CERN Yellow Reports: School Proceedings CERN-2021-005

Proceedings of the 2019 European School of High-Energy Physics

St Petersburg, Russia, 4–17 September 2019

Editors:
C. Duhr
M. Mulders

ISBN 978-92-9083-631-5 (paperback)
ISBN 978-92-9083-632-2 (PDF)
ISSN 2519-8041 (Print)
ISSN 2519-805X (Online)
DOI http://doi.org/10.23730/CYRSP-2021-005

Available online at http://publishing.cern.ch/ and http://cds.cern.ch/

This report should be cited as:

Proceedings of the 2019 European School of High-Energy Physics, St Petersburg, Russia, 4–17 September 2019, edited by C. Duhr and M. Mulders, CERN-2021-005 (CERN, Geneva, 2021), http://doi.org/10.23730/CYRSP-2021-005

A contribution in this report should be cited as:

[Author name(s)], in Proceedings of the 2019 European School of High-Energy Physics, St Petersburg, Russia, 4–17 September 2019, edited by C. Duhr and M. Mulders, CERN-2021-005 (CERN, Geneva, 2021), pp. [first page]–[last page], http://doi.org/10.23730/CYRSP-2021-005.[first page]

Corresponding editors: Claude.Duhr@cern.ch and Martijn.Mulders@cern.ch

Accepted by the CERN Reports Editorial Board (contact Carlos.Lourenco@cern.ch) in December 2021.

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

Indexed in the CERN Document Server, INSPIRE and in Scopus.

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

Abstract

The European School of High-Energy Physics is intended to give young physicists an introduction to the theoretical aspects of recent advances in elementary particle physics. These proceedings contain lecture notes on quantum field theory and the electroweak standard model, flavour physics and CP violation, neutrino physics, cosmology and dark matter, practical statistics for particle physics and LHC Run-2 highlights and future prospects.

Preface

The twenty-seventh event in the series of the European School of High-Energy Physics took place in St. Petersburg, Russia, from 4 to 17 September 2019. It was organized jointly by CERN, Geneva, Switzerland, and JINR, Dubna, Russia, with support from the Russian Academy of Sciences, the National Research Centre "Kurchatov Institute", and the Ministry of Science and Higher Education. The local organization team was chaired by Victor Kim (National Research Centre "Kurchatov Institute" – PNPI and SPbPU).

A total of 99 students of 35 different nationalities attended the school, mainly from institutes in member states of CERN and/or JINR, 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 New Peterhof Hotel, in the Peterhof district of St. Petersburg. According to the tradition of the School, the students shared twin rooms mixing participants of different nationalities.

A total of 30 lectures were complemented by daily discussion sessions led by six 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 Kate Shaw (University of Sussex), judged the presentations; other members of the jury were Veronica Sanz (University of Sussex), and Roger Barlow (University of Huddersfield). 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. The support of Inno-mir in the practical organization of the School is also gratefully acknowledged.

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.

We would like to express our strong appreciation to Fabiola Gianotti, Director-General of CERN, and Victor Matveev, Director of JINR, for their lectures on the scientific programmes of the two organizations and for discussing with the School participants. In addition to the rich academic programme, the participants enjoyed leisure and cultural activities in and around St. Petersburg. They attended a public outreach event followed by dinner in the city centre in historic buildings of the Russian Academy of Sciences. The outreach-event programme, which included a public lecture by Fabiola Gianotti followed by a round-table discussion on the role of science in society, was introduced by Victor Matveev with a video message from Grigoriy Trubnikov, First Deputy Minister of Science and Education of the Russian Federation. It attracted significant attention and brought together scientists and the general public, including researchers, university teachers and students, and also secondary-school pupils.

There was a half-day excursion to the beautiful Peterhof Palace and its famous fountains. A full-day excursion to St. Petersburg included a boat trip to the city centre and a guided tour of the Hermitage museum, followed by lunch and free time in the afternoon to explore the city, then dinner before returning to the hotel. On the final Saturday afternoon, the students were able to make use of the hotel facilities during free time or visit the city centre independently. The excursions provided an excellent environment for informal interactions between staff and students.

We are very grateful to the School Administrators, Kate Ross (CERN) and Tatyana Donskova (JINR), 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.

Nick Ellis
(On behalf of the Organizing Committee)

The 2019 European School of High-Energy Physics

The 2019 European School of High-Energy Physics
The 2019 European School of High-Energy Physics
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113

People in the photograph

1	Botho Paschen
2	Egor Frolov
3	Martijn Mulders
4	Mikhail Vysotskiy
5	Aleksei Chubykin
6	Armin Fehr
7	Rafael Eduardo Sosa Ricardo
8	Petr Kharlamov
9	Valery Rubakov
10	Henriette Petersen
11	Clara Elisabeth Leitgeb
12	Emin Nugaev
13	Yassine El Ghazali
14	Simone Meloni
15	Michael William O’Keefe
16	Alexandra Fell
17	Simona Ilieva
18	Cristina Ana Mantilla Suarez
19	Christos Vergis
20	Mohammed Faraj
21	Vyacheslav Moiseev
22	Konstantin Lehmann
23	Cristina Sánchez Gras
24	Hendrik Windel
25	Valerie Scheurer
26	Luis Ignacio Estevez Banos
27	Amanda Lynn Steinhebel
28	Giovanni Bartolini
29	Sergey Tolmachev
30	Eva Brottmann Hansen
31	Beatriz Garcia Plana
32	Milosz Zdybal
33	Joel Swallow
34	Mehrnoosh Moallemi
35	Nicole Michelle Hartman
36	Patrick Schwendimann
37	Eirik Hatlen
38	Krystsina Petukhova
39	Tobias Fitschen
40	Hanlin Xu
41	Victor Matveev
42	Guglielmo Frattari
43	Jonatan Adolfsson
44	Jean-Philippe Zopounidis
45	Lorenz Konrad Emberger
46	Kate Ross
47	Luigi Sabetta
48	Jakob Novak
49	James Mead
50	Alexander Bednyakov
51	Lesya Horyn
52	Alexander Olshevskiy
53	Luca Martinelli
54	Viktor Romanovskii
55	Michal Dragowski
56	Alexander Booth
57	Anastasia Kurova
58	Vasilije Perovic
59	Ioannis Xiotidis
60	Jonathan Gaunt
61	Vasilis Konstantinides
62	King Wai Kwok
63	Ryan Bernard Calladine
64	Iacopo Longarini
65	Michael Philipp Reichmann
66	Daniel Heuchel
67	Mario Grandi
68	Yvonne Ng
69	John Ellis
70	Alexander Huss
71	Terry WS Chan
72	Isabella Oceano
73	Davide Zuolo
74	Pepijn Johannes Bakker
75	Fabiola Gianotti
76	Viktor Sinetckii
77	Antonio Iuliano
78	Alessandro Da Rold
79	Jindrich Lidrych
80	Dias Kereibay
81	Maurizio De Santis
82	Dmitry Sosnov
83	Philip Daniel Keicher
84	Alexey Gladyshev
85	Valerio D’Amico
86	Janik Von Ahnen
87	Agostino De Iorio
88	Federico Vazzoler
89	Martin Klassen
90	Anton Shumakov
91	Laura Martikainen
92	Sebastian Bysiak
93	Denise Müller
94	Anatolii Egorov
95	Pu-Sheng Chen
96	Nick Ellis
97	Jakub Kandra
98	Deshan Kavishka Abhayasinghe
99	Feruzjon Ergashev
100	Daria Prokhorova
101	Tatyana Donskova
102	Louis Portales
103	Michele Veronesi
104	Jose Luis Munoz Martinez
105	Vitalii Okhotnikov
106	Lara Katharina Schildgen
107	Peter Major
108	Eleni Vryonidou
109	Mariia Didenko
110	Vasilii Plotnikov
111	Gogita Papalashvili
112	Karolina Juraskova
113	Victor Kim

PHOTOGRAPHS (MONTAGE)

The 2019 European School of High-Energy Physics
St. Petersburg, Russia, 4-17 September 2019
The 2019 European School of High-Energy Physics

ГОСТИНИЧНЫЙ КОМПЛЕКС
НОВЫЙ ПЕТЕРГОФ
ESHEP 2019

Contents

Proceedings of the 2019 European School of High-Energy Physics, St Petersburg, Russia, 14–27 September 2019, edited by C. Duhr and M. Mulders, CERN-2021-005 (CERN, Geneva, 2021)

Quantum field theory and the electroweak Standard Model

A.V. Bednyakov
Bogoliubov Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna, Russia

Abstract
In these lecture notes, we discuss the basics of quantum field theory, key ideas underlying the construction of the electroweak Standard Model, and some phenomenological manifestations of the latter. In addition, the present status, issues, and prospects of the SM are briefly covered.

Keywords
Quantum field theory; Standard Model; Electroweak interactions; Lectures.

1 Introduction

The Standard Model (SM), see Refs. [1–3], turns out to be an incredibly successful theory, which survived many stringent experimental tests. Even after almost half a century, no significant deviations from the SM predictions have been found. Moreover, the discovery, see Refs. [4,5], of the Higgs boson at the LHC in 2012 was the final step in finalizing the SM. It is fair to say that it fully deserves the following fancy name:

The Absolutely Amazing Theory of Almost Everything. [6]

Let us mention a few excellent lectures (e.g., Refs. [7–11]) and textbooks (e.g., Refs. [12–14]) that can convince the reader that it is indeed the case. Since the history of the SM is rather long, it is obvious that it is not possible to discuss all the peculiarities of the SM in the set of three lectures. So the main task of the course is to review some basic facts and underlying principles of the model and emphasize key features of the latter.

Let us start with a brief overview of the SM particle content. The SM particles fall into two categories: fermions (half-integer spin) from bosons (integer spin). The former traditionally[1] associated with “matter”, while the latter take the role of “force carriers” that mediate interactions between spin-1/2 particles.

In the SM, we have three *generations* involving two types of fermions - *quarks* and *leptons*. In total, there are

- 6 quarks of different flavour ($q = u, d, c, s, t, b$),
- 3 charged ($l = e, \mu, \tau$) and 3 neutral ($\nu_l = \nu_e, \nu_\mu, \nu_\tau$) leptons.

All of them participate in *weak* interactions. Both quarks q and charged leptons l take part in the electromagnetic interactions. In addition, quarks carry a *colour* charge and are influenced by the strong force. In the SM these interactions are due to the exchange of spin-1 (or vector) bosons:

- 8 gluons mediate the strong force between quarks;
- 4 electroweak (EW) bosons are responsible for the electromagnetic (photon - γ) and weak ($Z, W^\pm$) interactions.

