed a structure in the flavor parameters is strengthened by considering the values of the four SM parameters that are not flavor parameters, namely the three gauge couplings and the Higgs self-coupling:

$$g_s \sim 1, \quad g \sim 0.6, \quad e \sim 0.3, \quad \lambda \sim 0.12. \tag{8.2}$$

This set of values does seem to be a random distribution of order-one numbers, as one would naively expect.

A few examples of mechanisms that were proposed to explain the observed structure of the flavor parameters are the following:

- An approximate Abelian symmetry ("The Froggatt–Nielsen mechanism" [35]);
- An approximate non-Abelian symmetry (see e.g. Ref. [36]);
- Conformal dynamics ("The Nelson–Strassler mechanism" [37]);
- Location in an extra dimension [38];
- Loop corrections (see e.g. Ref. [39]).

We take as an example the Froggatt–Nielsen mechanism.

8.1 The Froggatt–Nielsen (FN) mechanism

Small numbers and hierarchies are often explained by approximate symmetries. For example, the small mass splitting between the charged and neutral pions finds an explanation in the approximate isospin (global $SU(2)$) symmetry of the strong interactions.

Approximate symmetries lead to selection rules which account for the size of deviations from the symmetry limit. Spurion analysis is particularly convenient to derive such selection rules. The Froggatt–Nielsen mechanism postulates a $U(1)_H$ symmetry, that is broken by a small spurion ϵ_H. Without loss of generality, we assign ϵ_H a $U(1)_H$ charge of $H(\epsilon_H) = -1$. Each SM field is assigned a $U(1)_H$ charge. In general, different fermion generations are assigned different charges, hence the term 'horizontal symmetry'. The rule is that each term in the Lagrangian, made of SM fields and the spurion, should be formally invariant under $U(1)_H$.

The approximate $U(1)_H$ symmetry thus leads to the following selection rules:

$$\begin{aligned} Y^u_{ij} &= \epsilon_H^{|H(\bar{Q}_i)+H(U_j)+H(\phi_u)|}, \\ Y^d_{ij} &= \epsilon_H^{|H(\bar{Q}_i)+H(D_j)+H(\phi_d)|}, \\ Y^e_{ij} &= \epsilon_H^{|H(\bar{L}_i)+H(E_j)-H(\phi_d)|}. \end{aligned} \tag{8.3}$$

As a concrete example, we take the following set of charges:

$$\begin{aligned} H(\bar{Q}_i) &= H(U_i) = H(E_i) = (2,1,0), \\ H(\bar{L}_i) &= H(D_i) = (0,0,0), \\ H(\phi_u) &= H(\phi_d) = 0. \end{aligned} \tag{8.4}$$

It leads to the following parametric suppressions of the Yukawa couplings:

$$Y^u \sim \begin{pmatrix} \epsilon^4 & \epsilon^3 & \epsilon^2 \\ \epsilon^3 & \epsilon^2 & \epsilon \\ \epsilon^2 & \epsilon & 1 \end{pmatrix}, \quad Y^d \sim (Y^e)^T \sim \begin{pmatrix} \epsilon^2 & \epsilon^2 & \epsilon^2 \\ \epsilon & \epsilon & \epsilon \\ 1 & 1 & 1 \end{pmatrix}. \tag{8.5}$$

We emphasize that for each entry we give the parametric suppression (that is the power of ϵ), but each entry has an unknown (complex) coefficient of order one, and there are no relations between the order one coefficients of different entries.

The structure of the Yukawa matrices dictates the parametric suppression of the physical observables:

$$\begin{aligned} y_t &\sim 1, \quad y_c \sim \epsilon^2, \quad y_u \sim \epsilon^4, \\ y_b &\sim 1, \quad y_s \sim \epsilon, \quad y_d \sim \epsilon^2, \\ y_\tau &\sim 1, \quad y_\mu \sim \epsilon, \quad y_e \sim \epsilon^2, \\ |V_{us}| &\sim \epsilon, \quad |V_{cb}| \sim \epsilon, \quad |V_{ub}| \sim \epsilon^2, \quad \delta_{\rm KM} \sim 1. \end{aligned} \tag{8.6}$$

For $\epsilon \sim 0.05$, the parametric suppressions are roughly consistent with the observed hierarchy. In particular, this set of charges predicts that the down and charged lepton mass hierarchies are similar, while the up hierarchy is the square of the down hierarchy. These features are roughly realized in Nature.

Exercise 8: *Derive the parametric suppression and approximate numerical values of Y^u, its eigenvalues, and the three angles of V_L^u, for $H(Q_i) = 4, 2, 0$, $H(U_i) = 3, 2, 0$ and $\epsilon_H = 0.2$.*

Could we explain any set of observed values with such an approximate symmetry? If we could, then the FN mechanism cannot be really tested. The answer however is negative. Consider, for example, the quark sector. Naively, we have 11 $U(1)_H$ charges that we are free to choose. However, the $U(1)_Y \times U(1)_B \times U(1)_{\rm PQ}$ symmetry implies that there are only 8 independent choices that affect the structure of the Yukawa couplings. On the other hand, there are 9 physical parameters. Thus, there should be a single relation between the physical parameters that is independent of the choice of charges. Assuming that the sum of charges in the exponents of Eq. (8.3) is of the same sign for all 18 combinations, the relation is

$$|V_{ub}| \sim |V_{us}V_{cb}|, \tag{8.7}$$

which is fulfilled to within a factor of 2. There are also interesting inequalities (here $i < j$):

$$|V_{ij}| \gtrsim m(U_i)/m(U_j), \; m(D_i)/m(D_j). \tag{8.8}$$

All six inequalities are fulfilled. Finally, if we order the up and the down masses from light to heavy, then the CKM matrix is predicted to be $\sim \mathbf{1}$, namely the diagonal entries are not parametrically suppressed. This structure is also consistent with the observed CKM structure.

8.2 The flavor of neutrinos

Five neutrino flavor parameters have been measured in recent years (see e.g. Ref. [40]): two mass-squared differences,

$$\Delta m^2_{21} = (7.4 \pm 0.2) \times 10^{-5}\ {\rm eV}^2, \quad |\Delta m^2_{32}| = (2.51 \pm 0.03) \times 10^{-3}\ {\rm eV}^2, \tag{8.9}$$

and the three mixing angles,

$$\sin^2\theta_{12} = 0.310 \pm 0.013, \;\; \sin^2\theta_{23} = 0.56 \pm 0.03, \;\; \sin^2\theta_{13} = 0.0224 \pm 0.0007. \tag{8.10}$$

These parameters constitute a significant addition to the thirteen SM flavor parameters and provide, in principle, tests of various ideas to explain the SM flavor puzzle.

The numerical values of the parameters show various surprising features:

- $|U_{\mu 3}| >$ any $|V_{ij}|$;
- $|U_{e2}| >$ any $|V_{ij}|$;
- $|U_{e3}|$ is not particularly small ($|U_{e3}| \not\ll |U_{e2}U_{\mu 3}|$);
- $m_2/m_3 \gtrsim 1/6 >$ any m_i/m_j for charged fermions.

These features can be summarized by the statement that, in contrast to the charged fermions, neither smallness nor hierarchy have been observed so far in the neutrino related parameters.

One way of interpretation of the neutrino data comes under the name of neutrino mass anarchy [41–43]. It postulates that the neutrino mass matrix has no special structure, namely all entries are of the same order of magnitude. Normalized to an effective neutrino mass scale, $v^2/\Lambda_{\rm seesaw}$, the various entries are random numbers of order one. Note that anarchy means neither hierarchy nor degeneracy.

If true, the contrast between neutrino mass anarchy and quark and charged lepton mass hierarchy may be a deep hint for a difference between the flavor physics of Majorana and Dirac fermions. The source of both anarchy and hierarchy might, however, be explained by a much more mundane mechanism. In particular, neutrino mass anarchy could be a result of a FN mechanism, where the three left-handed lepton doublets carry the same FN charge. In that case, the FN mechanism predicts parametric suppression of neither neutrino mass ratios nor leptonic mixing angles, which is quite consistent with (8.9) and (8.10). Indeed, the viable FN model presented in Section 8.1 belongs to this class.

Another possible interpretation of the neutrino data is to take $m_2/m_3 \sim |U_{e3}| \sim 0.15$ to be small, and require that they are parametrically suppressed (while the other two mixing angles are order one). Such a situation is impossible to accommodate in a large class of FN models [44].

The same data, and in particular the proximity of $(|U_{\mu 3}|, |U_{\tau 3}|)$ to $(1/\sqrt{2}, 1/\sqrt{2})$, and the proximity of $|U_{e2}|$ to $1/\sqrt{3} \simeq 0.58$, led to a very different interpretation. This interpretation, termed 'tribimaximal mixing' (TBM), postulates that the leptonic mixing matrix is parametrically close to the following special form [45]:

$$|U|_{\rm TBM} = \begin{pmatrix} \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{pmatrix}. \tag{8.11}$$

Such a form is suggestive of discrete non-Abelian symmetries, and indeed numerous models based on an A_4 symmetry have been proposed [46, 47]. A significant feature of of TBM is that the third mixing angle should be close to $|U_{e3}| = 0$. Until 2012, there have been only upper bounds on $|U_{e3}|$, consistent with the models in the literature. In recent years, however, a value of $|U_{e3}|$ close to the previous upper bound has been established [48], see Eq. (8.10). Such a large value (and the consequent significant deviation of $|U_{\mu 3}|$ from maximal bimixing) puts in serious doubt the TBM idea. Indeed, it is difficult in this framework, if not impossible, to account for $\Delta m^2_{12}/\Delta m^2_{23} \sim |U_{e3}|^2$ without fine-tuning [49].

9 Higgs physics: the new flavor arena

The SM relates the Yukawa couplings to the corresponding mass matrices:

$$Y^f = \sqrt{2}M_f/v. \tag{9.1}$$

Examining the Yukawa couplings in the mass basis, this simple equation implies four features:

1. *Proportionality:* $y_i \equiv Y^f_{ii} \propto m_i$;
2. *Factor* of proportionality: $y_i/m_i = \sqrt{2}/v$;
3. *Diagonality:* $Y^f_{ij} = 0$ for $i \neq j$.
4. *CP*: $\mathcal{I}m(y_i/m_i) = 0$.

In extensions of the SM, each of these four features might be violated. Thus, testing these features might provide a window to new physics and to allow progress in understanding the flavor puzzles.

The Higgs boson h was discovered by the ATLAS and CMS experiments at the LHC [50, 51]. The experiments normalize their results for Higgs production and decays to the SM rates:

$$\mu_f \equiv \frac{\sigma(pp \to h)\mathrm{BR}(h \to f)}{[\sigma(pp \to h)\mathrm{BR}(h \to f)]^{\mathrm{SM}}}. \tag{9.2}$$

The measurements give:

$$\begin{aligned}
\mu_{\gamma\gamma} &= 1.10 \pm 0.07, \\
\mu_{ZZ^*} &= 1.02 \pm 0.08, \\
\mu_{WW^*} &= 1.00 \pm 0.08, \\
\mu_{b\bar{b}} &= 0.99 \pm 0.12, \\
\mu_{\tau\tau} &= 0.91 \pm 0.09, \\
\mu_{\mu\mu} &= 1.21 \pm 0.33,
\end{aligned} \tag{9.3}$$

and the bounds [52–54]

$$\begin{aligned}
\mu_{c\bar{c}} &\in [1.2, 26], \\
\mu_{c\bar{c}} &\leq 20\mu_{b\bar{b}}, \\
\mathrm{BR}_{ee} &< 3.6 \times 10^{-4}.
\end{aligned} \tag{9.4}$$

Given that $m_b/m_c \simeq 4.58$ [55], and that the upper bound on $\mu_{c\bar{c}}/\mu_{b\bar{b}}$ implies that $\kappa_c/\kappa_b < 4.5$, it is now experimentally established that $y_c < y_b$. Given $\mathrm{BR}^{\mathrm{SM}}_{ee} = 5 \times 10^{-9}$, the latter translates into $\mu_{ee} < 7.2 \times 10^4$.