There is also a spin-0 Higgs boson h. As it will be obvious from the subsequent discussion it plays a very important role in the construction of the SM. It turns out that only gluons and photons (γ) are assumed to be massless. All other *elementary* particles are massive due to the *Higgs* mechanism.

[1] The distinction is outdated: the fermions also mediate interactions between bosons.

http://doi.org/10.23730/CYRSP-2021-005.1

In the SM the properties of the particle interactions can be read off the SM *Lagrangian* $\mathcal{L}_{SM}$. One can find its compact version on the famous CERN T-shirt. However, there is a lot of structure behind the short expression and it is *quantum field theory* or QFT (see, e.g., Refs. [14–18]) that allows us to derive the full Lagrangian and understand why the T-shirt Lagrangian is unique in a sense.

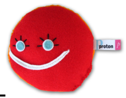
The form of $\mathcal{L}_{SM}$ is *restricted* by various kinds of (postulated) *symmetries*. Moreover, the SM is a *renormalizable* model. The latter fact allows us to use *perturbation theory* (PT) to make high-precision predictions for a vast number of observables and confront the model with experiment. All these peculiarities will be discussed during the lectures, which have the following structure.

We begin by introducing basics of quantum field theory in Section 2. Then we emphasize the role of symmetries in particles physics in Section 3. In Section 4 we use the *gauge principle* to construct the electroweak SM. The discussion of some experimental tests of the SM theoretical predictions can be found in Section 5. Final remarks and conclusions are provided in Section 6.

2 Basics of quantum field theory

2.1 Units, notation and all that

Before we begin our discussion of quantum fields, let us set up our notation. We work in natural units with the speed of light $c = 1$ and the (reduced) Planck constant $\hbar = 1$. As a consequence, all the quantities in particle physics are expressed in powers of electron-Volts (eV). To recover ordinary units, one uses the following convenient conversion factors:

[ ] $\hbar \simeq 6.58 \cdot 10^{-22}\ \text{MeV} \cdot s,\qquad \hbar c \simeq 1.97 \cdot 10^{-14}\ \text{GeV} \cdot \text{cm}$ []. (1)

In high-energy physics (HEP) we usually *require* that our theory should respect Lorentz *symmetry*. Due to this, a rotation or a boost in some direction, which can be parametrized by $\Lambda_{\mu\nu}$:

$$x_\mu \to x'_\mu = \Lambda_{\mu\nu} x_\nu, \tag{2}$$

does not change the value of scalar product

$$px \equiv p_\mu x_\mu = g_{\mu\nu} p_\mu x_\nu = p_0 x_0 - \mathbf{p}\cdot\mathbf{x}, \quad g_{\mu\nu} = \text{diag}(1,-1,-1,-1) \tag{3}$$

of any two four-vectors, e.g., space-time coordinates x_μ and energy-momenta p_μ

$$\begin{aligned} x_\mu &= \{x_0, \mathbf{x}\},\ \text{with time } t \equiv x_0, \\ p_\mu &= \{p_0, \mathbf{p}\},\ \text{with energy } E \equiv p_0. \end{aligned}$$

A well-known and very important example of a Lorentz invariant quantity is the particle *mass*. The latter corresponds to the "length" of the four-momentum vector $p^2 = E^2 - \mathbf{p}^2 = m^2$ and is the key property of a particle. Now let us switch to our main topic and discuss how fields are used to account for relativistic particles.

2.2 Quantum scalar field

A convenient way to deal with (quantum) fields is to consider the *Action* functional[2]. For the simplest (scalar) field, i.e., a function $\phi(x) \equiv \phi(t, \mathbf{x})$, the Action can be written as

$$\mathcal{A}[\phi(x)] = \int d^4x\ \underbrace{\mathcal{L}(\phi(x), \partial_\mu\phi)}_{\text{Lagrangian (density)}} = \int d^4x\ \underbrace{\left(\partial_\mu\phi^\dagger\partial_\mu\phi - m^2\phi^\dagger\phi\right)}_{\phi^\dagger\cdot K\cdot\phi}. \tag{4}$$

[2] Contrary to ordinary functions that produce numbers from numbers, a *functional* takes a function and produces a number.

Given the Lagrangian $\mathcal{L}$, one can derive the *equations of motions* (EOM) via the *Action Principle*, which we describe now. The variation of the action due to tiny (infinitesimal) shifts in the field $\phi'(x) = \phi(x) + \delta\phi(x)$ can be cast into

$$\underbrace{\mathcal{A}[\phi'(x)] - \mathcal{A}[\phi(x)]}_{\delta\mathcal{A}[\phi(x)]=0} = \int d^4x \left[\underbrace{\left(\partial_\mu \frac{\partial \mathcal{L}}{\partial\partial_\mu\phi} - \frac{\partial \mathcal{L}}{\partial\phi} \right)}_{(\partial_\mu^2+m^2)\phi=0} \delta\phi + \underbrace{\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial\partial_\mu\phi} \delta\phi \right)}_{\text{surface term=0}} \right]. \tag{5}$$

If we require that $\delta\mathcal{A}[\phi(x)] = 0$ for *any* variation $\delta\phi(x)$ of *some* $\phi(x)$, we will immediately deduce that this can be achieved only for *specific* $\phi(x)$ that satisfy EOM. These *particular* fields are usually called “on-mass-shell”. In the case of the scalar field $\phi(x)$ we derive the Klein-Gordon (KG) equation, which is related in a straightforward way to the quadratic form K in Eq. (4):

$$-K\phi(x) = \left(\partial_\mu^2 + m^2\right)\phi = \left(\partial_t^2 - \nabla^2 + m^2\right)\phi(x) = 0. \tag{6}$$

After Fourier transformation (FT) Eq. (6) leads to the energy-momentum relation for the non-interacting particle,i.e.,

$$\phi(x) = \frac{1}{(2\pi)^4}\int d^4p\phi(p)e^{-ipx} \Rightarrow \left(p^2 - m^2\right)\phi(p) = \left(p_0^2 - \mathbf{p}^2 - m^2\right)\phi(p) = 0. \tag{7}$$

General solution of the homogeneous KG equation can be written as a sum (integral) over plane waves with $p_0^2 = \vec{p}^2 + m^2$

$$\phi(t, \mathbf{x}) = \frac{1}{(2\pi)^{3/2}}\int \frac{d\mathbf{p}}{\sqrt{2\omega_p}}\left[a(\mathbf{p})e^{-i\omega_p t + i\mathbf{p}\mathbf{x}} + b(\mathbf{p})e^{+i\omega_p t - i\mathbf{p}\mathbf{x}}\right], \tag{8}$$

where $\omega_p \equiv +\sqrt{\mathbf{p}^2 + m^2}$. For further convenience we explicitly write the terms corresponding to $p_0 = +\omega_p$ and $p_0 = -\omega_p$. The *negative-energy* solution with $p_0 < 0$ poses a serious problem if the field Eq. (8) is interpreted as *a wave-function* of *single* particle in the context of relativistic quantum mechanics (RQM). A single-particle interpretation fails to account for the appearance of negative-energy modes, and a new formalism is required to deal with such situations (see, e.g., Refs. [14]). Moreover, in RQM space coordinates play a role of dynamical variables and are represented by operators, while time is an evolution parameter. Obviously, a *consistent* relativistic theory should treat space and time on equal footing.

In QFT we interpret $\phi(\mathbf{x}, t)$ satisfying Eq. (6) as a *quantum* field, i.e., an *operator*[3]. The space coordinates $\mathbf{x}$ can be treated as a *label* for infinitely many dynamical variables, and we are free to choose a system of reference, in which we evolve these variables. As a consequence, a single field can account for an infinite number of particles, which correspond to field *excitations*.

Rewriting Eq. (8) in the compact QFT notation $[a(\mathbf{p}) \to a^-_{\mathbf{p}}, b(\mathbf{p}) \to b^-_{\mathbf{p}}]$

$$\phi(x) = \frac{1}{(2\pi)^{3/2}}\int \frac{d\mathbf{p}}{\sqrt{2\omega_p}}\left[a^-_{\mathbf{p}}e^{-ipx} + b^+_{\mathbf{p}}e^{+ipx}\right], \tag{9}$$

we interpret $a^\pm_{\mathbf{p}}$ and $b^\pm_{\mathbf{p}}$ as *operators* obeying

$$a^-_{\mathbf{p}}a^+_{\mathbf{p}'} - a^+_{\mathbf{p}'}a^-_{\mathbf{p}} \equiv \left[a^-_{\mathbf{p}}, a^+_{\mathbf{p}'}\right] = \delta^3(\mathbf{p} - \mathbf{p}'), \quad \left[b^-_{\mathbf{p}}, b^+_{\mathbf{p}'}\right] = \delta^3(\mathbf{p} - \mathbf{p}'). \tag{10}$$

[3] We use the Heisenberg picture, in which operators $\mathcal{O}_H(t)$ depend on time, while in the Schrödinger picture it is the states that evolve: $\langle\psi(t)|\mathcal{O}_S|\psi(t)\rangle = \langle\psi|\mathcal{O}_H(t)|\psi\rangle$ with $\mathcal{O}_S = \mathcal{O}_H(t=0), |\psi\rangle = |\psi(t=0)\rangle$.

All other commutators are assumed to be zero, e.g., $\left[a^{\pm}_{\mathbf{p}}, a^{\pm}_{\mathbf{p}'}\right] = 0$. The operators satisfy $a^{\pm}_{\mathbf{p}} = (a^{\mp}_{\mathbf{p}})^{\dagger}$ and $b^{\pm}_{\mathbf{p}} = (b^{\mp}_{\mathbf{p}})^{\dagger}$, and for $a^{\pm}_{\mathbf{p}} \equiv b^{\pm}_{\mathbf{p}}$ the field is hermitian $\phi^{\dagger}(x) = \phi(x)$. The commutation relations, see Eq. (10), resemble the relations $[a^-, a^+] = 1$ for ladder operators $a^{\pm}$, which are usually considered to quantize harmonic oscillators. Following the analogy, we consider the *Fock* space that consists of a *vacuum* $|0\rangle$, which is *annihilated* by $a^-_{\mathbf{p}}$ (and $b^-_{\mathbf{p}}$) for every $\mathbf{p}$

$$\langle 0|0\rangle = 1, \quad a^-_{\mathbf{p}}|0\rangle = 0, \quad \langle 0|a^+_{\mathbf{p}} = (a^-_{\mathbf{p}}|0\rangle)^{\dagger} = 0,$$

and field excitations, which are *created* from the vacuum by acting with $a^+_{\mathbf{k}}$ (and/or $b^+_{\mathbf{k}}$) , e.g.,

$$|f_1\rangle = \int d\mathbf{k} \cdot f_1(\mathbf{k}) a^+_{\mathbf{k}}|0\rangle, \qquad \text{1-particle state;} \tag{11}$$

$$|f_2\rangle = \int d\mathbf{k}_1 d\mathbf{k}_2 \cdot f_2(\mathbf{k}_1, \mathbf{k}_2) a^+_{\mathbf{k}_1} a^+_{\mathbf{k}_2}|0\rangle \qquad \text{2-particle state,} \tag{12}$$

$$\ldots$$

Here various $f_i(\mathbf{k}, \ldots)$ are supposed to be square-integrable, so that, e.g., $\langle f_1|f_1\rangle = \int |f_1(\mathbf{k})|^2 d\mathbf{k} < \infty$. In spite of the fact that it is more appropriate to deal with such normalizable states, in QFT we usually consider (basis) states that have definite momentum $\mathbf{p}$, i.e., we assume that $f_1(\mathbf{k}) = \delta(\mathbf{k} - \mathbf{p})$.