As concerns quark flavor changing Higgs couplings, these have been searched for in $t \to qh$ decays ($q = c, u$) [56, 57]:

$$\mathrm{BR}(t \to ch) < 7.3 \times 10^{-4},$$

$$\mathrm{BR}(t \to uh) \quad < \quad 1.9 \times 10^{-4}. \tag{9.5}$$

As concerns lepton flavor violating Higgs decays, the current bounds are

$$\begin{aligned}\mathrm{BR}(h \to \tau\mu) &< 1.5 \times 10^{-3}, \\ \mathrm{BR}(h \to \tau e) &< 2.2 \times 10^{-3}, \\ \mathrm{BR}(h \to \mu e) &< 6.1 \times 10^{-5}.\end{aligned} \tag{9.6}$$

CPV has been searched for in the Higgs couplings to $t\bar{t}$ and to $\tau^+\tau^-$, yielding uppr bounds on the relative CP-odd fraction [58–60]:

$$\begin{aligned}\sin\theta_{htt} &= 0.00 \pm 0.33, \\ \sin\theta_{h\tau\tau} &= -0.02 \pm 0.32.\end{aligned} \tag{9.7}$$

The measurements quoted in Eq. (9.3) can be presented in the $y_i - m_i$ plane. We do so in Fig. 8. The first two features quoted above are already being tested. The upper bounds on flavor violating decays quoted in Eqs. (9.5) and (9.6) test the third feature. The allowed ranges in Eq. (9.7) test the fourth feature. We can make the following statements:

- $y_e \lesssim y_\mu < y_\tau$. This goes in the direction of proportionality.
- The third generation Yukawa couplings, y_t, y_b, y_τ, as well as the second generation y_μ, obey $y_i/m_i \approx \sqrt{2}/v$. This is in agreement with the predicted factor of proportionality.
- There are strong upper bounds on violation of diagonality: $Y_{tc} \lesssim 0.02$ and $Y_{\tau\mu} \lesssim 0.002$.
- There are upper bounds on CPV in y_t/m_t and y_τ/m_τ.

Beyond the search for new physics via Higgs decays, it is interesting to ask whether the measurements of the Higgs couplings to quarks and leptons can shed light on the Standard Model and/or new physics flavor puzzles. If eventually the values of y_b and/or y_τ deviate from their SM values, the most likely explanation of such deviations will be that there are more than one Higgs doublets, and that the doublet(s) that couple to the down and charged lepton sectors are not the same as the one that couples to the up sector. A more significant test of our understanding of flavor physics comes from the double ratio

$$X_{\mu^+\mu^-} \equiv \frac{\mathrm{BR}(h \to \mu^+\mu^-)/\mathrm{BR}(h \to \tau^+\tau^-)}{m_\mu^2/m_\tau^2}, \tag{9.8}$$

which is predicted within the SM with impressive theoretical cleanliness. To leading order, it is given by 1, and the corrections of order α_W and of order m_μ^2/m_τ^2 to this leading result are known, and reduce the value to 0.98. The current experimental value is given by

$$X_{\mu^+\mu^-} = 1.03 \pm 0.31, \tag{9.9}$$

consistent with the SM prediction (as well as with the predictions of 2HDMs with NFC, the MSSM and MFV models), and excluding the possibility that Y_μ and Y_τ arise from terms of different dimensions in the SMEFT [61]. It is also interesting to test diagonality via the search for the SM-forbidden decay

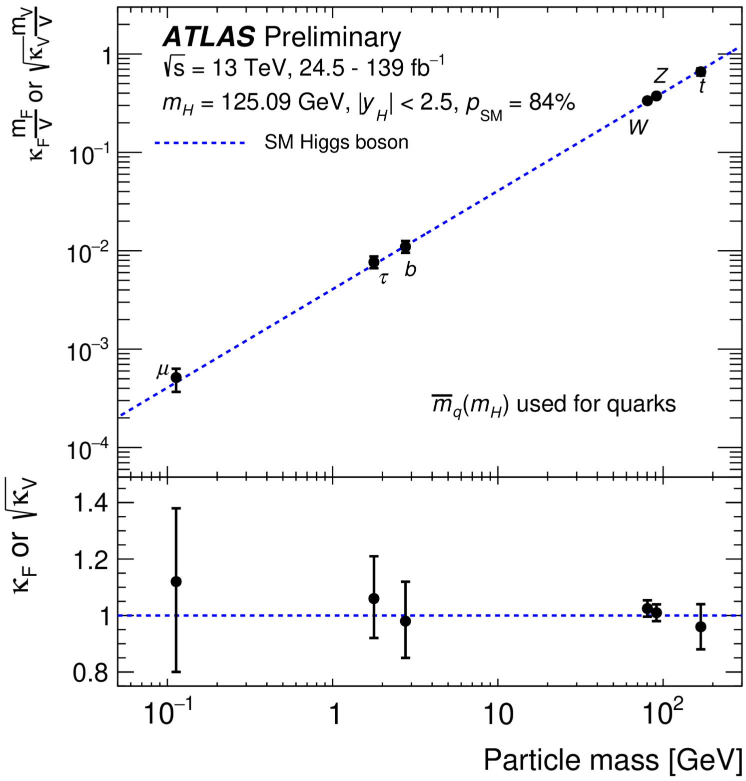


Fig. 8: The allowed ranges for the Higgs couplings. The SM prediction is presented by the dashed line.

modes, $h \to \mu^\pm\tau^\mp$. A measurement of, or an upper bound on

$$X_{\mu\tau} \equiv \frac{\mathrm{BR}(h \to \mu^+\tau^-) + \mathrm{BR}(h \to \mu^-\tau^+)}{\mathrm{BR}(h \to \tau^+\tau^-)}, \tag{9.10}$$

would provide additional information relevant to flavor physics. The current experimental value is given by

$$X_{\mu\tau} < 0.04. \tag{9.11}$$

We demonstrate below the potential power of Higgs flavor physics to lead to progress in our understanding of the flavor puzzles by focusing on the measurements of $\mu_{\tau^+\tau^-}$, $X_{\mu^+\mu^-}$ and $X_{\mu\tau}$ [61].

Let us take as an example how we can use the set of these three measurements if there is a single light Higgs boson. A violation of the SM relation $Y_{ij}^{\rm SM} = \frac{\sqrt{2}m_i}{v}\delta_{ij}$, is a consequence of non renormalizable terms. The leading ones are the $d=6$ terms. In the interaction basis, we have

$$\begin{aligned} \mathcal{L}_Y^{d=4} &= -\lambda_{ij}\overline{f_L^i}f_R^j\phi + \text{h.c.}, \\ \mathcal{L}_Y^{d=6} &= -\frac{\lambda'_{ij}}{\Lambda^2}\overline{f_L^i}f_R^j\phi(\phi^\dagger\phi) + \text{h.c.}\,, \end{aligned} \tag{9.12}$$

where expanding around the vacuum we have $\phi = (v+h)/\sqrt{2}$. Defining $V_{L,R}$ via

$$\sqrt{2}m = V_L \left(\lambda + \frac{v^2}{2\Lambda^2}\lambda'\right) V_R^\dagger v, \tag{9.13}$$

where $m = \text{diag}(m_e, m_\mu, m_\tau)$, and defining $\hat{\lambda}$ via

$$\hat{\lambda} = V_L \lambda' V_R^\dagger, \tag{9.14}$$

we obtain

$$Y_{ij} = \frac{\sqrt{2}m_i}{v}\delta_{ij} + \frac{v^2}{\Lambda^2}\hat{\lambda}_{ij}. \tag{9.15}$$

To proceed, one has to make assumptions about the structure of $\hat{\lambda}$. In what follows, we consider first the assumption of minimal flavor violation (MFV) and then a Froggatt–Nielsen (FN) symmetry.

Exercise 9: *Find the predictions of models with Natural Flavor Conservation (NFC) for* $\mu_{\tau^+\tau^-}$, $X_{\mu^+\mu^-}$ *and* $X_{\tau\mu}$.

9.1 MFV

MFV requires that the leptonic part of the Lagrangian is invariant under an $SU(3)_L \times SU(3)_E$ global symmetry, with the left-handed lepton doublets transforming as $(3,1)$, the right-handed charged lepton singlets transforming as $(1,3)$ and the charged lepton Yukawa matrix Y is a spurion transforming as $(3,\bar{3})$.

Specifically, MFV means that, in Eq. (9.12),

$$\lambda' = a\lambda + b\lambda\lambda^\dagger\lambda + \mathcal{O}(\lambda^5), \tag{9.16}$$

where a and b are numbers. Note that, if V_L and V_R are the diagonalizing matrices for λ, $V_L\lambda V_R^\dagger = \lambda^{\text{diag}}$, then they are also the diagonalizing matrices for $\lambda\lambda^\dagger\lambda$: $V_L\lambda\lambda^\dagger\lambda V_R^\dagger = (\lambda^{\text{diag}})^3$. Then, Eqs. (9.13), (9.14), and (9.15) become

$$\begin{aligned}
\frac{\sqrt{2}m}{v} &= \left(1 + \frac{av^2}{2\Lambda^2}\right)\lambda^{\text{diag}} + \frac{bv^2}{2\Lambda^2}(\lambda^{\text{diag}})^3, \\
\hat{\lambda} &= a\lambda^{\text{diag}} + b(\lambda^{\text{diag}})^3 = a\frac{\sqrt{2}m}{v} + \frac{2\sqrt{2}bm^3}{v^3}, \\
Y_{ij} &= \frac{\sqrt{2}m_i}{v}\delta_{ij}\left[1 + \frac{av^2}{\Lambda^2} + \frac{2bm_i^2}{\Lambda^2}\right],
\end{aligned} \tag{9.17}$$

where, in the expressions for $\hat{\lambda}$ and Y, we included only the leading universal and leading non-universal corrections to the SM relations.

We learn the following points about the Higgs-related lepton flavor parameters in this class of models:

1. h has no flavor off-diagonal couplings:

$$Y_{\mu\tau}, Y_{\tau\mu} = 0. \tag{9.18}$$

2. The values of the diagonal couplings deviate from their SM values. The deviation is small, of order v^2/Λ^2:

$$y_\tau \approx \left(1 + \frac{av^2}{\Lambda^2}\right) \frac{\sqrt{2}m_\tau}{v}. \tag{9.19}$$

3. The ratio between the Yukawa couplings to different charged lepton flavors deviates from its SM value. The deviation is, however, very small, of order m_ℓ^2/Λ^2:

$$\frac{y_\mu}{y_\tau} = \frac{m_\mu}{m_\tau}\left(1 - \frac{2b(m_\tau^2 - m_\mu^2)}{\Lambda^2}\right). \tag{9.20}$$

The predictions of the SM with MFV non-renormalizable terms are then the following:

$$\begin{aligned} \mu_{\tau^+\tau^-} &= 1 + 2av^2/\Lambda^2, \\ X_{\mu^+\mu^-} &= 1 - 4bm_\tau^2/\Lambda^2, \\ X_{\tau\mu} &= 0. \end{aligned} \tag{9.21}$$

Thus, MFV will be excluded if experiments observe the $h \to \mu\tau$ decay. On the other hand, MFV allows for a universal deviation of $\mathcal{O}(v^2/\Lambda^2)$ of the flavor-diagonal dilepton rates, and a smaller non-universal deviation of $\mathcal{O}(m_\tau^2/\Lambda^2)$.

9.2 FN

An attractive explanation of the smallness and hierarchy in the Yukawa couplings is provided by the Froggatt–Nielsen (FN) mechanism [35]. In this framework, a $U(1)_H$ symmetry, under which different generations carry different charges, is broken by a small parameter ϵ_H. Without loss of generality, ϵ_H is taken to be a spurion of charge -1. Then, various entries in the Yukawa mass matrices are suppressed by different powers of ϵ_H, leading to smallness and hierarchy.

Specifically for the leptonic Yukawa matrix, taking the Higgs field to be neutral under $U(1)_H$, $H(\phi) = 0$, we have

$$\lambda_{ij} \propto \epsilon_H^{H(E_j)-H(L_i)}. \tag{9.22}$$

We emphasize that the FN mechanism dictates only the parametric suppression. Each entry has an arbitrary order-one coefficient. The resulting parametric suppression of the masses and leptonic mixing angles is given by [62]

$$m_{\ell_i}/v \sim \epsilon_H^{H(E_i)-H(L_i)}, \quad |U_{ij}| \sim \epsilon_H^{H(L_j)-H(L_i)}. \tag{9.23}$$

Since $H(\phi^\dagger\phi) = 0$, the entries of the matrix λ' have the same parametric suppression as the corresponding entries in λ [63], though the order-one coefficients are different:

$$\lambda'_{ij} = \mathcal{O}(1) \times \lambda_{ij}. \tag{9.24}$$

This structure allows us to estimate the entries of $\hat{\lambda}_{ij}$ in terms of physical observables:

$$\begin{aligned}
\hat{\lambda}_{33} &\sim m_\tau/v, \\
\hat{\lambda}_{22} &\sim m_\mu/v, \\
\hat{\lambda}_{23} &\sim |U_{23}|(m_\tau/v), \\
\hat{\lambda}_{32} &\sim (m_\mu/v)/|U_{23}|.
\end{aligned} \tag{9.25}$$

We learn the following points about the Higgs-related lepton flavor parameters in this class of models:

1. h has flavor off-diagonal couplings:

$$\begin{aligned}
Y_{\mu\tau} &= \mathcal{O}\left(\frac{|U_{23}|vm_\tau}{\Lambda^2}\right), \\
Y_{\tau\mu} &= \mathcal{O}\left(\frac{vm_\mu}{|U_{23}|\Lambda^2}\right).
\end{aligned} \tag{9.26}$$

2. The values of the diagonal couplings deviate from their SM values:

$$y_\tau \approx \frac{\sqrt{2}m_\tau}{v}\left[1 + \mathcal{O}\left(\frac{v^2}{\Lambda^2}\right)\right]. \tag{9.27}$$

3. The ratio between the Yukawa couplings to different charged lepton flavors deviates from its SM value:

$$\frac{y_\mu}{y_\tau} = \frac{m_\mu}{m_\tau}\left[1 + \mathcal{O}\left(\frac{v^2}{\Lambda^2}\right)\right]. \tag{9.28}$$

The predictions of the SM with FN-suppressed non-renormalizable terms are then the following:

$$\begin{aligned}
\mu_{\tau^+\tau^-} &= 1 + \mathcal{O}(v^2/\Lambda^2), \\
X_{\mu^+\mu^-} &= 1 + \mathcal{O}(v^2/\Lambda^2), \\
X_{\tau\mu} &= \mathcal{O}(v^4/\Lambda^4).
\end{aligned} \tag{9.29}$$

Thus, FN will be excluded if experiments observe deviations from the SM of the same size in both flavor-diagonal and flavor-changing h decays. On the other hand, FN allows non-universal deviations of $\mathcal{O}(v^2/\Lambda^2)$ in the flavor-diagonal dilepton rates, and a smaller deviation of $\mathcal{O}(v^4/\Lambda^4)$ in the off-diagonal rate.

10 New physics?

In this section we discuss two sets of recent measurements of flavor changing processes that arouse much interest: lepton flavor universality in semileptonic B decays, $R(D^{(*)})$, and direct CP violation in

D decays, ΔA_{CP}.

10.1 $B \to D^{(*)}\tau\nu$

Within the Standard Model (SM), the electroweak interactions of the leptons are flavor universal. Violation of lepton flavor universality arises from Yukawa interactions, that are negligible in this context, and from phase space effects, which are calculable. A test of the SM prediction of lepton flavor universality between the τ-lepton and the light ℓ-leptons ($\ell = e, \mu$) is provided by the ratios

$$R(D^{(*)}) \equiv \frac{\Gamma(B \to D^{(*)}\tau\bar{\nu})}{\Gamma(B \to D^{(*)}\ell\bar{\nu})}, \quad (\ell = e \text{ or } \mu). \tag{10.1}$$

The SM predictions, derived by naive averaging [22] over the results reported in Refs. [64–67], are

$$\begin{aligned} R(D) &= 0.299 \pm 0.003, \\ R(D^*) &= 0.258 \pm 0.005. \end{aligned} \tag{10.2}$$

The current world averages for $R(D)$ and $R(D^*)$, combining the results reported in Refs. [68–76] are as follows [22]:

$$\begin{aligned} R(D) &= 0.340 \pm 0.027 \pm 0.013, \\ R(D^*) &= 0.295 \pm 0.011 \pm 0.008. \end{aligned} \tag{10.3}$$

The difference of the experimental measurements from the SM predictions corresponds to about 3.1σ (p-value of 2.7×10^{-3}). We thus aim to explain

$$R(D^{(*)})/R(D^{(*)})^{\rm SM} \approx 1.14 \pm 0.05. \tag{10.4}$$

In this section we entertain the idea that a deviation from the SM will indeed be established. We describe the analysis of Ref. [77]. The quark transition via which the $B \to D^{(*)}\tau\nu$ proceeds is $b \to c\tau\nu$. Note, however, that the flavor of the neutrino is, of course, unobservable. It could be ν_τ, in which case the process respects the accidental lepton flavor symmetry of the SM. There is no reason, however, that the symmetry is respected by new physics, particularly when the new physics violates lepton flavor universality, so that the neutrino could also be ν_μ or ν_e or some combination of the three flavors [77, 78]. Let us ask two questions:

- Could the $R(D^{(*)})$ puzzle be solved via new physics contributions to $b \to c\tau\nu_{e,\mu}$?
- If not, how precise should the alignment of ν with ν_τ be?

We assume that the new physics contributions originate at a scale $\Lambda \gg v$, and consider the following two terms in the SMEFT Lagrangian [78]:

$$\mathcal{L}_{\rm NP} = \frac{C_1^{ilkm}}{\Lambda^2}(\overline{L_i}\gamma_\sigma L_l)(\overline{Q_k}\gamma^\sigma Q_m) + \frac{C_3^{ilkm}}{\Lambda^2}(\overline{L_i}\gamma_\sigma \tau^a L_l)(\overline{Q_k}\gamma^\sigma \tau^a Q_m), \tag{10.5}$$

where L is the $SU(2)$-doublet lepton field, Q is the $SU(2)$-doublet quark field, and i, l, k, m are flavor indices. For the sake of definiteness, and to avoid the strongest constraints from flavor changing neutral current (FCNC) processes, we take $i = \tau$, $k = s$, and $m = b$, while l runs over e, μ, τ. We denote $C_{1,3}^{\tau lsb}$ by $C_{1,3}^{l}$. The $C_{1,3}^{l}$-dependent terms can be rewritten as follows:

$$\begin{aligned}
\Lambda^2 \mathcal{L}_{\rm NP} &= (C_1^l + C_3^l) V_{is} V_{jb}^* (\overline{u_{Li}} \gamma^\mu u_{Lj})(\overline{\nu_\tau} \gamma_\mu \nu_l) + (C_1^l + C_3^l)(\overline{s_L} \gamma^\mu b_L)(\overline{\tau_L} \gamma_\mu l_L) \\
&+ (C_1^l - C_3^l) V_{is} V_{jb}^* (\overline{u_{Li}} \gamma^\mu u_{Lj})(\overline{\tau_L} \gamma_\mu l_L) + (C_1^l - C_3^l)(\overline{s_L} \gamma^\mu b_L)(\overline{\nu_\tau} \gamma_\mu \nu_l) \\
&+ 2C_3^l V_{is} (\overline{u_{Li}} \gamma^\mu b_L)(\overline{\tau_L} \gamma_\mu \nu_l) + 2C_3^l V_{jb} (\overline{u_{Lj}} \gamma^\mu s_L)(\overline{\tau_L} \gamma_\mu \nu_l) + \text{h.c.}. \qquad (10.6)
\end{aligned}$$

Thus, the SMEFT Lagrangian terms that contribute to $b \to c\tau\nu$ are

$$\mathcal{L} = \left(\frac{4G_F V_{cb} \delta_{l\tau}}{\sqrt{2}} + \frac{2C_3^l V_{cs}}{\Lambda^2} \right) (\overline{c_L} \gamma^\mu b_L)(\overline{\tau_L} \gamma_\mu \nu_l). \qquad (10.7)$$

We obtain:

$$\frac{R(D^{(*)})}{R(D^{(*)})^{\rm SM}} = 1 + \frac{\sqrt{2}}{G_F} \mathcal{R}e \left(\frac{V_{cs}}{V_{cb}} \frac{C_3^\tau}{\Lambda^2} \right) + \frac{\sum_{\ell=e,\mu} |C_3^\ell|^2}{2G_F^2 \Lambda^4} \left| \frac{V_{cs}}{V_{cb}} \right|^2, \qquad (10.8)$$

where we assume that the contribution of the term quadratic in C_3^τ is negligible compared to the term linear in C_3^τ.

Thus, to account for the $R(D^{(*)})$ puzzle by purely $b \to c\tau\nu_\ell$, $\ell = e, \mu$, we need

$$\left(\frac{\sum_{\ell=e,\mu} |C_3^\ell|^2}{\Lambda^4} \right)^{1/2} = (0.24 \pm 0.04)\ \text{TeV}^{-2} = \frac{1}{[(2.0 \pm 0.2)\ \text{TeV}]^2}. \qquad (10.9)$$

On the other hand, to account for the $R(D^{(*)})$ puzzle by purely $b \to c\tau\nu_\tau$, we need

$$\frac{C_3^\tau}{\Lambda^2} = (0.046 \pm 0.016)\ \text{TeV}^{-2} \approx \frac{1}{[(4.7 \pm 0.8)\ \text{TeV}]^2}. \qquad (10.10)$$

If the $R(D^{(*)})$ puzzle is accounted for by purely $b \to c\tau\bar{\nu}_\ell$, Eq. (10.9) implies that we need $|C_3^\ell|/\Lambda^2 \sim 1/(2\ \text{TeV})^2$. Eq. (10.6) implies that the C_3^ℓ term contributes, via four Fermi operators with the flavor structures $\bar{s}b\bar{\tau}\ell$ and $b\bar{s}\bar{\nu}_\tau\nu_\ell$, to various flavor changing neutral current and lepton flavor violating processes which are forbidden in the SM. The strongest constraints are the following:

– The experimental upper bound on $\text{BR}(B^+ \to K^+\tau^+\mu^-)$ [79] implies

$$\frac{|C_1^\mu + C_3^\mu|}{\Lambda^2} < 0.058\ \text{TeV}^{-2}. \qquad (10.11)$$

– The experimental upper bound on $\text{BR}(B^+ \to K^+e^-\tau^+)$ [79] implies

$$\frac{|C_1^e + C_3^e|}{\Lambda^2} < 0.044\ \text{TeV}^{-2}. \qquad (10.12)$$

– The experimental upper bound on $\text{BR}(B^+ \to K^+\nu\bar{\nu})$ [80, 81] implies

$$\frac{|C_1^\ell - C_3^\ell|}{\Lambda^2} < 0.031\ \text{TeV}^{-2}. \qquad (10.13)$$

The effective operators of Eq. (10.5) will contribute to the scattering process $pp \to \tau^\pm \mu^\mp X_h$, where X_h stands for final hadrons. At present, however, these bounds are not competitive with the ones extracted from the LFV B decays.

From the upper bounds on $|C_1^\ell \pm C_3^\ell|$ we can obtain upper bounds on C_3^ℓ alone:

$$\frac{|C_3^\mu|}{\Lambda^2} < 0.044 \text{ TeV}^{-2}, \quad \frac{|C_3^e|}{\Lambda^2} < 0.037 \text{ TeV}^{-2}. \tag{10.14}$$

We reach the following conclusions:

– Given that, to account for the central value of $R(D^{(*)})$, it is required that $|C_3^\ell|/\Lambda^2 \simeq 0.24 \text{ TeV}^{-2}$, but other constraints require that $|C_3^\mu|/\Lambda^2 < 0.044 \text{ TeV}^{-2}$, the contribution of $b \to c\tau\nu_\ell$, with $\ell = e, \mu$, to $R(D^{(*)})/R(D^{(*)})^{\rm SM} - 1$ cannot exceed about 4% of the required shift.