The two set of operators $a^{\pm}$ and $b^{\pm}$ correspond to particles and antiparticles. It is worth emphasizing that in QFT all the particles of certain kind are excitations of a *single* field, and due to $a^+_{\mathbf{p}} a^+_{\mathbf{k}} = a^+_{\mathbf{k}} a^+_{\mathbf{p}}$, they are *indistinguishable* by construction.

Exploiting again the analogy with harmonic oscillators, we can introduce the Hamiltonian operator

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_{part} + \hat{\mathcal{H}}_{antipart} = \int d\mathbf{p}\, \omega_p \left[n_{\mathbf{p}} + \bar{n}_{\mathbf{p}}\right] \tag{13}$$

with $\bar{n}_{\mathbf{p}} \equiv b^+_{\mathbf{p}} b^-_{\mathbf{p}}$ and $n_{\mathbf{p}} \equiv a^+_{\mathbf{p}} a^-_{\mathbf{p}}$. The interpretation of the terms in Eq. (13) is straightforward: ($\bar{n}_{\mathbf{p}}$) $n_{\mathbf{p}}$ *counts* (anti-)particles with definite momentum $\mathbf{p}$ and there is a sum over the corresponding energies. In writing Eq. (13) we omit the infinite constant (zero-point energies) and from the very beginning assume that all the operators in $\hat{\mathcal{H}}$ are normal-ordered, i.e., all creation operators go before the annihilation ones. This corresponds to the assumption that the vacuum state has zero-energy $\hat{\mathcal{H}}|0\rangle = 0$.

It is easy to check that $[\hat{\mathcal{H}}, a^{\pm}_{\mathbf{p}}] = \pm\omega_{\mathbf{p}} a^{\pm}_{\mathbf{p}}$ and $[\hat{\mathcal{H}}, b^{\pm}_{\mathbf{p}}] = \pm\omega_{\mathbf{p}} b^{\pm}_{\mathbf{p}}$. As a consequence, single-particle states with definite momentum $\mathbf{p}$

$$|\mathbf{p}\rangle = a^+_{\mathbf{p}}|0\rangle, \quad \hat{\mathcal{H}}|\mathbf{p}\rangle = \omega_p|\mathbf{p}\rangle, \qquad |\bar{\mathbf{p}}\rangle = b^+_{\mathbf{p}}|0\rangle, \quad \hat{\mathcal{H}}|\bar{\mathbf{p}}\rangle = \omega_p|\bar{\mathbf{p}}\rangle \tag{14}$$

are eigenvectors of the Hamiltonian with *positive* energies, and we avoid introduction of negative energies in our formalism from the very beginning. One can generalize Eq. (13) and "construct" the momentum $\hat{\mathbf{P}}$ and charge $\hat{Q}$ operators[4]:

$$\hat{\mathbf{P}} = \int d\mathbf{p}\, \mathbf{p} \left[n_{\mathbf{p}} + \bar{n}_{\mathbf{p}}\right], \quad \hat{\mathbf{P}}|0\rangle = 0|0\rangle, \qquad \hat{\mathbf{P}}|\mathbf{p}\rangle = \mathbf{p}|\mathbf{p}\rangle \qquad \hat{\mathbf{P}}|\mathbf{p}\rangle = \mathbf{p}|\bar{\mathbf{p}}\rangle, \tag{15}$$

$$\hat{Q} = \int d\mathbf{p} \left[n_{\mathbf{p}} - \bar{n}_{\mathbf{p}}\right], \quad \hat{Q}|0\rangle = 0|0\rangle, \qquad \hat{Q}|\mathbf{p}\rangle = +|\mathbf{p}\rangle \qquad \hat{Q}|\bar{\mathbf{p}}\rangle = -|\bar{\mathbf{p}}\rangle. \tag{16}$$

The charge operator $\hat{Q}$ distinguishes particles from anti-particles. One can show that the field $\phi^{\dagger}$ (ϕ) increases (decreases) the charge of a state

$$\left[\hat{Q}, \phi^{\dagger}(x)\right] = +\phi^{\dagger}(x), \quad \left[\hat{Q}, \phi(x)\right] = -\phi(x)$$

and consider the following amplitudes:

[4] It is worth pointing here that by construction both $\hat{Q}$ and $\hat{\mathbf{P}}$ do not depend on time and commute. In the next section, we look at this fact from a different perspective and connect it to various symmetries.

$$t_2 > t_1: \quad \langle 0|\underbrace{\phi(x_2)}_{a^-}\underbrace{\phi^\dagger(x_1)}_{a^+}|0\rangle \qquad t_1 > t_2: \quad \langle 0|\underbrace{\phi^\dagger(x_1)}_{b^-}\underbrace{\phi(x_2)}_{b^+}|0\rangle$$

Particle (charge $+1$) propagates from x_1 to x_2

Antiparticle (charge -1) propagates from x_2 to x_1

We can take both possibilities into account in one function:

$$\underbrace{\langle 0|T[\phi(x_2)\phi^\dagger(x_1)]|0\rangle}_{-iD_c(x-y)} \equiv \theta(t_2-t_1)\langle 0|\phi(x_2)\phi^\dagger(x_1)|0\rangle + \theta(t_1-t_2)\langle 0|\phi^\dagger(x_1)\phi(x_2)|0\rangle, \qquad (17)$$

with T being the *time-ordering* operation ($\theta(t) = 1$ for $t \geq 0$ and zero otherwise).

Equation (17) give rise to the famous Feynman propagator, which has the following momentum representation:

$$D_c(x-y) = \frac{-1}{(2\pi)^4}\int d^4p \frac{e^{-ip(x-y)}}{p^2 - m^2 + i\epsilon}. \qquad (18)$$

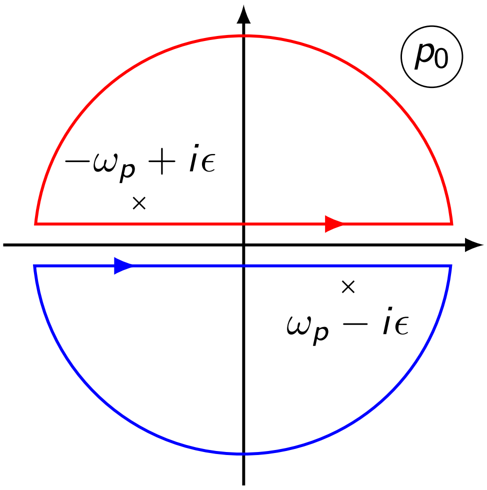


Fig. 1: Integration contours in p_0 plane.

The $i\epsilon$-prescription ($\epsilon \to 0$) picks up certain poles in the complex p_0 plane (see Fig. 1) and leads to the time-ordered expression Eq. (17). The propagator plays a key role in the construction of perturbation theory for interacting fields (see Section 2.5).

For the moment, let us mention a couple of facts about $D_c(x)$. It is a Green-function for the KG equation, i.e.,

$$\left(\partial_x^2 + m^2\right) D_c(x-y) = \delta(x-y). \qquad (19)$$

This gives us an alternative way to find the expression Eq. (18) by inverting the quadratic form introduced in Eq. (4). One can also see that $D_c(x-y)$ is a Lorentz and translational invariant function.

The propagator of particles can be connected to the force between two classical static sources $J_i(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_i)$ located at $\mathbf{x}_i = (\mathbf{x}_1, \mathbf{x}_2)$. The sources disturb the vacuum $|0\rangle \to |\Omega\rangle$, since the Hamiltonian of the system is modified $\mathcal{H} \to \mathcal{H}_0 + J\cdot\phi$. Assuming for simplicity that $\phi = \phi^\dagger$, we can find the energy of the disturbed vacuum from

$$\langle\Omega|e^{-i\mathcal{H}T}|\Omega\rangle \equiv e^{-iE_0(J)T} \Rightarrow \text{in the limit } T \to \infty$$
$$= e^{\frac{i^2}{2!}\int dxdy J(x)\langle 0|T(\phi(x)\phi(y))|0\rangle J(y)} = e^{+\frac{i}{2}\int dxdy J(x) D_c(x-y) J(y)}$$

Evaluating the integral for $J(x) = J_1(x) + J_2(x)$ and neglecting “self-interactions“, we get the contribution δE_0 to $E_0(J)$ due to interactions between two sources

$$\lim_{T\to\infty} \delta E_0 T = -\int dxdy J_1(x) D_c(x-y) J_2(y)$$
$$\delta E_0 = -\int \frac{d\mathbf{p}}{(2\pi)^3}\frac{e^{+i\mathbf{p}(\mathbf{x}_1-\mathbf{x}_2)}}{\mathbf{p}^2+m^2} = -\frac{1}{4\pi r}e^{-mr}, \qquad r = |\mathbf{x_1} - \mathbf{x_2}|$$

This is nothing else but the *Yukawa* potential. It is *attractive* and *falls off* exponentially over the distance scale $1/m$. Obviously, for $m = 0$ we get a Coulomb-like potential.