– Given that, to account for the central value of $R(D^{(*)})$, it is required that $|C_3^\tau|/\Lambda^2 \simeq 0.046 \text{ TeV}^{-2}$, but phenomenological constraints require that $|C_3^\mu|/\Lambda^2 < 0.044 \text{ TeV}^{-2}$, and $|C_3^e|/\Lambda^2 < 0.037 \text{ TeV}^{-2}$, we learn that no special alignment with the τ-direction is needed to explain the $R(D^{(*)})$ puzzle.

– Conversely, if operators of the form

$$\frac{C_3^l}{\Lambda^2}(\overline{L_\tau}\gamma_\sigma\tau^a L_l)(\overline{Q_s}\gamma^\sigma\tau^a Q_b) \tag{10.15}$$

have C_3^τ, C_3^μ and C_3^e all of the same order of magnitude, $C_3^l/\Lambda^2 \sim 0.04 \text{ TeV}^{-2}$, then the shift in $R(D^{(*)})$ will be dominated by a factor of order 30 by C_3^τ, and all phenomenological constraints satisfied.

10.2 Direct CP violation in charm decays

Direct CP violation can be measured in charm decays to final CP eigenstates [82] via

$$\Delta A_{CP} = A_{CP}(K^+K^-) - A_{CP}(\pi^+\pi^-), \tag{10.16}$$

where

$$A_{CP}(f) = \frac{\Gamma(D^0 \to f) - \Gamma(\overline{D}^0 \to f)}{\Gamma(D^0 \to f) + \Gamma(\overline{D}^0 \to f)}. \tag{10.17}$$

The LHCb collaboraion measured [83, 84]

$$\begin{aligned} \Delta A_{CP} &= (-1.54 \pm 0.29) \times 10^{-3}, \\ A_{CP}(K^+K^-) &= (+0.77 \pm 0.57) \times 10^{-3}, \\ A_{CP}(\pi^+\pi^-) &= (+2.32 \pm 0.61) \times 10^{-3}. \end{aligned} \tag{10.18}$$

The CP asymmetry arises from interference between a strong penguin and tree diagrams. It is thus loop suppressed, and CKM suppressed by a factor of

$$2\mathcal{I}m\left(\frac{V_{ub}V_{cb}^*}{V_{us}V_{cs}^*}\right) \approx 1.4 \times 10^{-3}. \tag{10.19}$$

Within the SM, and assuming the U-spin symmetry of QCD, one can then estimate the size of the asymmetry:

$$\begin{aligned} \Delta A_{CP}^{\rm SM} &\approx -2.8 \times 10^{-3} \times (\alpha_s/\pi) r_{QCD}, \\ A_{CP}(K^+K^-) &= -A_{CP}(\pi^+\pi^-). \end{aligned} \tag{10.20}$$

Thus, to accommodate the experimental results within the SM, two surprising features should arise in the relevant strong interactions:

– The ratio of penguin to tree should be enhanced by a surprisingly large factor, $r_{QCD} \sim 10$ [85].
– U-spin should be strongly broken. with $A_{USV}/A_{USC} \sim 1.7$ [86].

If these features are not realized in nature, the measured values call for new physics (see e.g. Ref. [87, 88]). Relevant models include the 2HDM, the MSSM, vector-like up quarks and Z' models. Within the SMEFT, the scale of the CP violating new physics is bounded:

$$\Lambda_{\rm NP} \lesssim 40 \text{ TeV}. \tag{10.21}$$

11 Conclusions

(i) The symmetry principles that define the Standard Model have a very strong predictive power concerning flavor physics. They predict that the photon-, gluon- and Z-mediated interactions are flavor universal, that the W-mediated interactions in the lepton sector are flavor universal, and in the quark sector depend on a unitary matrix, and that the Higgs mediated interactions are flavor diagonal.

(ii) Experimental results are consistent with all of these predictions, except for the lepton flavor universality of the leptonic W interactions. The observed lepton flavor transitions established that the neutrinos are massive.

(iii) Measurements of CP violating B-meson decays have established that the Kobayashi–Maskawa mechanism is the dominant source of the observed CP violation.

(iv) Measurements of flavor changing B-meson decays have established the the Cabibbo–Kobayashi–Maskawa mechanism is the dominant source of the observed quark flavor violation.

(v) The consistency of all these measurements with the CKM predictions sharpens the new physics flavor puzzle: If there is new physics at, or below, the TeV scale, then its flavor structure must be highly non-generic.

(vi) Measurements of neutrino flavor parameters have not only not clarified the Standard Model flavor puzzle, but actually deepened it. Whether they imply an anarchical structure, or a tribimaximal mixing, it seems that the neutrino flavor structure is very different from that of quarks.

(vii) If the LHC experiments, ATLAS and CMS, discover new particles that couple to the Standard

Model fermions, then, in principle, they will be able to measure new flavor parameters. Consequently, the new physics flavor puzzle is likely to be understood.

(viii) If the flavor structure of such new particles is affected by the same physics that sets the flavor structure of the Yukawa couplings, then the LHC experiments (and future flavor factories) may be able to shed light also on the Standard Model flavor puzzle.

(ix) The Higgs program provides an opportunity to make progress in our understanding of the flavor puzzle(s).

(x) Extensions of the SM where new particles couple to quark- and/or lepton-pairs are constrained by flavor.

(xi) There are experimental hints that lepton flavor universality is violated in B decays. These hints will be further tested in the coming years.

The huge progress in flavor physics in recent years has provided answers to many questions. At the same time, new questions arise. The LHC experiments, Belle-II and neutrino experiments are likely to provide more answers and more questions.

Acknowledgments

YN is the Amos de-Shalit chair of theoretical physics, and is supported by grants from the Israel Science Foundation (grant number 1124/20), the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel (grant number 2018257), by the Minerva Foundation (with funding from the Federal Ministry for Education and Research), and by the Yeda-Sela (YeS) Center for Basic Research.

Appendices

A SM calculations of the mixing amplitude

We present the SM calculation of the mixing amplitude $M_{B\overline{B}}$ and its generalization to the other meson systems. The leading diagrams that contribute to $M_{B\overline{B}}$ are one loop diagrams that are called "box diagrams" and are displayed in Fig. 4. We can write the transition amplitude as

$$\mathcal{A}_{B^0\to\overline{B}^0} = C_{\rm SM}(\bar{d}_L\gamma_\mu b_L)(\bar{d}_L\gamma^\mu b_L). \tag{A.1}$$

The normalized matrix element is related to the amplitude via

$$M_{B\overline{B}} = \frac{1}{2m_B}\langle B^0|\mathcal{A}_{B^0\to\overline{B}^0}|\overline{B}^0\rangle, \tag{A.2}$$

and thus

$$M_{B\overline{B}} = \frac{C_{\rm SM}}{2m_B}\langle B^0|(\bar{d}_L\gamma_\mu b_L)(\bar{d}_L\gamma^\mu b_L)|\overline{B}^0\rangle. \tag{A.3}$$

The non-perturbative QCD effects are encoded in the hadronic matrix element, which we parameterize as follows:

$$\langle B^0|(\bar{d}_L\gamma_\mu b_L)(\bar{d}_L\gamma^\mu b_L)|\overline{B}^0\rangle = -\frac{1}{3}m_B^2 B_B f_B^2, \tag{A.4}$$

where B_B is a number and f_B is the B-meson decay constant. Lattice calculations give $\sqrt{B_B}f_B \approx 0.22$ GeV. This is where the hadronic uncertainties lie.

The weak interactions effects are encoded in $C_{\rm SM}$, which is calculated from the box diagrams:

$$C_{\rm SM} = \frac{G_F^2 m_W^2}{2\pi^2} \times \left[(V_{cb}V_{cd}^*)^2 S(x_c,x_c) + (V_{tb}V_{td}^*)^2 S(x_t,x_t) + (V_{cb}V_{cd}^*)(V_{tb}V_{td}^*)S(x_t,x_c)\right], \tag{A.5}$$

where $x_i = m_i^2/m_W^2$, we approximate $x_u = 0$, and S is the loop function:

$$\begin{aligned} S(x_i,x_j) \quad = \quad & x_i x_j \left[-\frac{3}{4(1-x_i)(1-x_j)} + \frac{\log x_i}{(x_i-x_j)(1-x_i)^2}\left(1-2x_i+\frac{x_i^2}{4}\right) + \right. \\ & \left. \frac{\log x_j}{(x_j-x_i)(1-x_j)^2}\left(1-2x_j+\frac{x_j^2}{4}\right)\right]. \end{aligned} \tag{A.6}$$

Note that $S(0,x) = 0$. Taking into account the values of the quark masses and CKM elements, we conclude that the term proportional to $S(x_t,x_t)$ dominates over those proportional to $S(x_t,x_c)$ and $S(x_c,x_c)$, and thus

$$M_{B\overline{B}} \approx \frac{G_F^2}{12\pi^2} m_B m_W^2 (B_B f_B^2)(V_{tb}V_{td}^*)^2 S(x_t,x_t). \tag{A.7}$$

This result is subject to known radiative correction that are of $O(1)$.

Eq. (A.7) can be straightforwardly generalized to other systems. For $M_{B_s\overline{B_s}}$, we replace $d \to s$:

$$M_{B_s\overline{B_s}} \approx \frac{G_F^2}{12\pi^2} m_{B_s} m_W^2 (B_{B_s} f_{B_s}^2)(V_{tb}V_{ts}^*)^2 S(x_t,x_t)\,. \tag{A.8}$$

Table A.1: The experimental values of the neutral meson mixing parameters. In all cases (including the K meson system) we define x and y as in Eqs. (B.12). For the K^0 system, the error on y is well below a permill and thus we do not include an error. For the B^0 system, there is only an upper bound on $|y|$.

P	m [GeV]	Γ [GeV]	x	y
K^0	0.498	3.68×10^{-15}	0.945 ± 0.001	-0.997
D^0	1.86	1.60×10^{-13}	0.0039 ± 0.0018	$+0.0065 \pm 0.0009$
B^0	5.28	4.33×10^{-13}	0.775 ± 0.006	-0.007 ± 0.009
B_s	5.37	4.34×10^{-13}	26.82 ± 0.23	-0.061 ± 0.008

The ratio $\Delta m_B/\Delta m_{B_s}$ is particularly interesting:

$$\frac{\Delta m_B}{\Delta m_{B_s}} = \frac{m_B B_B f_B^2}{m_{B_s} B_{B_s} f_{B_s}^2} \left|\frac{V_{td}}{V_{ts}}\right|^2 . \tag{A.9}$$

In the $SU(3)_F$ limit, the hadronic matrix elements of B and B_s are the same. Consequently, in the ratio of Eq. (A.9), the hadronic uncertainty is only in the correction to the $SU(3)_F$ limit, and is therefore small. Thus, the ratio $\Delta m_B/\Delta m_{B_s}$ provides an excellent measurement of $|V_{td}/V_{ts}|$.

For $M_{K\overline{K}}$, we replace $b \to s$:

$$M_{K\overline{K}} = \frac{G_F^2}{12\pi^2} m_K m_W^2 (B_K f_K^2) \left[(V_{cs}V_{cd}^*)^2 S(x_c,x_c) + (V_{ts}V_{td}^*)^2 S(x_t,x_t) + (V_{cs}V_{cd}^*)(V_{ts}V_{td}^*) S(x_t,x_c)\right] . \tag{A.10}$$

Lattice results gives $B_K = 0.86 \pm 0.24$.

For the four systems, $P = B, B_s, D, K$, the calculation of $M_{P\overline{P}}$ translates into the calculation of the mass splitting $\Delta M_P = 2|M_{P\overline{P}}|$ (in the D system, however, the calculation of $M_{D\overline{D}}$ is complicated, and we do not discuss it here).

The numerical values of the mixing parameters are presented in Table A.1. The SM calculations outlined above agree well with the data.

B Neutral meson oscillations

We define decay amplitudes of B (which could be charged or neutral) and its CP conjugate $\overline{B}$ to a multi-particle final state f and its CP conjugate $\overline{f}$ as

$$A_f = \langle f|\mathcal{H}|B\rangle \quad , \quad \overline{A}_f = \langle f|\mathcal{H}|\overline{B}\rangle \quad , \quad A_{\overline{f}} = \langle \overline{f}|\mathcal{H}|B\rangle \quad , \quad \overline{A}_{\overline{f}} = \langle \overline{f}|\mathcal{H}|\overline{B}\rangle , \tag{B.1}$$

where $\mathcal{H}$ is the Hamiltonian governing weak interactions. The action of CP on these states introduces phases ξ_B and ξ_f according to

$$\begin{aligned} CP\,|B\rangle &= e^{+i\xi_B}\,|\overline{B}\rangle \quad , \quad CP\,|f\rangle = e^{+i\xi_f}\,|\overline{f}\rangle , \\ CP\,|\overline{B}\rangle &= e^{-i\xi_B}\,|B\rangle \quad , \quad CP\,|\overline{f}\rangle = e^{-i\xi_f}\,|f\rangle , \end{aligned} \tag{B.2}$$

so that $(CP\ )^2 = 1$. The phases ξ_B and ξ_f are arbitrary and unphysical because of the flavor symmetry of the strong interaction. If CP is conserved by the dynamics, $[CP\ ,\mathcal{H}] = 0$, then A_f and $\overline{A}_{\overline{f}}$ have the same magnitude and an arbitrary unphysical relative phase

$$\overline{A}_{\overline{f}} = e^{i(\xi_f - \xi_B)}\, A_f\ . \tag{B.3}$$

A state that is initially a superposition of B^0 and $\overline{B}^0$, say

$$|\psi(0)\rangle = a(0)|B^0\rangle + b(0)|\overline{B}^0\rangle\ , \tag{B.4}$$

will evolve in time acquiring components that describe all possible decay final states $\{f_1, f_2, \ldots\}$, that is,

$$|\psi(t)\rangle = a(t)|B^0\rangle + b(t)|\overline{B}^0\rangle + c_1(t)|f_1\rangle + c_2(t)|f_2\rangle + \cdots\ . \tag{B.5}$$

If we are interested in computing only the values of $a(t)$ and $b(t)$ (and not the values of all $c_i(t)$), and if the times t in which we are interested are much larger than the typical strong interaction scale, then we can use a much simplified formalism [89]. The simplified time evolution is determined by a 2×2 effective Hamiltonian $\mathcal{H}$ that is not Hermitian, since otherwise the mesons would only oscillate and not decay. Any complex matrix, such as $\mathcal{H}$, can be written in terms of Hermitian matrices M and Γ as

$$\mathcal{H} = M - \frac{i}{2}\,\Gamma\ . \tag{B.6}$$

M and Γ are associated with $(B^0, \overline{B}^0) \leftrightarrow (B^0, \overline{B}^0)$ transitions via off-shell (dispersive) and on-shell (absorptive) intermediate states, respectively. Diagonal elements of M and Γ are associated with the flavor-conserving transitions $B^0 \to B^0$ and $\overline{B}^0 \to \overline{B}^0$ while off-diagonal elements are associated with flavor-changing transitions $B^0 \leftrightarrow \overline{B}^0$.

The eigenvectors of $\mathcal{H}$ have well defined masses and decay widths. We introduce complex parameters p and q to specify the components of the strong interaction eigenstates, B^0 and $\overline{B}^0$, in the light (B_L) and heavy (B_H) mass eigenstates:

$$|B_{L,H}\rangle = p|B^0\rangle \pm q|\overline{B}^0\rangle \tag{B.7}$$

with the normalization $|p|^2 + |q|^2 = 1$. The special form of Eq. (B.7) is related to the fact that CPT imposes $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$. Solving the eigenvalue problem gives

$$\left(\frac{q}{p}\right)^2 = \frac{M_{12}^* - (i/2)\Gamma_{12}^*}{M_{12} - (i/2)\Gamma_{12}}\ . \tag{B.8}$$

If either CP or T is a symmetry of $\mathcal{H}$, then M_{12} and Γ_{12} are relatively real, leading to

$$\left(\frac{q}{p}\right)^2 = e^{2i\xi_B} \quad \Rightarrow \quad \left|\frac{q}{p}\right| = 1\ , \tag{B.9}$$

where ξ_B is the arbitrary unphysical phase introduced in Eq. (B.2).

The real and imaginary parts of the eigenvalues of $\mathcal{H}$ corresponding to $|B_{L,H}\rangle$ represent their masses and decay-widths, respectively. The mass difference Δm_B and the width difference $\Delta\Gamma_B$ are defined as follows:

$$\Delta m_B \equiv M_H - M_L, \quad \Delta\Gamma_B \equiv \Gamma_H - \Gamma_L. \tag{B.10}$$

Note that here Δm_B is positive by definition, while the sign of $\Delta\Gamma_B$ is to be experimentally determined. The average mass and width are given by

$$m_B \equiv \frac{M_H + M_L}{2}, \quad \Gamma_B \equiv \frac{\Gamma_H + \Gamma_L}{2}. \tag{B.11}$$

It is useful to define dimensionless ratios x and y:

$$x \equiv \frac{\Delta m_B}{\Gamma_B}, \quad y \equiv \frac{\Delta\Gamma_B}{2\Gamma_B}. \tag{B.12}$$

Solving the eigenvalue equation gives

$$(\Delta m_B)^2 - \frac{1}{4}(\Delta\Gamma_B)^2 = (4|M_{12}|^2 - |\Gamma_{12}|^2), \quad \Delta m_B \Delta\Gamma_B = 4\mathcal{R}e(M_{12}\Gamma_{12}^*). \tag{B.13}$$

All CP-violating observables in B and $\overline{B}$ decays to final states f and $\overline{f}$ can be expressed in terms of phase-convention-independent combinations of A_f, $\overline{A}_f$, $A_{\overline{f}}$ and $\overline{A}_{\overline{f}}$, together with, for neutral-meson decays only, q/p. CP violation in charged-meson decays depends only on the combination $|\overline{A}_{\overline{f}}/A_f|$, while CP violation in neutral-meson decays is complicated by $B^0 \leftrightarrow \overline{B}^0$ oscillations and depends, additionally, on $|q/p|$ and on

$$\lambda_f \equiv (q/p)(\overline{A}_f/A_f). \tag{B.14}$$

For neutral D, B, and B_s mesons, $\Delta\Gamma/\Gamma \ll 1$ and so both mass eigenstates must be considered in their evolution. We denote the state of an initially pure $|B^0\rangle$ or $|\overline{B}^0\rangle$ after an elapsed proper time t as $|B^0{}_{\rm phys}(t)\rangle$ or $|\overline{B}^0_{\rm phys}(t)\rangle$, respectively. Using the effective Hamiltonian approximation, we obtain

$$\begin{aligned} |B^0{}_{\rm phys}(t)\rangle &= g_+(t)\,|B^0\rangle - \frac{q}{p}\,g_-(t)|\overline{B}^0\rangle, \\ |\overline{B}^0_{\rm phys}(t)\rangle &= g_+(t)\,|\overline{B}^0\rangle - \frac{p}{q}\,g_-(t)|B^0\rangle\,, \end{aligned} \tag{B.15}$$

where

$$g_\pm(t) \equiv \frac{1}{2}\left(e^{-im_H t - \frac{1}{2}\Gamma_H t} \pm e^{-im_L t - \frac{1}{2}\Gamma_L t}\right). \tag{B.16}$$

One obtains the following time-dependent decay rates:

$$\begin{aligned} \frac{d\Gamma[B^0{}_{\rm phys}(t)\to f]/dt}{e^{-\Gamma t}\mathcal{N}_f} &= \left(|A_f|^2 + |(q/p)\overline{A}_f|^2\right)\cosh(y\Gamma t) + \left(|A_f|^2 - |(q/p)\overline{A}_f|^2\right)\cos(x\Gamma t) \\ &+ 2\,\mathcal{R}e((q/p)A_f^*\overline{A}_f)\sinh(y\Gamma t) - 2\,\mathcal{I}m((q/p)A_f^*\overline{A}_f)\sin(x\Gamma t)\,, \end{aligned} \tag{B.17}$$

$$\begin{aligned} \frac{d\Gamma[\overline{B}^0_{\rm phys}(t)\to f]/dt}{e^{-\Gamma t}\mathcal{N}_f} &= \left(|(p/q)A_f|^2 + |\overline{A}_f|^2\right)\cosh(y\Gamma t) - \left(|(p/q)A_f|^2 - |\overline{A}_f|^2\right)\cos(x\Gamma t) \\ &+ 2\,\mathcal{R}e((p/q)A_f\overline{A}_f^*)\sinh(y\Gamma t) - 2\,\mathcal{I}m((p/q)A_f\overline{A}_f^*)\sin(x\Gamma t)\,, \end{aligned} \tag{B.18}$$

where $\mathcal{N}_f$ is a common normalization factor. Decay rates to the CP-conjugate final state $\overline{f}$ are obtained analogously, with $\mathcal{N}_f = \mathcal{N}_{\overline{f}}$ and the substitutions $A_f \to A_{\overline{f}}$ and $\overline{A}_f \to \overline{A}_{\overline{f}}$ in Eqs. (B.17,B.18). Terms proportional to $|A_f|^2$ or $|\overline{A}_f|^2$ are associated with decays that occur without any net $B \leftrightarrow \overline{B}$ oscillation, while terms proportional to $|(q/p)\overline{A}_f|^2$ or $|(p/q)A_f|^2$ are associated with decays following a net oscillation. The $\sinh(y\Gamma t)$ and $\sin(x\Gamma t)$ terms of Eqs. (B.17,B.18) are associated with the interference between these two cases. Note that, in multi-body decays, amplitudes are functions of phase-space variables. Interference may be present in some regions but not others, and is strongly influenced by resonant substructure.

C CP violation in neutral meson decays

CP asymmetries arise when two processes related by CP conjugation differ in their rates. Given the fact that CP violation is related to a phase in the Lagrangian, all CP asymmetries must arise from interference effects.

To date, CP violation has been observed (at a level higher than 5σ) in about thirty different hadron decay modes, involving b or c or s decays. It has not been established in other quark decays, nor in the leptonic sector, nor in flavor diagonal processes. Here we present the formalism relevant to measuring CP asymmetries in meson decays.

C.1 Notations and formalism

We discuss here the specific case of B-meson decays, but our discussion applies to all meson decays. Our starting points are Eqs. (B.17,B.18), which give the time-dependent decay rates of B^0 and $\overline{B}^0$. We also use the parameter λ_f, defined in Eq. (B.14).

Consider A_f, the $B \to f$ decay amplitude, and $\overline{A}_{\overline{f}}$, the amplitude of the CP conjugate process, $\overline{B} \to \overline{f}$. There are two types of phases that may appear in these decay amplitudes:

– CP-odd phases, also known as weak phases. They are complex parameters in any Lagrangian term that contributes to A_f, and appear in a complex conjugate form in $\overline{A}_{\overline{f}}$. In other words, CP violating phases change sign between A_f and $\overline{A}_{\overline{f}}$. In the SM, these phases appear only in the couplings of the $W^{\pm}$-bosons, hence the CP violating phases are called "weak phases".
– CP-even phases, also known as strong phases. Phases can appear in decay amplitudes even when the Lagrangian parameters are all real. They arise from contributions of intermediate on-shell states, and can be identified with the e^{-iHt} term in the time evolution Schrödinger equation. These CP conserving phases appear with the same sign in A_f and $\overline{A}_{\overline{f}}$. In meson decays, the intermediate states are typically hadronic state with the same flavor QN as the final state, and their dynamics is driven by strong interactions, hence the CP conserving phases are called "strong phases".