2.3 More degrees of freedom?

Up to now we were discussing simple scalar particles, which describe only one (in the case of hermitian field) degree of freedom (per space point). We can extend our formalism to account for fields involving several degrees of freedom by adding more (and more) *indices* to $\phi(x) \to \Phi^i_\alpha(x)$, and treating the latter as components of generalized $\Phi(x)$ in some field space. One conveniently splits the indices into two groups: *space-time* (α) and *internal* (i). Under space-time, e.g., Lorentz Eq. (2), transformations $x \to x'$, we also have $\Phi^i_\alpha(x) \to \Phi'^i_\alpha(x')$, where

$$\Phi'^i_\alpha(\Lambda x) = S_{\alpha\beta}(\Lambda)\Phi^i_\beta(x) \quad \text{(Lorentz transform)} \tag{20}$$

is a linear combination of "old" fields Φ^i_α having the same index i. Analogously, one considers rotations in the "internal" space $\Phi^i_\alpha(x) \to \Phi'^i_\alpha(x)$

$$\Phi'^i_\alpha(x) = \ U^{ij}\Phi^j_\alpha(x) \qquad \text{(Internal transform)} \tag{21}$$

that are characterized by some matrix U_{ij}, which acts only on internal indices. A quantum field in this case is represented as

$$\Phi^i_\alpha(x) = \frac{1}{(2\pi)^{3/2}} \sum_s \int \frac{d\mathbf{p}}{\sqrt{2\omega_p}} \left[u^s_\alpha(\mathbf{p})\,(a^-_\mathbf{p})^i_s\, e^{-ipx} + v^{*s}_\alpha(\mathbf{p})(b^+_\mathbf{p})^i_s\, e^{+ipx}\right]. \tag{22}$$

Here the factors $e^{\pm ipx}$ with $p_0 = \omega_\mathbf{p}$ (plane waves) guarantee that every component of Φ^i_α satisfies the KG equation. The sum in Eq. (22) is over all polarization states, which are characterized by polarization "vectors" for particles $u^s_\alpha(\mathbf{p})$ annihilated by $(a^-_\mathbf{p})^i_s$, and anti-particles $v^{*s}_\alpha(\mathbf{p})$ created by $(b^+_\mathbf{p})^i_s$. The conjugated field $(\Phi^i_\alpha)^\dagger$ involves (conjugated) polarization vectors for (anti) particles that are (annihilated) created. Let us give a couple of examples:

– Quarks are *coloured fermions* Ψ^i_α and, e.g., $(a^-_\mathbf{p})^b_s$ annihilates the "blue" quark in a spin state s. The latter is characterized by a spinor $u^s_\alpha(\mathbf{p})$;
– There are *eight vector* gluons G^a_μ. So $(a^-_\mathbf{p})^a_s$ annihilates a gluon a in spin state s having polarization $u^s_\alpha(\mathbf{p}) \to \epsilon^s_\mu(\mathbf{p})$.

One can notice that the Lorentz transformations plays a key role in QFT. We can classify our fields as different *representations* of the corresponding *group*. Since in the SM (and, actually, in other *renormalizable* four-dimensional QFT models) we only deal with spin-0, spin-1/2, and spin-1 fields, let us elaborate on the formalism used to describe vector (spin-1) and fermion (spin-1/2) particles.

2.3.1 Massive vector fields

A charged vector field (e.g., a W-boson) can be written as

$$W_\mu(x) = \frac{1}{(2\pi)^{3/2}} \sum_{\lambda=1}^{3} \int \frac{d\mathbf{p}}{\sqrt{2\omega_p}} \left[\left(\epsilon^\lambda_\mu(\mathbf{p}) a^-_\lambda(\mathbf{p}) e^{-ipx} + \epsilon^{*\lambda}_\mu(\mathbf{p}) b^+_\lambda(\mathbf{p})\, e^{+ipx}\right)\right]. \tag{23}$$

A massive spin-1 particle has *3* independent polarization vectors, which satisfy

$$p_\mu \epsilon^\lambda_\mu(\mathbf{p}) = 0, \quad \epsilon^\lambda_\mu(\mathbf{p})\epsilon^{*\lambda'}_\mu(\mathbf{p}) = -\delta^{\lambda\lambda'}, \quad \sum_{\lambda=1}^{3} \epsilon^\lambda_\mu \epsilon^{*\lambda}_\nu = -\left(g_{\mu\nu} - \frac{p_\mu p_\nu}{m^2}\right) \quad [p_0 = \omega_\mathbf{p}].$$

The Feynman propagator can be found by considering time-ordered product of two fields

$$\langle 0|T(W_\mu(x) W^\dagger_\nu(y))|0\rangle = \frac{1}{(2\pi)^4} \int d^4 p e^{-ip(x-y)} \left[\frac{-i\left(g_{\mu\nu} - \frac{p_\mu p_\nu}{m^2}\right)}{p^2 - m^2 + i\epsilon} \right] \quad [p_0 - \text{arbitrary}] \tag{24}$$

or by inverting the quadratic form of the (free) Lagrangian

$$\mathcal{L} = -\frac{1}{2} W^\dagger_{\mu\nu} W_{\mu\nu} + m^2 W^\dagger_\mu W_\mu, \quad W_{\mu\nu} \equiv \partial_\mu W_\nu - \partial_\nu W_\mu.$$

One can show that one of the polarization vectors $\epsilon^L_\mu \simeq p_\mu/m + \mathcal{O}(m)$ and *diverges* in the limit $p_\mu \to \infty$ $(m \to 0)$. This indicates that one should be careful when constructing models with massive vector fields. We will return to this issue later.

2.3.2 Massless vector fields

Massless (say photon) vectors are usually represented by

$$A_\mu(x) = \frac{1}{(2\pi)^{3/2}} \sum_{\lambda=0}^{3} \int \frac{d\mathbf{p}}{\sqrt{2\omega_p}} \left[\epsilon^\lambda_\mu(\mathbf{p}) a^-_\lambda(\mathbf{p}) e^{-ipx} + \text{h.c.} \right]. \tag{25}$$

with

$$\epsilon^\lambda_\mu(\mathbf{p}) \epsilon^{*\lambda'}_\mu(\mathbf{p}) = g^{\lambda\lambda'}, \quad \epsilon^\lambda_\mu(\mathbf{p}) \epsilon^{*\lambda}_\nu(\mathbf{p}) = g_{\mu\nu}, \quad [a^-_\lambda(\mathbf{p}), a^+_{\lambda'}(\mathbf{p}')] = -g_{\lambda\lambda'} \delta_{\mathbf{p},\mathbf{p}'}.$$

The corresponding Feynman propagator can be given by

$$\langle 0 | T(A_\mu(x) A_\nu(y)) | 0 \rangle = \frac{1}{(2\pi)^4} \int d^4 p e^{-ip(x-y)} \left[\frac{-i g_{\mu\nu}}{p^2 + i\epsilon} \right]. \tag{26}$$

In spite of the fact that we sum over four polarizations in Eq. (25) only *two* of them are *physical*! This reflects the fact that the vector-field Lagrangian in the massless case $m = 0$

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F_{\mu\nu}, \quad F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$$

is invariant under $A_\mu \to A_\mu + \partial_\mu \alpha(x)$ for arbitrary $\alpha(x)$ (*gauge* symmetry). Additional *conditions* (gauge-fixing) are needed to get rid of unphysical states.

2.3.3 Fermion fields

Spin-1/2 fermion fields (e.g., leptons) are represented by[5]

$$\psi^\alpha(x) = \frac{1}{(2\pi)^{3/2}} \int \frac{d\mathbf{p}}{\sqrt{2\omega_p}} \sum_{s=1,2} \left[u^\alpha_s(\mathbf{p}) a^-_s(\mathbf{p}) e^{-ipx} + v^\alpha_s(\mathbf{p}) b^+_s(\mathbf{p}) e^{+ipx} \right],$$

where we explicitly write the *spinor* (Dirac) index α for u_s, v_s and the quantum operator ψ. The former satisfy the 4×4 matrix (Dirac) equations

$$(\hat{p} - m) u_s(\mathbf{p}) = 0, \quad (\hat{p} + m) v_s(\mathbf{p}) = 0, \quad \hat{p} \equiv \gamma_\mu p_\mu, \quad p_0 \equiv \omega_{\mathbf{p}} \tag{27}$$

and correspond to particles (u_s) or antiparticles (v_s). In Eq. (27) we use gamma-matrices

$$\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu \equiv [\gamma_\mu, \gamma_\nu]_+ = 2 g_{\mu\nu} \mathbf{1} \quad \Rightarrow \gamma_0^2 = \mathbf{1}, \quad \gamma_1^2 = \gamma_2^2 = \gamma_3^2 = -\mathbf{1}$$

to account for *two* spin states ($s = 1, 2$) of particles and antiparticles. Fermion fields transform under the Lorentz group $x' = \Lambda x$ as (*cf.* Eq. (20))

$$\psi'(x') = \mathcal{S}_\Lambda \psi(x), \qquad \psi'(x')^\dagger = \psi(x) \mathcal{S}^\dagger_\Lambda. \tag{28}$$

[5] There exists a charge-conjugation matrix $C = i\gamma_2$, which relates spinors for particles u and antiparticles v, e.g., $v = Cu^*$.

It turns out that the 4×4 matrix $\mathcal{S}_\Lambda^\dagger \neq \mathcal{S}_\Lambda^{-1}$ but $\mathcal{S}^{-1} = \gamma_0 \mathcal{S}^\dagger \gamma_0$. Due to this, it is convenient to introduce a *Dirac-conjugated* spinor $\bar\psi(x) \equiv \psi^\dagger\gamma_0$. The latter enters into

$$\begin{aligned} \bar\psi'(x')\psi'(x') &= \bar\psi(x)\psi(x), && \text{Lorentz } \textit{scalar}; \\ \bar\psi'(x')\gamma_\mu\psi'(x') &= \Lambda_{\mu\nu}\bar\psi(x)\gamma_\nu\psi(x), && \text{Lorentz } \textit{vector}. \end{aligned}$$

This allows us to convince ourselves that the Dirac Lagrangian

$$\mathcal{L} = \bar\psi\left(i\hat\partial - m\right)\psi$$

is also a Lorentz scalar, i.e., respects Lorentz symmetry. Dirac-conjugated spinors are used to impose Lorentz-invariant normalization on u and v:

$$\bar u_s(\mathbf{p})u_r(\mathbf{p}) = 2m\delta_{rs}, \qquad \bar v_s(\mathbf{p})v_r(\mathbf{p}) = -2m\delta_{rs},$$

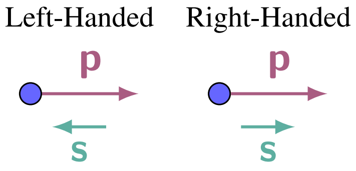
An important fact about quantum fermion fields is that, contrary to the case of scalar or vector (*boson*) fields, the creation/annihilation operators for fermions $a^\pm_{s,\mathbf{p}}$ and antifermions $b^\pm_{s,\mathbf{p}}$ *anticommute*:

$$\begin{aligned} \left[a^-_{r,\mathbf{p}}, a^+_{s,\mathbf{p}'}\right]_+ &= \left[b^-_{r,\mathbf{p}}, b^+_{s,\mathbf{p}'}\right]_+ = \delta_{sr}\delta(\mathbf{p}-\mathbf{p}') \\ \left[a^\pm_{r,\mathbf{p}}, a^\pm_{s,\mathbf{p}'}\right]_+ &= \left[b^\pm_{r,\mathbf{p}}, b^\pm_{s,\mathbf{p}'}\right]_+ = \left[a^\pm_{r,\mathbf{p}}, b^\pm_{s,\mathbf{p}'}\right]_+ = 0. \end{aligned}$$

Due to this, fermions obey the *Pauli principle*, e.g., $a^+_{r,\mathbf{p}}a^+_{r,\mathbf{p}} = 0$. Moreover, one can explicitly show that quantization of bosons (integer spin) with anticommutators or fermions (half-integer spin) with commutators leads to inconsistencies (violates the *Spin-Statistics* theorem).