It is useful to factorize an amplitude into three parts: the magnitude a_i, the weak phase ϕ_i, and the strong phase δ_i. If there are two such contributions we write

$$A_f = a_1 e^{i(\delta_1+\phi_1)} + a_2 e^{i(\delta_2+\phi_2)}, \qquad \overline{A}_{\overline{f}} = a_1 e^{i(\delta_1-\phi_1)} + a_2 e^{i(\delta_2-\phi_2)}. \tag{C.1}$$

where we always can choose $a_1 \geq a_2$. It is further useful to define

$$\phi_f \equiv \phi_2 - \phi_1, \qquad \delta_f \equiv \delta_2 - \delta_1, \qquad r_f \equiv \frac{a_2}{a_1}. \tag{C.2}$$

For neutral meson mixing, it is useful to write

$$M_{B\overline{B}} = |M_{B\overline{B}}|e^{i\phi_M}, \qquad \Gamma_{B\overline{B}} = |\Gamma_{B\overline{B}}|e^{i\phi_\Gamma} , \tag{C.3}$$

and define

$$\theta_B = \phi_M - \phi_\Gamma. \tag{C.4}$$

Note that each of the phases appearing in Eqs. (C.1) and (C.3) is convention dependent, but combinations such as $\delta_1 - \delta_2$, $\phi_1 - \phi_2$, $\phi_M - \phi_\Gamma$, are physical.

In neutral meson decays, the phenomenology of CP violation is particularly rich thanks to the fact that meson mixing can contribute to the CP violating interference effects. One distinguishes three types of CP violation in meson decays, depending on which amplitudes interfere:

1. In decay: The interference is between two decay amplitudes. It corresponds to interference between a_1 and a_2.
2. In mixing: The interference is between the absorptive and dispersive mixing amplitudes. It corresponds to interference between $M_{B\overline{B}}$ and $\Gamma_{B\overline{B}}$.
3. In interference of decays with and without mixing: The interference is between the direct decay amplitude and a first-mix-then-decay amplitude. It corresponds to interference between $\overline{A}_f$ and $M_{B\overline{B}}A_f$.

We discuss these three types below.

For the discussion of CP violation in the $K^0 - \overline{K}^0$ system, we use a somewhat different notation. The reason is that, since the lifetimes of K_S and K_L are so different, experiments often identify these mass eigenstates, rather than the flavor-tagged decays, as done in most measurements of CP violation in the $B^0 - \overline{B}^0$ system. Thus, for K-mesons, we define

$$\epsilon_f \equiv \frac{1-\lambda_f}{1+\lambda_f}. \tag{C.5}$$

The converse relation reads

$$\lambda_f \equiv \frac{1-\epsilon_f}{1+\epsilon_f}. \tag{C.6}$$

Historically, CP violation was first observed in the $K_L \to \pi^+\pi^-$ decay and thus we denote $\epsilon_{\pi^+\pi^-} = \epsilon_K$. For modes with $|\bar{A}_f/A_f| - 1 \ll |q/p| - 1$, as is the case for $f = \pi^+\pi^-$, we can set $|\bar{A}_f/A_f| = 1$ and then we have $|q/p| = |\lambda_f|$.

C.2 CP violation in decay

CP violation in decay corresponds to

$$|\overline{A}_{\overline{f}}/A_f| \neq 1. \tag{C.7}$$

In charged particle decays, this is the only possible contribution to the CP asymmetry:

$$\mathcal{A}_f \equiv \frac{\Gamma(B^- \to f^-) - \Gamma(B^+ \to f^+)}{\Gamma(B^- \to f^-) + \Gamma(B^+ \to f^+)} = \frac{|\overline{A}_{f^-}/A_{f^+}|^2 - 1}{|\overline{A}_{f^-}/A_{f^+}|^2 + 1}. \tag{C.8}$$

Using Eq. (C.1), we obtain, for $r_f \ll 1$,

$$\mathcal{A}_f = 2 r_f \sin\phi_f \sin\delta_f. \tag{C.9}$$

This result shows explicitly that we need two decay amplitudes, that is, $r_f \neq 0$, with different weak phases, $\phi_f \neq 0, \pi$ and different strong phases, $\delta_f \neq 0, \pi$.

A few comments are in order:

1. In order to have a large CP asymmetry, we need each of the three factors in (C.9) not to be small.
2. A similar expression holds for the contribution of CP violation in decay in neutral meson decays. In this case there are, however, additional contributions from mixing, as discussed below.
3. Another complication with regard to neutral meson decays is that it is not always possible to tell the flavor of the decaying meson, that is, if it is B^0 or $\overline{B}^0$. This can be a problem or a virtue.
4. In general, the strong phase is not calculable since it is related to QCD. This is not a problem if the aim is just to demonstrate CP violation, but it is if we want to extract the weak parameter ϕ_f. In some cases, however, the strong phase can be independently measured, eliminating this particular source of theoretical uncertainty.

C.3 CP violation in mixing

CP violation in mixing corresponds to

$$|q/p| \neq 1\ . \tag{C.10}$$

In decays of neutral mesons into flavor specific final states ($\overline{A}_f = 0$ and, consequently, $\lambda_f = 0$), and, in particular, semileptonic neutral meson decays, this is the only source of CP violation:

$$\mathcal{A}_{\rm SL}(t) \equiv \frac{\hat\Gamma[\overline{B}^0(t) \to \ell^+ X] - \hat\Gamma[B^0(t) \to \ell^- X]}{\hat\Gamma[\overline{B}^0(t) \to \ell^+ X] + \hat\Gamma[B^0(t) \to \ell^- X]} = \frac{1 - |q/p|^4}{1 + |q/p|^4}. \tag{C.11}$$

Using Eq. (B.8), we obtain, for $|\Gamma_{B\overline{B}}/M_{B\overline{B}}| \ll 1$,

$$\mathcal{A}_{\rm SL} = -\left|\Gamma_{B\overline{B}}/M_{B\overline{B}}\right| \sin(\phi_M - \phi_\Gamma). \tag{C.12}$$

Two comments are in order:

1. Eq. (C.11) implies that $\mathcal{A}_{\rm SL}(t)$, which is an asymmetry of time-dependent decay rates, is actually time independent.
2. The calculation of $|\Gamma_{B\overline{B}}/M_{B\overline{B}}|$ is difficult, since it depends on low-energy QCD effects. Hence, the extraction of the value of the CP violating phase $\phi_M - \phi_\Gamma$ from a measurement of $\mathcal{A}_{\rm SL}$ involves, in general, large hadronic uncertainties.

CP violation in $K^0 - \overline{K}^0$ mixing is measured via a semileptonic asymmetry which is defined as follows:

$$\delta_L \equiv \frac{\Gamma(K_L \to \ell^+\nu_\ell\pi^-) - \Gamma(K_L \to \ell^-\nu_\ell\pi^+)}{\Gamma(K_L \to \ell^+\nu_\ell\pi^-) + \Gamma(K_L \to \ell^-\nu_\ell\pi^+)} = \frac{1-|q/p|^2}{1+|q/p|^2} \approx 2\mathcal{R}e(\epsilon_K), \tag{C.13}$$

where we use Eq. (C.6) and the fact that $|\epsilon_K| \ll 1$. This asymmetry is different from the one defined in Eq. (C.11) in that the decaying meson is the neutral mass eigenstate, rather than the flavor eigenstate, hence the different dependence on $|q/p|$.

C.4 CP violation in interference of decays with and without mixing

CP violation in interference of decays with and without mixing corresponds to

$$\mathcal{I}m(\lambda_f) \neq 0. \tag{C.14}$$

A particular simple case is the CP asymmetry in decays into final CP eigenstates. Moreover, a situation that is relevant in many cases is when one can neglect the effects of CP violation in decay and in mixing, that is when $|\overline{A}_{f_{CP}}/A_{f_{CP}}| \approx 1$ and $|q/p| \approx 1$. In this case, $\lambda_{f_{CP}}$ is, to a good approximation, a pure phase, $|\lambda_{f_{CP}}| = 1$. We further consider the case where we can neglect y ($|y| \ll 1$). Then,

$$\mathcal{A}_{f_{CP}}(t) \equiv \frac{\Gamma[\overline{B}^0(t) \to f_{CP}] - \Gamma[B^0(t) \to f_{CP}]}{\Gamma[\overline{B}^0(t) \to f_{CP}] + \Gamma[B^0(t) \to f_{CP}]} = \mathcal{I}m(\lambda_{f_{CP}})\sin(\Delta m_B t). \tag{C.15}$$

The approximations made above are valid in cases that $|\Gamma_{B\overline{B}}/M_{B\overline{B}}| \ll 1$ and $a_2 \ll a_1$, which lead to

$$\frac{q}{p} = \frac{M^*_{B\overline{B}}}{|M_{B\overline{B}}|} = e^{-i\phi_M}, \qquad \frac{\overline{A}_{f_{CP}}}{A_{f_{CP}}} = e^{-2i\phi_A}, \tag{C.16}$$

where ϕ_M is defined in Eq. (C.3), and $\phi_A = \phi_1$ is defined in Eq. (C.1). We then get

$$\mathcal{I}m(\lambda_{f_{CP}}) = \mathcal{I}m\left(\frac{M^*_{B\overline{B}}}{|M_{B\overline{B}}|}\frac{\overline{A}_{f_{CP}}}{A_{f_{CP}}}\right) = -\sin(\phi_M + 2\phi_A). \tag{C.17}$$

We learn that a measurement of a CP asymmetry in a process where these approximations are valid provides a direct probe of the weak phase between the mixing amplitude and the decay amplitude.

For the case where we measure decays of the K_L and K_S mass eigenstates into final CP-even eigenstates, one obtains

$$\mathcal{A}^{\rm mass}_{f_{CP}} \equiv \frac{\Gamma(K_L \to f_{CP})}{\Gamma(K_S \to f_{CP})} = \left|\frac{1-\lambda_{f_{CP}}}{1+\lambda_{f_{CP}}}\right|^2 = |\epsilon_{f_{CP}}|^2. \tag{C.18}$$

In particular, for $f_{CP} = \pi^+\pi^-$ we have

$$\mathcal{A}^{\rm mass}_{\pi^+\pi^-} = |\epsilon_K|^2. \tag{C.19}$$

References

[1] R. L. Workman *et al.* [Particle Data Group], *PTEP* **2022** (2022) 083C01, doi:10.1093/ptep/ptac097.

[2] N. Cabibbo, *Phys. Rev. Lett.* **10** (1963) 531–533, doi:10.1103/PhysRevLett.10.531.

[3] M. Kobayashi and T. Maskawa, *Prog. Theor. Phys.* **49** (1973) 652–657, doi:10.1143/PTP.49.652.

[4] L. Chau and W. Keung, *Phys. Rev. Lett.* **53** (1984) 1802–1805, doi:10.1103/PhysRevLett.53.1802.

[5] H. Harari and M. Leurer, *Phys. Lett. B* **181** (1986) 123–128, doi:10.1016/0370-2693(86)91268-2.

[6] L. Wolfenstein, *Phys. Rev. Lett.* **51** (1983) 1945–1947, doi:10.1103/PhysRevLett.51.1945.

[7] A.J. Buras, M.E. Lautenbacher, and G. Ostermaier, *Phys. Rev.* **D50** (1994) 3433–3446, doi:10.1103/PhysRevD.50.3433, arXiv:hep-ph/9403384.

[8] C. Jarlskog, *Phys. Rev. Lett.* **55** (1985) 1039–1042, doi:10.1103/PhysRevLett.55.1039.

[9] C. Jarlskog, *Z. Phys. C* **29** (1985) 491–497, doi:10.1007/BF01565198.

[10] CKMfitter Group (J. Charles *et al.*), *Eur. Phys. J. C***41** (2005) 1, doi:10.1140/epjc/s2005-02169-1, [hep-ph/0406184], doi:10.48550/arXiv.hep-ph/0406184, updated results and plots available at: `http://ckmfitter.in2p3.fr`.

[11] S.L. Glashow and S. Weinberg, *Phys. Rev. D* **15** (1977) 1958–1965, doi:10.1103/PhysRevD.15.1958.

[12] G.C. Branco, *Phys. Rev. Lett.* **44** (1980) 504–507, doi:10.1103/PhysRevLett.44.504.

[13] G.C. Branco, *Phys. Rev. D* **22** (1980) 2901–2905, doi:10.1103/PhysRevD.22.2901.