Let us emphasize an important difference between the notions of *chirality* and *helicity*. Two independent solutions for *massive* fermions ($u_{1,2}$) can be chosen to correspond to two different *helicities* — projections of spin vector $\mathbf{s}$ onto direction of $\mathbf{p}$:

$$\mathcal{H} = \mathbf{s}\cdot\mathbf{n}, \quad \mathbf{n} = \mathbf{p}/|\mathbf{p}|. \tag{29}$$



In *free* motion it is *conserved* and serves as a good quantum number. However, it is not a Lorentz-invariant quantity. Indeed, we can flip the sign of particle momentum by moving with speed faster than $v = |\mathbf{p}|/p_0$. As a consequence, $\mathbf{n}\to-\mathbf{n}$ and $\mathcal{H}\to-\mathcal{H}$. However, *helicity* for a *massless* particle is the same for all inertial observers and coincides with *chirality*, which is a *Lorentz-invariant* concept.

By definition Left (ψ_L) and Right (ψ_R) *chiral* spinors are eigenvectors of

$$\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3 \Rightarrow [\gamma_\mu,\gamma_5]_+ = 0, \quad \gamma_5^2 = 1, \quad \gamma_5^\dagger = \gamma_5, \tag{30}$$

where

$$\gamma_5\psi_L = -\psi_L, \qquad \gamma_5\psi_R = +\psi_R. \tag{31}$$

Any spinor ψ can be decomposed as

$$\psi = \psi_L + \psi_R, \qquad \psi_{L/R} = P_{L/R}\psi, \quad P_{L/R} = \frac{1\mp\gamma_5}{2}. \tag{32}$$

To illustrate this fact, let us use the Dirac representation of 4×4 γ-matrices:

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}, \quad \gamma^5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad P_{L/R} = \frac{1}{2}\begin{pmatrix} 1 & \mp 1 \\ \mp 1 & 1 \end{pmatrix}$$

to re-write the spinors for particles u_λ and antiparticles v_λ of positive $\lambda = +1$ and negative $\lambda = -1$ helicities as a sum of left and right *chiral* parts ($\beta^+ = \sqrt{E+m}$):

$$u_\lambda(\mathbf{p}) = \frac{1}{2}\beta_+ \left(1 - \lambda\frac{p}{E+m}\right)\begin{pmatrix} \chi_\lambda \\ -\chi_\lambda \end{pmatrix} + \frac{1}{2}\beta^+ \left(1 + \lambda\frac{p}{E+m}\right)\begin{pmatrix} \chi_\lambda \\ \chi_\lambda \end{pmatrix}, \tag{33}$$

$$v_\lambda(\mathbf{p}) = \frac{1}{2}\beta^+ \left(1 - \lambda\frac{p}{E+m}\right)\begin{pmatrix} \chi_{-\lambda} \\ \chi_{-\lambda} \end{pmatrix} + \frac{1}{2}\beta^+ \left(1 + \lambda\frac{p}{E+m}\right)\begin{pmatrix} -\chi_{-\lambda} \\ \chi_{-\lambda} \end{pmatrix} \tag{34}$$

with $\mathbf{p} = p(\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)$ and

$$\chi_1 = \begin{pmatrix} \cos\frac{\theta}{2} \\ e^{i\phi}\sin\frac{\theta}{2} \end{pmatrix}, \; \chi_{-1} = \begin{pmatrix} -e^{-i\phi}\sin\frac{\theta}{2} \\ \cos\frac{\theta}{2} \end{pmatrix}.$$

One can easily see that in the massless case[6]

$P_L u_+ = 0, \quad P_R u_- = 0,$ for *particle*, the *spinor* chirality *coincides* with helicity,
$P_L v_- = 0, \quad P_R v_+ = 0,$ for *antiparticle*, the *spinor* chirality is *opposite* to helicity.

Moreover, we can rewrite the Dirac Lagrangian it terms of chiral components (Weyl spinors)

$$\mathcal{L} = i(\underbrace{\bar\psi_L\hat\partial\psi_L + \bar\psi_R\hat\partial\psi_R}_{\text{conserve chirality}}) - m(\underbrace{\bar\psi_L\psi_R + \bar\psi_R\psi_L}_{\text{break chirality}}), \tag{35}$$

and see that, indeed, it is the mass term that mixes two chiralities. Due to this, it violates *chiral* symmetry corresponding to the independent rotation of left and right components

$$\psi \to e^{i\gamma_5\alpha}\psi. \tag{36}$$

Consequently, if we drop the mass term, the symmetry of the Lagrangian is enhanced.

Up to now we were discussing the so-called Dirac mass term. For *neutral* fermions (e.g., neutrino) there is another possibility — a *Majorana* mass. Since charge-conjugation applied to fermion fields, $\psi \to \psi^c$, *flips* chirality, we can use ψ_L^c in place of ψ_R to write

$$\mathcal{L} = \frac{1}{2}(i\bar\psi_L\hat\partial\psi_L - m\bar\psi_L\psi_L^c). \tag{37}$$

As a consequence, to describe Majorana particles, we only need two components instead of four since antiparticles coincide with particles in this case. At the moment, the nature of neutrinos is unclear, and we refer to Ref. [20] for more elaborated discussion.

2.4 From free to interacting fields

Fields that describe non-interacting particles seems to be an abstraction. Nevertheless, all we have an intuition that in many cases we can neglect all (or some) interactions and treat real particles as free. Indeed, in HEP, a typical collision/scattering experiment deals with *"free"* initial and final states and

[6] One can also define *chirality of an antiparticle*, which is opposite to that of the corresponding *spinor* v, i.e., introduce v_R, such as $P_L v_R = v_R$, but $P_R v_R = 0$. In this case, for $m \to 0$ the chirality precisely corresponds to helicity and $v_R \to v_+$, etc.

considers *transitions* between these states due to *interactions*. To account for this in a quantum theory, one introduces the S-*matrix* with matrix elements

$$\mathcal{M} = \langle\beta|S|\alpha\rangle, \qquad \mathcal{M} = \delta_{\alpha\beta} + (2\pi)^4\delta^4(p_\alpha - p_\beta)iM_{\alpha\beta} \tag{38}$$

corresponding to the amplitudes for possible transitions between *in* $|\alpha\rangle$ and *out* $|\beta\rangle$ states:

$$|\alpha\rangle = \tilde{a}^+_{\mathbf{p}_1}...\tilde{a}^+_{\mathbf{p}_r}|0\rangle, \quad |\beta\rangle = \tilde{a}^+_{\mathbf{k}_1}...\tilde{a}^+_{\mathbf{k}_s}|0\rangle, \quad \tilde{a}^+_{\mathbf{p}} = (2\pi)^{3/2}\sqrt{2\omega_{\mathbf{p}}}a^+_{\mathbf{p}}, \tag{39}$$

where for convenience (see Eq. (50)) we re-scale our creation/annihilation operators. Given the matrix element $M_{\alpha\beta}$, one can calculate the differential probability (per unit volume per unit time) to evolve from $|\alpha\rangle$ to $|\beta\rangle$:

$$dw = \frac{n_1...n_r}{(2\omega_{p_1})...(2\omega_{p_r})}|M_{\alpha\beta}|^2 d\Phi_s. \tag{40}$$

Here n_i correspond to initial-state particle densities, and an element of phase space is given by

$$d\Phi_s = (2\pi)^4\delta^4\left(p_{in} - k_{out}\right)\frac{d\mathbf{k}_1}{(2\pi)^3(2\omega_{k_1})}\cdots\frac{d\mathbf{k}_i}{(2\pi)^3(2\omega_{k_i})} \tag{41}$$

with $p_{in} = \sum p_i$ and $k_{out} = \sum k_i$. Since we are usually interested in processes involving one (decay) or two particles (e.g., collision of two beams) in the initial state, it is more convenient to consider the differential decay width $d\Gamma$ in the rest frame of a particle with mass m, or cross-section $d\sigma$ of a process $2 \to s$:

$$d\Gamma = \Phi_\Gamma|M_{1\to s}|^2 d\Phi_s, \qquad \Phi_\Gamma = \frac{1}{2m}, \tag{42}$$

$$d\sigma = \Phi_\sigma|M|^2 d\Phi_s, \qquad \Phi_\sigma = \frac{1}{4\sqrt{(p_1p_2)^2 - p_1^2p_2^2}} \quad . \tag{43}$$

In Eq. (43) the factor Φ_σ is *Lorentz-invariant* and is expressed in terms of four-momenta of initial particles p_1 and p_2. The total width Γ and total cross-section σ can be obtained by integration over the momenta of final particles restricted by energy-momentum conservation due to the four-dimensional δ-function in Eq. (41).

In QFT, the S-matrix is written in terms of the time-ordered exponent

$$S = Te^{-i\int d^4x\mathcal{H}_I(x)} = Te^{i\int d^4x\mathcal{L}_I(x)}, \tag{44}$$

which involve the interaction Hamiltonian $\mathcal{H}_I$ (Lagrangian $\mathcal{L}_I$).

The interaction Lagrangian $\mathcal{L}_I = \mathcal{L}_{full} - \mathcal{L}_0$ is a sum of *Lorentz-invariant* terms having more than *two* fields and more ∂_μ than in the quadratic part $\mathcal{L}_0$, which, if considered alone, describes free particles. It is worth noting that in Eq. (44) we treat $\mathcal{L}_I$ ($\mathcal{H}_I$) as an operator built from *free*[7] quantum fields (i.e., certain combinations of $a^\pm$ and $b^\pm$).

The *time-ordering* operation, which was used to define particle propagators, is generalized in Eq. (44) to account for more than two fields originating from $\mathcal{L}_I$

$$T\Phi_1(x_1)...\Phi_n(x_n) = (-1)^k\Phi_{i_1}(x_{i_1})...\Phi_{i_n}(x_{i_n}), \qquad x^0_{i_1} > ... > x^0_{i_n}. \tag{45}$$

The factor $(-1)^k$ appears due to k possible permutations of *fermion* fields.

To conserve probability the (interaction) Lagrangian should be hermitian. Any scalar combination of quantum fields can, in principle, be included in $\mathcal{L}_I$, e.g.,

$$\mathcal{L}_I: \quad g\phi^3(x), \qquad \lambda\phi^4(x), \qquad y\bar{\psi}(x)\psi(x)\phi(x)$$

[7] More precisely, operators in the *interaction/Dirac* picture.

$$e\bar{\psi}(x)\gamma_\mu\psi(x)A_\mu(x), \qquad G\left[(\bar{\psi}_1\gamma_\mu\psi_2)\,(\bar{\psi}_3\gamma_\mu\psi_4) + \text{h.c.}\right] \quad .$$

The parameters (couplings) g, λ, e, y, and G are related to the "*strength*" of the interactions. An important property of any coupling in the QFT model is its (mass) *dimension*. The latter can be deduced from the fact that in the natural units the Action is dimensionless and $[\mathcal{L}] = 4$. One can notice that all the couplings (hidden) in the T-shirt Lagrangian are *dimensionless*. As it will be clear from subsequent discussion, this has crucial consequences for the self-consistency of the SM model.

2.5 Perturbation theory

In an interacting theory it is very hard, if not impossible, to calculate the S-matrix, see Eq. (44), exactly. Usually, we make an assumption that the couplings in $\mathcal{L}_I$ are small allowing us to treat the terms in $\mathcal{L}_I$ as *perturbations* to $\mathcal{L}_0$. As a consequence, we expand the T-exponent and restrict ourselves to a finite number of terms. In the simplest case of $\mathcal{L}_I = -\lambda\phi^4/4!$ we have at the nth order

$$\frac{i^n}{n!}\left[\frac{\lambda}{4!}\right]^n \langle 0|\tilde{a}^-_{\mathbf{k}_1}...\tilde{a}^-_{\mathbf{k}_s} \int dx_1...dx_n T\left[\phi(x_1)^4...\phi(x_n)^4\right] \tilde{a}^+_{\mathbf{p}_1}...\tilde{a}^+_{\mathbf{p}_r}|0\rangle. \tag{46}$$

To proceed, one utilizes the *Wick* theorem:

$$T\Phi_1...\Phi_n = \sum(-1)^\sigma \langle 0|T(\Phi_{i_1}\Phi_{i_2})|0\rangle...\langle 0|T(\Phi_{i_{m-1}}\Phi_{i_m})|0\rangle :\Phi_{i_{m+1}}...\Phi_{i_n}:, \tag{47}$$

where the sum goes over all possible ways to pair the fields. The Wick theorem Eq. (47) expresses *time-ordered* products of fields in terms of *normal-ordered* ones and propagators. As it was mentioned earlier the normal-ordered operation puts *all* annihilation operators originating from different Φs to the right. It also cares about fermions, e.g.,

$$:a_1^- a_2^+ a_3^- a_4^- a_5^+ a_6^- := (-1)^\sigma a_2^+ a_5^+ a_1^- a_3^- a_4^- a_6^-\,, \tag{48}$$

with σ corresponding to the number of fermion permutations (*cf.* Eq. (45)). In Fig. 2 a cartoon, which illustrates Eq. (47) for one of the contributions to $T[\mathcal{L}_I(x)\mathcal{L}_I(y)]$, is provided.

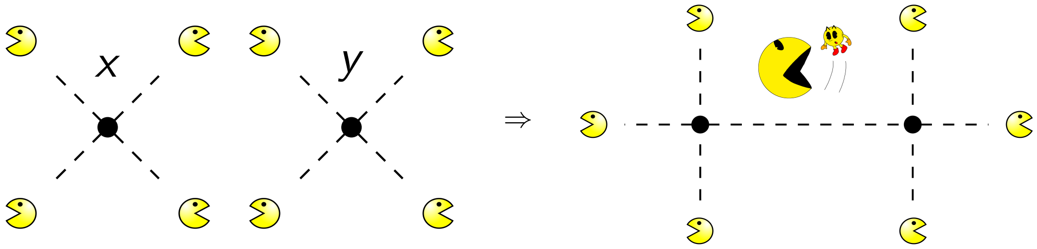


Fig. 2: The Wick theorem at work: one of the contributions.

After application of the Wick theorem we are left with the matrix elements of the form

$$\langle 0|\tilde{a}^-_{\mathbf{k}_1}...\tilde{a}^-_{\mathbf{k}_s} :\Phi_{i_{m+1}}...\Phi_{i_n}: \tilde{a}^+_{\mathbf{p}_1}...\tilde{a}^+_{\mathbf{p}_r}|0\rangle. \tag{49}$$

To get a *non-zero* result, all $a^-(a^+)$ in the normal product of fields from the Lagrangian have to be "killed" by (commuted with) $a^+(a^-)$ from the external states. For our *generalized* field, Eq. (22), we have

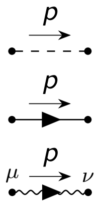
$$\left[\Phi^i_\alpha(x), (a^+_\mathbf{p})^i_s\right] = \underbrace{\frac{e^{-ipx}}{(2\pi)^{3/2}\sqrt{2\omega_p}}}_{\text{common to all fields}} u^s_\alpha(\mathbf{p}), \qquad \text{initial state polarization (particle);}$$

$$\left[(b^-_{\mathbf{p}})^i_s, \Phi^i_\alpha(x)\right] = \frac{e^{+ipx}}{(2\pi)^{3/2}\sqrt{2\omega_p}} v^{*s}_\alpha(\mathbf{p}), \qquad \text{final state polarization (antiparticle).} \tag{50}$$

and one clearly sees that the factors in the denominators, Eq. (50), are avoided when the re-scaled $\tilde{a}^\pm$ (or $\tilde{b}^\pm$) operators, Eq. (39), are used.

All this machinery can be implemented in a set of *Feynman rules*, which are used to draw (and evaluate) *Feynman diagrams*. Every Feynman diagram involves *vertices*, *external* and *internal* lines. Internal lines connect two vertices and correspond to propagators. The expression for propagators can be derived from $\mathcal{L}_0$, e.g.,

$$\left.\begin{array}{r} \langle 0|T(\phi(x)\phi^\dagger(y)|0\rangle \\ \langle 0|T(\psi(x)\bar{\psi}(y)|0\rangle \\ \langle 0|T(W_\mu(x)W^\dagger_\nu(y)|0\rangle \end{array}\right\} = \int \frac{d^4p}{(2\pi)^4}\frac{ie^{-ip(x-y)}}{p^2 - m^2 + i\epsilon}\left\{\begin{array}{lr} 1 & \phi; \\ \hat{p} + m & \psi; \\ -g_{\mu\nu} + p_\mu p_\nu/m^2 & W_\mu. \end{array}\right. \tag{51}$$



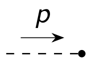
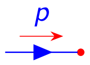
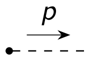
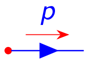
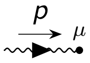
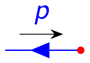
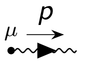
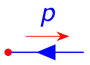
One can notice that all the dependence on x_i of the integrand in Eq. (46) comes from either Eq. (50) or Eq. (51). As a consequence, it is possible to carry out the integration for *every* x_i

$$\int d^4x_i e^{-ix_i(p_1 + \ldots + p_n)} = (2\pi)^4\delta^4(p_1 + \ldots + p_n) \tag{52}$$

and obtain a δ-function reflecting energy-momentum conservation at the corresponding vertex.

Depending on the direction of momenta, the external lines represent incoming or outgoing particles (see Table 1). Again, the corresponding factors (="polarization vectors") are derived from $\mathcal{L}_0$. Notice that we explicitly write the Lorentz indices for vector particles and suppress the Dirac indices for fermions. To keep track of the index contractions in the latter case, one uses *arrows* on the fermion lines.[8]

Table 1: Feynman rules for external states.

incoming scalar	1		*incoming* fermion	$u_s(\mathbf{p})$	
outgoing scalar	1		*outgoing* fermion	$\bar{u}_s(\mathbf{p})$	
incoming vector	$\epsilon^\lambda_\mu(\mathbf{p})$		*incoming* antifermion	$\bar{v}_s(\mathbf{p})$	
outgoing vector	$\epsilon^{*\lambda}_\mu(\mathbf{p})$		*outgoing* antifermion	$v_s(\mathbf{p})$	

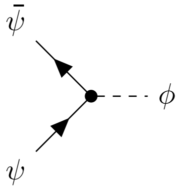
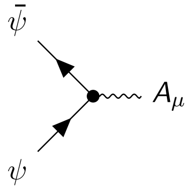
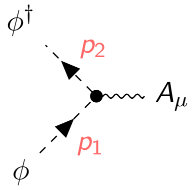
Let us turn to interaction vertices. The corresponding Feynman rules are derived from $\mathcal{A}_I = \int d^4\mathcal{L}_I$. It is convenient to do this by carrying out a Fourier transform to "convert" coordinate derivatives to momenta and considering variations of the action. In the case of $\mathcal{L}_I = -\lambda\phi^4/4!$ we have (all momenta are assumed to be incoming)

$$i\frac{\delta^4\mathcal{A}_I[\phi]}{\delta\phi(p_1)\delta\phi(p_2)\delta\phi(p_3)\delta\phi(p_4)}\bigg|_{\phi=0} \Rightarrow \underbrace{(2\pi)^4\delta^4(p_1 + p_2 + p_3 + p_4)}_{\text{conservation of energy-momentum}} \times [-i\lambda]. \tag{53}$$

[8] There are subtleties when interactions involve Majorana fermions.

In a typical diagram all $(2\pi)^4\delta(...)$ factors (but *one*, which is accounted for in the definition, Eq. (38), of $M_{\alpha\beta}$) that reflect the energy-momentum conservation at each vertex, are removed by the momentum integration originating from propagators, Eq. (51). Due to this, we also omit these factors (see, Table 2 for examples).

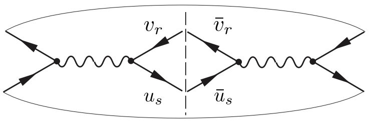
Table 2: Vertex Feynman rules. Derivatives in $\mathcal{L}_I$ correspond to particle momenta.

$\mathcal{L}_I = -y\bar{\psi}\psi\phi$	$\mathcal{L}_I = e\bar{\psi}\gamma_\mu\psi A_\mu$	$\mathcal{L}_I = ieA_\mu\left(\phi^\dagger\partial_\mu\phi - \phi\partial_\mu\phi^\dagger\right)$
		
$-iy$	$ie\gamma_\mu$	$ie(p_1+p_2)_\mu$

Given Feynman rules, one can draw all possible diagrams that contribute to a process and evaluate the amplitude. We do not provide the precise prescription here (see Refs. [14–18] for details) but just mention the fact that one should keep in mind various *symmetry* factors and relative *signs* that can appear in real calculations.