[14] S.L. Glashow, J. Iliopoulos and L. Maiani, *Phys. Rev. D* **2** (1970) 1285–1292, doi:10.1103/PhysRevD.2.1285.

[15] A.B. Carter and A.I. Sanda, *Phys. Rev. Lett.* **45** (1980) 952–954, doi:10.1103/PhysRevLett.45.952; *Phys. Rev. D* **23** (1981) 1567–1579, doi:10.1103/PhysRevD.23.1567.

[16] I.I.Y. Bigi and A.I. Sanda, *Nucl. Phys. B* **193** (1981) 85–108, doi:10.1016/0550-3213(81)90519-8.

[17] G. Buchalla, A.J. Buras, and M.E. Lautenbacher, *Rev. Mod. Phys.* **68** (1996) 1125–1244, doi:10.1103/RevModPhys.68.1125, arXiv:hep-ph/9512380.

[18] Y. Grossman, A. L. Kagan and Z. Ligeti, *Phys. Lett. B* **538** (2002) 327–334, doi:10.1016/S0370-2693(02)02027-0, arXiv:hep-ph/0204212.

[19] H. Boos, T. Mannel and J. Reuter, *Phys. Rev. D* **70** (2004) 036006, doi:10.1103/PhysRevD.70.036006, arXiv:hep-ph/0403085.

[20] H.-n. Li and S. Mishima, *JHEP* **03** (2007) 009, doi:10.1088/1126-6708/2007/03/009, arXiv:hep-ph/0610120.

[21] M. Gronau and J.L. Rosner, *Phys. Lett.* **B672** (2009) 349–353, doi:10.1016/j.physletb.2009.01.049, arXiv:0812.4796 [hep-ph].

[22] Y.S. Amhis *et al.* [HFLAV], *Eur. Phys. J. C* **81** (2021) 226, doi:10.1140/epjc/s10052-020-8156-7 and online update at `https://hflav.web.cern.ch/`.

[23] G.C. Branco, L. Lavoura and J.P. Silva, *CP violation,* Clarendon Press, Oxford (1999), doi:10.1093/oso/9780198503996.001.0001.

[24] I.I.Y. Bigi and N.G. Uraltsev, *Nucl. Phys. B* **592** (2001) 92–106, doi:10.1016/S0550-3213(00)00604-0, [arXiv:hep-ph/0005089], doi:10.48550/arXiv.hep-ph/0005089.

[25] A.F. Falk, Y. Grossman, Z. Ligeti and A.A. Petrov, *Phys. Rev. D* **65** (2002) 054034, doi:10.1103/PhysRevD.65.054034, arXiv:hep-ph/0110317.

[26] A.F. Falk, Y. Grossman, Z. Ligeti, Y. Nir and A.A. Petrov, *Phys. Rev. D* **69** (2004) 114021, doi:10.1103/PhysRevD.69.114021, arXiv:hep-ph/0402204.

[27] G. Raz, *Phys. Rev. D* **66** (2002) 037701, doi:10.1103/PhysRevD.66.037701, arXiv:hep-ph/0205310.

[28] G. Isidori, Y. Nir and G. Perez, *Ann. Rev. Nucl. Part. Sci.* **60** (2010) 355–380, doi:10.1146/annurev.nucl.012809.104534, arXiv:1002.0900 [hep-ph].

[29] N. Arkani-Hamed and S. Dimopoulos, *JHEP* **06** (2005) 073, doi:10.1088/1126-6708/2005/06/073, arXiv:hep-th/0405159.

[30] M. Ciuchini *et al.*, *Phys. Lett. B* **655** (2007) 162, doi:10.1016/j.physletb.2007.08.055, arXiv:hep-ph/0703204.

[31] Y. Nir, *JHEP* **05** (2007) 102, doi:10.1088/1126-6708/2007/05/102, arXiv:hep-ph/0703235.

[32] O. Gedalia *et al.*, *Phys. Lett. B* **714** (2012) 55, doi:10.1016/j.physletb.2012.06.050, arXiv:1202.5038 [hep-ph].

[33] K. Blum, Y. Grossman, Y. Nir and G. Perez, *Phys. Rev. Lett.* **102** (2009) 211802, doi:10.1103/PhysRevLett.102.211802, arXiv:0903.2118 [hep-ph].

[34] G. D'Ambrosio *et al.*, *Nucl. Phys. B* **645** (2002) 155–187, doi:10.1016/S0550-3213(02)00836-2, arXiv:hep-ph/0207036.

[35] C.D. Froggatt and H.B. Nielsen, *Nucl. Phys. B* **147** (1979) 277–298, doi:10.1016/0550-3213(79)90316-X.

[36] M. Dine, R.G. Leigh and A. Kagan, *Phys. Rev. D* **48** (1993) 4269–4274, doi:10.1103/PhysRevD.48.4269, arXiv:hep-ph/9304299.

[37] A.E. Nelson and M.J. Strassler, *JHEP* **09** (2000) 030, doi:10.1088/1126-6708/2000/09/030, arXiv:hep-ph/0006251.

[38] N. Arkani-Hamed and M. Schmaltz, *Phys. Rev. D* **61** (2000) 033005, doi:10.1103/PhysRevD.61.033005, arXiv:hep-ph/9903417.

[39] W. Altmannshofer, C. Frugiuele and R. Harnik, *JHEP* **12** (2014) 180, doi:10.1007/JHEP12(2014)180, arXiv:1409.2522 [hep-ph].

[40] M.C. Gonzalez-Garcia, M. Maltoni and T. Schwetz, *Universe* **7** (2021) 459, doi:10.3390/universe7120459.

[41] L.J. Hall, H. Murayama and N. Weiner, *Phys. Rev. Lett.* **84** (2000) 2572, doi:10.1103/PhysRevLett.84.2572, arXiv:hep-ph/9911341.

[42] N. Haba and H. Murayama, *Phys. Rev. D* **63** (2001) 053010, doi:10.1103/PhysRevD.63.053010, arXiv:hep-ph/0009174.

[43] A. de Gouvea and H. Murayama, *Phys. Lett. B* **573** (2003) 94–100, doi:10.1016/j.physletb.2003.08.045, arXiv:hep-ph/0301050; *Phys. Lett. B* **747** (2015) 479–483, doi:10.1016/j.physletb.2015.06.028, arXiv:1204.1249 [hep-ph].

[44] S. Amitai, arXiv:1211.6252 [hep-ph] (2012).

[45] P.F. Harrison, D.H. Perkins and W.G. Scott, *Phys. Lett. B* **530** (2002) 167–173, doi:10.1016/S0370-2693(02)01336-9, arXiv:hep-ph/0202074.

[46] E. Ma and G. Rajasekaran, *Phys. Rev. D* **64** (2001) 113012, doi:10.1103/PhysRevD.64.113012, arXiv:hep-ph/0106291.

[47] G. Altarelli and F. Feruglio, *Rev. Mod. Phys.* **82** (2010) 2701–2729, doi:10.1103/RevModPhys.82.2701, arXiv:1002.0211 [hep-ph]; *Nucl. Phys. B* **741** (2006) 215–235, doi:10.1016/j.nuclphysb.2006.02.015, arXiv:hep-ph/0512103.

[48] F.P. An *et al.* [DAYA-BAY Collaboration], *Phys. Rev. Lett.* **108** (2012) 171803, doi:10.1103/PhysRevLett.108.171803, arXiv:1203.1669 [hep-ex].

[49] S. Amitai, arXiv:1212.5165 [hep-ph] (2012).

[50] G. Aad *et al.* [ATLAS Collaboration], *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214 [hep-ex].

[51] S. Chatrchyan *et al.* [CMS Collaboration], *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235 [hep-ex].

[52] A. Tumasyan *et al.* [CMS Collaboration], *Phys. Rev. Lett.* **131** (2023) 061801, doi:10.1103/PhysRevLett.131.061801.

[53] G. Aad *et al.* [ATLAS Collaboration], *Eur. Phys. J. C* **82** (2022) 717, doi:10.1140/epjc/s10052-022-10588-3.

[54] G. Aad *et al.* [ATLAS Collaboration], *Phys. Lett. B* **801** (2020) 135148, doi:10.1016/j.physletb.2019.135148.

[55] A. Bazavov *et al.* [Fermilab Lattice, MILC and TUMQCD Collaborations], *Phys. Rev. D* **98** (2018) 054517, doi:10.1103/PhysRevD.98.054517.

[56] A. Tumasyan *et al.* [CMS Collaboration], *Phys. Rev. Lett.* **129** (2022) 032001, doi:10.1103/PhysRevLett.129.032001.

[57] A. Tumasyan *et al.* [CMS Collaboration], *JHEP* **02** (2022) 169, doi:10.1007/JHEP02(2022)169.

[58] A.M. Sirunyan *et al.* [CMS Collaboration], *Phys. Rev. Lett.* **125** (2020) 061801, doi:10.1103/PhysRevLett.125.061801.

[59] G. Aad *et al.* [ATLAS Collaboration], *Phys. Rev. Lett.* **125** (2020) 061802, doi:10.1103/PhysRevLett.125.061802.

[60] A. Tumasyan *et al.* [CMS Collaboration], *JHEP* **06** (2022) 012, doi:10.1007/JHEP06(2022)012.

[61] A. Dery *et al.*, *JHEP* **05** (2013) 039, doi:10.1007/JHEP05(2013)039, arXiv:1302.3229 [hep-ph].

[62] *Nucl. Phys.* **B448** (1995) 30–50, doi:10.1016/0550-3213(95)00203-5, arXiv:hep-ph/9502418.

[63] M. Leurer, Y. Nir and N. Seiberg, *Nucl. Phys. B* **420** (1994) 468, doi:10.1016/0550-3213(94)90074-4, arXiv:hep-ph/9310320.

[64] D. Bigi and P. Gambino, *Phys. Rev. D* **94** (2016) 094008, doi:10.1103/PhysRevD.94.094008 arXiv:1606.08030 [hep-ph].

[65] F.U. Bernlochner *et al.*, *Phys. Rev. D* **95** (2017) 115008, doi:10.1103/PhysRevD.95.115008, [erratum: *Phys. Rev. D* **97** (2018) 059902, doi:10.1103/PhysRevD.97.059902].

[66] D. Bigi, P. Gambino and S. Schacht, *JHEP* **11** (2017) 061, doi:10.1007/JHEP11(2017)061.

[67] S. Jaiswal, S. Nandi and S.K. Patra, *JHEP* **12** (2017) 060, doi:10.1007/JHEP12(2017)060.

[68] A. Abdesselam *et al.* [Belle Collaboration], arXiv:1904.08794 [hep-ex] (2019).

[69] R. Aaij *et al.* [LHCb Collaboration], *Phys. Rev. D* **97** (2018) 072013, doi:10.1103/PhysRevD.97.072013.

[70] R. Aaij *et al.* [LHCb Collaboration], *Phys. Rev. Lett.* **120** (2018) 171802, doi:10.1103/PhysRevLett.120.171802.

[71] S. Hirose *et al.* [Belle Collaboration], *Phys. Rev. D* **97** (2018) 012004, doi:10.1103/PhysRevD.97.012004.

[72] S. Hirose *et al.* [Belle Collaboration], *Phys. Rev. Lett.* **118** (2017) 211801, doi:10.1103/PhysRevLett.118.211801, arXiv:1612.00529 [hep-ex].

[73] R. Aaij *et al.* [LHCb Collaboration], *Phys. Rev. Lett.* **115** (2015) 111803, doi:10.1103/PhysRevLett.115.111803, [erratum: *Phys. Rev. Lett.* **115** (2015) 159901, doi:10.1103/PhysRevLett.115.159901].

[74] M. Huschle *et al.* [Belle Collaboration], *Phys. Rev. D* **92** (2015) 072014, doi:10.1103/PhysRevD.92.072014, arXiv:1507.03233 [hep-ex].

[75] J.P. Lees *et al.* [BaBar Collaboration], *Phys. Rev. D* **88** (2013) 072012, doi:10.1103/PhysRevD.88.072012, arXiv:1303.0571 [hep-ex].