In order to get probabilities, we have to *square* matrix elements, e.g.,

$$|M|^2 = MM^\dagger \Rightarrow \tag{54}$$



Sometimes we do not care about polarization states of initial or final particles. As a consequence, we have to *sum* the probabilities corresponding to different *final* polarizations, and *average* over the *initial* ones. That is where *spin-sum* formulas,e.g.,

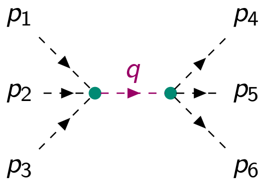
$$\sum_s u_s(\mathbf{p_1})\bar{u}_s(\mathbf{p_1}) = \hat{p}_1 + m, \qquad \sum_s v_s(\mathbf{p_2})\bar{v}_s(\mathbf{p_2}) = \hat{p}_2 - m \tag{55}$$

become useful

$$MM^\dagger \to \sum_{s,r}(\bar{u}_s A v_r)(\bar{v}_r A^\dagger u_s) = \mathrm{Tr}\left[(\hat{p}_1+m)A(\hat{p}_2-m)A^\dagger\right]. \tag{56}$$

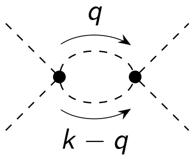
In this case we avoid explicit manipulations with spinors and utilize well-known and efficient machinery for gamma-matrix traces.

Let us continue by mentioning that only in *tree* graphs, such as



$$\Rightarrow \quad (2\pi)^4\delta^4\left(\sum_{i=1}^3 p_i - \sum_{i=4}^6 p_i\right)[-i\lambda]^2\,\frac{i}{q^2-m^2+i\epsilon},$$

all the integrations (due to propagators) are “killed” by vertex δ-functions. However, nothing forbids us from forming *loops*. In this case, we have *integrals* over unconstrained momenta, e.g., in the ϕ^4-theory



$$: \quad I_2(k) \equiv \int \frac{d^4 q}{[q^2+i\epsilon][(k-q)^2+i\epsilon]} \sim \int^\infty \frac{|q|^3 d|q|}{|q|^4} \sim \ln\infty,$$

which can lead to *divergent* (meaningless?) results. This is a manifestation of the ultraviolet (or *UV*) divergences due to *large* momenta ("small distances").

A natural question arises: Do we have to abandon QFT? Since we still use it, there are reasons *not* to do this. Indeed, we actually do not know physics up to infinitely small scales and our extrapolation can not be adequate in this case. To make sense of the integrals, we can *regularize* them, e.g., introduce a *"cut-off"* $|q| < \Lambda$,

$$I_2^\Lambda(k) = i\pi^2\left[\ln\frac{\Lambda^2}{k^2}+1\right]+\mathcal{O}\left(\frac{k^2}{\Lambda^2}\right) = i\pi^2\left[\ln\frac{\Lambda^2}{\mu^2}-\ln\frac{k^2}{\mu^2}+1\right]+\mathcal{O}\left(\frac{k^2}{\Lambda^2}\right) \tag{57}$$

or use another convenient possibility — *dimensional* regularization, when $d = 4$ space-time is formally continued to $d = 4 - 2\varepsilon$ dimensions:

$$I_2^{4-2\varepsilon}(k) = \mu^{2\varepsilon}\int \frac{d^{4-2\varepsilon}q}{q^2(k-q)^2} = i\pi^2\left(\frac{1}{\varepsilon}-\ln\frac{k^2}{\mu^2}+2\right)+\mathcal{O}(\varepsilon). \tag{58}$$

Both the regularized integrals are now convergent[9] and share the same logarithmic dependence on external momentum k. One can also notice a new (renormalization) scale μ, which appears in regularized integrals, and a (one-to-one) correspondence between a *logarithmically* divergent contribution $\log\Lambda^2/\mu^2$ in Eq. (57) and the pole term $1/\varepsilon$ in Eq. (58). However, the constant terms are *different*. How do we make sense of this ambiguity?

The crucial observation here is that the divergent pieces, which blow up when we try to remove the regulators ($\Lambda \to \infty$ or $\varepsilon \to 0$), are *local*, i.e., depend polynomially on external kinematic variables. This fact allows us to *cancel* them by the so-called counterterm (CT) vertices. The latter can be interpreted as new terms in $\mathcal{L}_I$. Moreover, in a *renormalizable* QFT model additional (divergent) contributions have the same form as the initial Lagrangian and thus can be "absorbed' into redefinition of fields and parameters.

One can revert the reasoning and assume that the initial Lagrangian is written in terms of the so-called *bare* (unobservable) quantities. The predictions of the model are finite since the explicit dependence of Feynman integrals on the cut-off Λ (or ε) is actually compensated by the implicit dependence of bare fields and parameters. In some sense the latter quantities represent our ignorance of dynamics at tiny scales: physical fields and parameters are always "dressed" by clouds of virtual particles.

It is obvious that working with *bare* quantities is not very convenient. One usually makes the dependence on Λ (or ε) explicit by introducing divergent Z-factors for *bare* fields (ϕ_B), masses (m_B^2), and couplings (λ_B), e.g.,

$$\mathcal{L}_{full} = \frac{1}{2}(\partial\phi_B)^2 - \frac{m_B^2}{2}\phi_B^2 + \frac{\lambda_B\phi_B^4}{4!} = \frac{Z_2}{2}(\partial\phi)^2 - \frac{Z_m m^2}{2}Z_2\phi^2 + \frac{Z_\lambda\lambda}{4!}(Z_2\phi^2)^2 \tag{59}$$

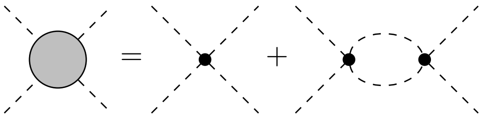
$$= \frac{(\partial\phi)^2}{2} - \frac{m^2\phi^2}{2} + \frac{\lambda\phi^4}{4!} + \underbrace{\frac{(Z_2-1)}{2}(\partial\phi)^2 - \frac{(Z_m Z_2 - 1)m^2}{2}\phi^2 + (Z_4 Z_2^2 - 1)\frac{\lambda\phi^4}{4!}}_{\text{counterterms}}. \tag{60}$$

Here ϕ, m and λ denote *renormalized* (finite) quantities. Since we can always subtract something finite from infinity, there is a certain freedom in this procedure. The different constant terms in Eq. (57) and

[9] We do not discuss the issue of possible infrared (IR) divergences here.

Eq. (58) are just a manifestation of this fact. So we have to impose additional *conditions* on Z, i.e., define a *renormalization* scheme. For example, in the minimal (MS) schemes we subtract only the divergent terms, e.g., only poles in ε, while in the so-called momentum-subtraction (MOM) schemes we require certain amplitudes (more generally *Green functions*) to have specific values at some fixed kinematics.

As an illustration, let us consider a scattering amplitude $2 \to 2$ in the ϕ^4 model calculated in perturbation theory:



$$+ \text{ permutations} + \text{more loops} \qquad (61)$$

$$= \lambda_B(\Lambda) - \frac{\lambda_B(\Lambda)^2}{2(16\pi^2)}\left(\ln\frac{\Lambda^2}{\mu^2} - \ln\frac{k^2}{\mu^2} + \ldots\right) + \ldots \qquad (62)$$

$$= \left[\lambda(\mu) + \frac{3}{2}\frac{\lambda^2(\mu)}{16\pi^2}\ln\frac{\Lambda^2}{\mu^2}\right] - \frac{\lambda(\mu)^2}{2(16\pi^2)}\left(\ln\frac{\Lambda^2}{\mu^2} - \ln\frac{k^2}{\mu^2} + \ldots\right) + \ldots \qquad (63)$$

$$= \lambda(\mu) + \frac{\lambda(\mu)^2}{2(16\pi^2)}\left(\ln\frac{k^2}{\mu^2} + \ldots\right) + \ldots. \qquad (64)$$

In Eq. (61) the tree-level and one-loop diagrams contributing to the matrix element are presented. The corresponding expression in terms of the bare coupling $\lambda_B(\Lambda)$ that implicitly depends on the regularization parameter Λ is given in Eq. (62). We introduce a renormalized[10] coupling $\lambda(\mu)$ in Eq. (63) to make the dependence explicit:

$$\lambda_B(\Lambda) = \lambda(\mu) Z_\lambda = \lambda(\mu)\left(1 + \frac{3}{2}\frac{\lambda(\mu)}{16\pi^2}\ln\frac{\Lambda^2}{\mu^2} + ...\right). \qquad (65)$$

The final result, Eq. (64), is finite (when $\Lambda \to \infty$) and can be confronted with experiment. It seems to depend on the auxiliary scale μ. The crucial point here is that *observables* (if all orders of PT are taken into account) actually do *not* depend on μ. Changing μ corresponds to a certain reshuffling of the PT series: some terms from loop corrections are absorbed into the re-scaled (*running*) couplings. This allows one to improve the "convergence"[11] of the finite-order result by a convenient choice of μ.

The scale-dependence of the *running* couplings is governed by renormalization-group equations (RGE). In the considered case we have

$$\lambda(\mu_0) \to \lambda(\mu), \qquad \frac{d}{d\ln\mu}\lambda = \beta_\lambda(\lambda), \qquad \beta_\lambda = \frac{3}{2}\frac{\lambda^2}{16\pi^2} + ... \quad . \qquad (66)$$

The *beta-function* β_λ can be calculated order-by-order in PT. However, the (initial) value $\lambda(\mu_0)$ needed to solve Eq. (66) is *not predicted* and has to be extracted from experiment ("measured").

It is worth pointing out here that two different numerical values of the *renormalized* self-coupling, λ_1 and λ_2, do not necessarily correspond to different physics. Indeed, if they are fitted from measurements at different scales, e.g., μ_0 and μ, and are related by means of RGE, they represent the *same* physics (see Fig. 3).

2.6 Renormalizable or non-renormalizable?

Let us stress again that the model is called *renormalizable* if *all* the UV divergences that appear in loop integrals can be canceled by local counterterms due to renormalization of bare parameters and couplings of the *initial* Lagrangian $\mathcal{L}_{full}$. But what happens if there is a UV divergent amplitude but the structure

[10] We use minimal subtractions here and the factor of three comes from the fact that all three one-loop graphs (s, t and u) give rise to the same *divergent* term.

[11] Actually, the PT series are *asymptotic* (divergent) and we speak about the behavior of a limited number of first terms here.

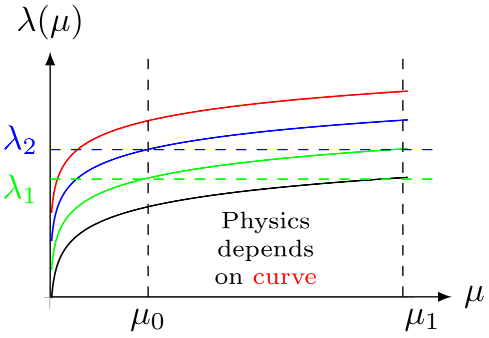


Fig. 3: Solutions of RGE for different boundary conditions.

of the required subtraction does not have a counter-part in $\mathcal{L}_{full}$, i.e., we do not have a coupling to absorb the infinity? Obviously, we can modify $\mathcal{L}_{full}$ and *add* the required term (and the coupling).