[76] J.P. Lees *et al.* [BaBar Collaboration], textitPhys. Rev. Lett. **109** (2012) 101802, 10.1103/PhysRevLett.109.101802, arXiv:1205.5442 [hep-ex].

[77] S. Bressler, F.D. Halevy and Y. Nir, *JHEP* **07** (2022) 077, doi:10.1007/JHEP07(2022)077.

[78] F. Feruglio, P. Paradisi and A. Pattori, *Phys. Rev. Lett.* **118** (2017) 011801, doi:10.1103/PhysRevLett.118.011801, arXiv:1606.00524 [hep-ph].

[79] J.P. Lees *et al.* [BaBar Collaboration], *Phys. Rev. D* **86** (2012) 012004, doi:10.1103/PhysRevD.86.012004, arXiv:1204.2852 [hep-ex].

[80] J.P. Lees *et al.* [BaBar Collaboration], *Phys. Rev. D* **87** (2013) 112005, doi:10.1103/PhysRevD.87.112005, arXiv:1303.7465 [hep-ex].

[81] J. Grygier *et al.* [Belle Collaboration], *Phys. Rev. D* **96** (2017) 091101, doi:10.1103/PhysRevD.96.091101 [erratum *Phys. Rev. D* **97** (2018) 099902, doi:10.1103/PhysRevD.97.099902], arXiv:1702.03224 [hep-ex].

[82] Y. Grossman, A.L. Kagan and Y. Nir, *Phys. Rev. D* **75** (2007) 036008, doi:10.1103/PhysRevD.75.036008, arXiv:hep-ph/0609178 [hep-ph].

[83] R. Aaij *et al.* [LHCb Collaboration], *Phys. Rev. Lett.* **122** (2019) 211803, doi:10.1103/PhysRevLett.122.211803.

[84] R. Aaij *et al.* [LHCb Collaboration], *Phys. Rev. Lett.* **131** (2023) 091802, doi:10.1103/PhysRevLett.131.091802.

[85] Y. Grossman and S. Schacht, *JHEP* **07** (2019) 020, doi:10.1007/JHEP07(2019)020.

[86] S. Schacht, *JHEP* **03** (2023) 205, doi:10.1007/JHEP03(2023)205.

[87] A. Dery and Y. Nir, *JHEP* **12** (2019) 104, doi:10.1007/JHEP12(2019)104.

[88] R. Bause *et al.*, *Phys. Rev. D* **108** (2023) 035005, doi:10.1103/PhysRevD.108.035005.

[89] V. Weisskopf and E.P. Wigner, *Z. Phys.* **63** (1930) 54–73, doi:10.1007/BF01336768; *Z. Phys.* **65** (1930) 18–29, doi:10.1007/BF01397406, [See also P.K. Kabir, *The CP puzzle: Strange decays of the neutral kaon*, Academic Press (1968), Appendix A, p. 99, Internet Archive.]

Scientific programme[1]

Field Theory and the Standard Model
    Luis Alvarez-Gaume (Stony Brook)

Neutrino Physics
    Pilar Hernandez (IFIC, Univ. Valencia-CSIC)

Practical Statistics
    Eilam Gross (Weizmann Inst.)

Light and Ultra-Light Dark-Matter Physics
    Yonit Hochberg (Hebrew Univ.)

Higgs Physics
    Fabio Maltoni (Louvain)

Flavour Physics
    Yossi Nir (Weizmann Inst.)

Physics Beyond the Standard Model
    Riccardo Rattazzi (EPFL)

Cosmology
    Geraldine Servant *(DESY, Univ. Hamburg)*

Outlook for Experimental HEP
    Guy Wilkinson (Oxford)

Heavy-Ion Physics
    Itzhak Tserruya (Weizmann Inst.)

[1] Slides available at https://indico.cern.ch/event/940219/.

Organizing committees

Standing committee

Tatyana Donskova (JINR)
Claude Duhr (CERN)
Nick Ellis (CERN)
Markus Elsing (CERN)
Alexander Huss (CERN)
Martijn Mulders (CERN), Chair
Alexander Olshevskiy (JINR)
Kate Ross (CERN)
Sascha Stahl (CERN)

Local Organizing committee

Liron Barak (Tel Aviv Univ.)
Zvi Citron (Ben Gurion Univ.)
Eilam Gross (Weizmann Inst.)
Yonit Hochberg (Hebrew Univ.)
Eric Kuflik (Hebrew Univ.)
Gilad Perez (Weizmann Inst.), Chair
Yoram Rozen (Technion)
Hadar Alper (Weizmann Inst.), Administrator
Livnat Moses (Weizmann Inst.), Administrator

International advisors

Fabiola Gianotti (CERN)
Victor Matveev (JINR)
Grigori Trubnikov (JINR)
Valery Rubakov (INR)

List of lecturers

Luis Alvarez-Gaume (Stony Brook)
Fabiola Gianotti (CERN)
Eilam Gross (Weizmann Inst.)
Pilar Hernandez (IFIC, Univ. Valencia-CSIC)
Yonit Hochberg (Hebrew Univ.)
Fabio Maltoni (Louvain)
Yossi Nir (Weizmann Inst.)
Riccardo Rattazzi (EPFL)
Geraldine Servant (DESY, Univ. Hamburg)
Itzhak Tserruya (Weizmann Inst.)
Guy Wilkinson (Oxford)

List of discussion leaders

Jasmine Brewer (CERN)
Gauthier Durieux (CERN)
Marco Gorghetto (Weizmann)
Hyungjin Kim (DESY)
Davide Pagani (INFN Bologna)

Outreach training

Chris Jameson (Company Inside Edge)
Tony Prideaux (Company Inside Edge)

List of students

Vladimir ALLAKHVERDIAN
Pascal Raffael BAERTSCHI
Zhaoyu BAI
Lisa Marie BALTES
Patrick BAUER
Filippo BELLONI
Valentina BIANCACCI
Lukas BIERWIRTH
Marek BIROS
Denis BODROV
Maximilian BURKART
Mattia CAMPANA
Maria CARNESALE
Charlotte CAVANAGH
Elizaveta CHEREPANOVA
Federica COLOMBINA
Antoine COOLS
Ta'el COREN
Gonzalo Díaz LÓPEZ
Martin HABEDANK
Marcel HOHMANN
Huiling HUA
Shaked IRONI
Anna IVINA
Weijie JIN
Nilotpal KAKATI
Alexandra KALLITSOPOULOU
Nikolay KARPUSHKIN
Dmitrii KOBYLIANSKII
Ondrej KOVANDA
Martin KRIVOS
Jake LANE
Blaž LEBAN
Jingyan LI
Adelina LINTULUOTO
Marta LISOWSKA
Andrej LOZAR
Dvij MANKAD
Michele MANTOVANO
Giulia MARINELLI
Oliver MATONOHA
Arnaud Jean MAURY
Peter MEIRING
Jona MOTTA
Gadi NINIO
Julian NORTHEY
Giovanni PADOVANO
Santosh PARAJULI
Gujju Gurunadha REDDY
Markus REIF
Beatriz RIBEIRO LOPES
Charline ROUGIER
Adrian RUBIO JIMENEZ
Giovanna SALVI
Lorenzo SANTI
Roy SCHIMMEL BRENER
Caspar SCHMITT
Vladislav SHALAEV
Omer SHTAIF
Svetlana SILAEVA
Sahibjeet SINGH
Nathalie SOYBELMAN
Nadav Michael TAMIR
Nick THAMM
Raffaella TRAMONTANO
Jan VAN DER LINDEN
Dedu VLAD-GEORGE
Craig WELLS
Robert Stephen WHITE
Tianyu Justin YANG
Ilia ZHIZHIN

List of listeners

Abhishek BANERJEE
Yuval Yitzhak FRID
Inbar SAVORAY

List of posters

Poster title	Presenter
Smearing weights in the muon momentum scale calibration for the CMS W boson mass measurement	T.J.YANG
Searching for new phenomena in final states with leptons and jets using the ATLAS detector	R.S.BRENER, N.TAL HOD, A.SANTRA
Testing Lepton Flavor Universality with the First Measurement of R(K*) in ATLAS	D.C.MANKAD
The NA61/SHINE at the SPS, BM at Nica and CBM at FAIR	N. KARPUSHKIN
Differential ttbb measurement at the CMS experiment	J.VAN DER LINDEN
Tracking with a Graph Neutral Network	C.ROUGIER
First simultaneous differential measurement of tZq and ttZ with the CMS detector	F.COLOMBINA
Search for central exclusive production of top quark pairs with the CMS and TOTEM experiments	B.RIBEIRO LOPES
Search for New Physics in Leptonic Decays with the CMS Experiment at the LHC	V.SHALAEV, I.ZHIZHIN, S.SHMATOV
Using CMS Data Scouting to improve our understanding of the Higgs Boson	A.LINTULUOTO
Search for single production of a vector-like T quark decaying into a Higgs boson and top quark with fully hadronic final states using the ATLAS detector	J.FOO, J.HALEY, P.SINERVO, S.SINGH
The FASER Experiment: LHC Run 3 Collision Data	C.CAVANAGH
Measurement of jet p_T correlations in Xe+Xe collisions at $\sqrt{s_{NN}} = 5.44$ TeV with the ATLAS detector	M.KRIVOS
Search for doubly charged Higgs boson production in multi-lepton final states using 135 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector	B.LEBAN
Search for invisible particles together with single-top quarks using ATLAS Run 2 data	A.RUBIO JIMENEZ
Search for Rare Electroweak Decay B $\to$ K$^{(*)}\nu\bar{\nu}$ in New Physics Models at Belle II	C.SCHMITT
Charmless B-Decays at Belle II	M.REIF

Poster title	Presenter
New Gas Monitoring Chambers for T2K	T.Radermacher, S.Roth, D.Smyczek, N.Thamm
Development and firmware implementation of a machine learning based hadronic tau lepton Level-1 Trigger algorithm in CMS for the HL-LHC	J.Motta
Standard model precision physics with tau leptons with the ATLAS experiment at the LHC	G.Padovano, F.Morodei, M.Bauce, M.Corradi, S.Giagu
On a possibility to measure Michel parameters describing muon polarization in $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$	D.Bodrov, P.Pakhlov
Search for a Higgs boson associated with bottom quarks and decaying in tau leptons	P.Bärtschi
Dark matter searches in invisible Higgs boson decays with CMS	R.White
Improving the Forward Electron Performance of the ATLAS Detector	C.Wells
A search for additional Higgs bosons in $\tau\tau$ final states in proton-proton collisions at $\sqrt{s} = 13$ TeV	M.Burkart
Active volume determination of HPGe detectors and its impact in GERDA and LEGEND experiments	V.Biancacci
Measurement of the electroweak Wjj production with the ATLAS experiment	L.M.Baltes
Search for the single production of a vector-like quark Y→Wb in lepton+jet final states	G.Reddy
Search for heavy diboson resonances in semileptonic final states	E.Cherepanova
Instrumental Asymmetries	M.Mantovano
A novel method to measure the relavie strong phase between D^0 and $\bar{D}^0$ in $D \to K_s^0 \pi^+ \pi^-$ and its application to measurements of the CKM angle γ in $B^\pm \to D^0 \bar{D}^0 K^\pm$ decays	J.Lane, E.Gersabeck, J.Rademacker
The perturbation series solution of the time dependent radiative transfer equation	A.Vladimir, N.Dmitry
Fissile isotopes antineutrino spectra: Summation method and direct expleriment	P.Naumov, S.Silaeva, V.Sinev, A.Vlasenko
Sub-25 ps timing measurements in 10×10 cm^2 PICOSEC Micromegas detectors	M.Lisowska

Poster title	Presenter
Development of a PICOSEC Micromegas Detector for ENUBET	A.KALLITSOPOULOU
Neutron and beta imaging with optical readout Micromegas detector	A.COOLS ET AL.
Search for Higgs Boson Pair Production in the Multi-lepton Final State using Proton-Proton Collision Data at $\sqrt{s} = 13$ TeV from the ATLAS Detector	S.PARAJULI