An example of such a situation can be found in the model with a scalar ϕ (e.g., Higgs) coupled to a fermion ψ (e.g., top quark) via the Yukawa interaction characterized by the coupling y

$$\mathcal{L}_I \ni \delta\mathcal{L}_Y = -y \cdot \bar{\psi}\psi\phi. \tag{67}$$

Let us assume for the moment that we set the self-coupling to zero $\lambda = 0$ and want to calculate the Higgs-scattering amplitude due to virtual top quarks (see, Fig. 4). We immediately realize that the contribution is divergent and without $\delta\mathcal{L}_4 = -\lambda\phi^4/4!$ we are not able to make sense of our model. Due to this, we are forced to consider the ϕ^4 term in a consistent theory.

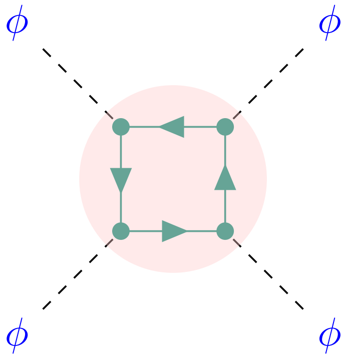


Fig. 4: One-loop correction to higgs self-interaction.

Since we modified $\mathcal{L}_{full}$, we have to re-calculate all the amplitudes form scratch. In principle, new terms in $\mathcal{L}_I$ will generate new diagrams, which can require new interactions to be added to $\mathcal{L}_I$. Will this process terminate? In the case of *renormalizable* models the answer is positive. We just need to make sure that $\mathcal{L}_I$ include *all* possible terms with *dimensionless* couplings[12], or, *equivalently*, local dimension-4 *operators* built from quantum fields and their derivatives.

On the contrary, if we have an avalanche of new terms with increasing dimensions, this is a signal of a *non-renormalizable* model. It looks like that we have to abandon such models since we need to measure an infinite number of couplings to predict something in this situation! However, it should be stressed that non-renormalizable models, contrary to renormalizable ones, involve couplings G_i with *negative* mass dimension $[G_i] < 0$! Due to this, not all of them are important at *low* energies, which satisfy

$$G_i E^{-[G_i]} \ll 1. \tag{68}$$

[12] Remember the T-shirt Lagrangian?

This explains the success of the *Fermi model* involving the dimension-6 four-fermion operator

$$-\mathcal{L}_I = G\bar{\Psi}_p\gamma_\rho\Psi_n \cdot \bar{\Psi}_e\gamma_\rho\Psi_\nu + \text{h.c.} \tag{69}$$

in the description of the β-decay $n \to p + e^- + \bar{\nu}_e$. The model turns out to be the harbinger of the modern electroweak theory. Although being non-renormalizable and not self-consistent, it provides us with the important information about the *electroweak* scale. The latter turns out to be related to the *measured* value of the Fermi constant G and corresponds to the scale, at which some new dynamics should appear to cure the inconsistencies.

Let us summarize what we have learned so far. In QFT we describe particles and their interactions by considering an *Action/Lagrangian*. We assume that general Lagrangian $\mathcal{L}$ is

– Lorentz (Poincare) invariant (a sum of Lorentz scalars),
– Local (involve finite number of partial derivatives),
– Real (hermitian) (respects unitarity=conservation of probability)

We split $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_I$ into

- the free part $\mathcal{L}_0$ that determines *Feynman rules* for propagators and polarization vectors,
- the interaction Lagrangian $\mathcal{L}_I$ that gives rise to *Feynman rules* for interaction vertices.

Given Feynman rules we evaluate amplitudes and probabilities in *perturbation theory* (PT). Depending on the *dimension* of the couplings in $\mathcal{L}_I$ we distinguish *renormalizable* (self-consistent) and *non-renormalizable* (not self-consistent yet useful) models. To make sense of predictions of *renormalizable* models we utilize *regularization/renormalization*. The parameters of such models depend on scale and *RG* can be used to relate predictions at different scales. *Non-renormalizable* models, on the contrary, are treated only as low-energy *effective* approximations and give us a hint for a "breakdown" or *"new physics"* scale [19].

In principle, we provide (almost) all the necessary information that allows one, given some $\mathcal{L}$, to carry out *calculations* and confront the model with experiments. We put some important, yet very general, restrictions on $\mathcal{L}$. We can try to construct new models by trial and error, but it always nice to have some guiding principle. It is fair to say that modern physics is built around *symmetries*. Anticipating their role in the construction of the SM, let us consider this topic in more detail.

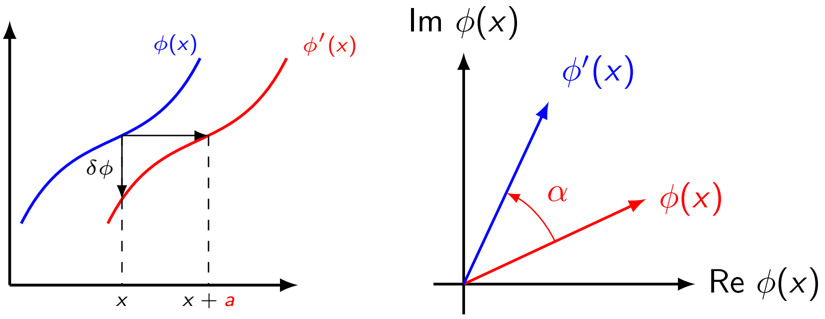
3 An ode to symmetry

The beauty of symmetries and their usefulness in ordinary life are beyond doubt. For example, an architect can design only half of the building (and use mirror symmetry to get the rest), or we can save a lot of time if employ symmetry arguments for cutting paper snowflakes.

Symmetries are intimately connected with *transformations*, which leave something *invariant*. The transformations can be *discrete*, such as (switching back to QFT)

$$\begin{aligned}
\text{Parity}: \phi'(\mathbf{x},t) &= P\phi(\mathbf{x},t) = \phi(-\mathbf{x},t),\\
\text{Time-reversal}: \phi'(\mathbf{x},t) &= T\phi(\mathbf{x},t) = \phi(\mathbf{x},-t),\\
\text{Charge-conjugation}: \phi'(\mathbf{x},t) &= C\phi(\mathbf{x},t) = \phi^\dagger(\mathbf{x},t),
\end{aligned}$$

or depend on *continuous* parameters. One distinguishes *space-time* from *internal* transformations (*cf.* with the distinction between two sets of indices that we attached to our *generalized* field, Eq. (22)). Lorentz boosts, rotations, and translations are typical examples of the former, while phase transformations belong to the latter (see Fig. 5). At the moment, let us consider *global* symmetries with parameters independent of space-time coordinates and postpone the discussion of x-dependent or *local* (*gauge*) transformations to Section 3.2.



$\phi'(x) = \phi(x+a)$ $\qquad$ $\phi'(x) = e^{i\alpha}\phi(x)$

Fig. 5: Translations (left) and phase transformations (right).

3.1 Global symmetries

A convenient way to deal with symmetries in (quantum) field theories is to consider again the Action functional $\mathcal{A}[\phi]$. We can *define* a *symmetry* as *particular* infinitesimal variations $\delta\phi(x)$ that for *any* $\phi(x)$ leave $\mathcal{A}[\phi]$ invariant up to a surface term (*cf.* the Action Principle)

$$\mathcal{A}[\phi'(x)] - \mathcal{A}[\phi(x)] = \int d^4x\, \partial_\mu \mathcal{K}_\mu, \quad \phi'(x) \equiv \phi(x) + \delta\phi(x).$$

If we compare this with the general expression

$$\mathcal{A}[\phi'(x)] - \mathcal{A}[\phi(x)] = \int d^4x \left[\left(\partial_\mu \frac{\partial \mathcal{L}}{\partial \partial_\mu \phi} - \frac{\partial \mathcal{L}}{\partial \phi} \right) \delta\phi + \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \partial_\mu \phi} \delta\phi \right) \right].$$

and require in addition that ϕ *satisfy* EOM[13], we get a *local conservation law*

$$\partial_\mu J_\mu = 0, \quad J_\mu \equiv \mathcal{K}_\mu - \frac{\partial \mathcal{L}}{\partial \partial_\mu \phi} \delta\phi. \tag{70}$$

The integration of Eq. (70) over *space* leads to *conserved* charge:

$$\frac{d}{dt} Q = 0, \qquad Q = \int d\mathbf{x} J_0. \tag{71}$$

If $\delta\phi = \rho_i \delta_i \phi$ depends on parameters ρ_i, we have a conservation law for every ρ_i. This is the essence of the *first Noether theorem* [21].

By means of the Noether theorem we can get almost at no cost the expressions for energy-momentum $P_\mu = (\mathcal{H}, \mathbf{P})$ and charge Q, which we used in Section 2.2. For example, P_μ is nothing else but the conserved "charges", which correspond to space-time translations. Indeed, the Noether current in this case is just the energy-momentum tensor $T_{\mu\nu}$

$$\phi'(x+a) = \phi(x), \qquad \text{expand in } a \Rightarrow \delta\phi(x) = -a_\nu \partial_\nu \phi(x), \tag{72}$$

$$\delta\mathcal{L}(\phi(x), \partial_\mu\phi(x)) = \partial_\nu\left(-a_\nu \mathcal{L}\right) \Rightarrow J_\mu = -a_\mu \mathcal{L} + a_\nu \frac{\partial \mathcal{L}}{\partial \partial_\mu \phi} \partial_\nu \phi = a_\nu T_{\mu\nu}. \tag{73}$$

According to Eq. (71), for every a_μ we have $P_\nu = \int d\mathbf{x} T_{0\nu}$, i.e., conserved total energy-momentum. In the same way, we can apply the Noether theorem to phase transformations of our *complex* field and get

$$\phi'(x) = e^{i\alpha}\phi(x), \quad \delta\phi(x) = \; i\alpha\phi(x), \quad J_\mu = i(\phi^\dagger \partial_\mu \phi - \phi \partial_\mu \phi^\dagger), \quad Q = \int d\mathbf{x} J_0. \tag{74}$$

[13] This requirement is crucial.

The corresponding quantum operators, i.e., $\hat{\mathcal{H}}$ in Eq. (13) or $\hat{Q}$ in Eq. (16), are obtained (modulo ordering issues) from these (classical) expressions by plugging in quantum field $\hat{\phi}$ from Eq. (9).

After quantization the operators corresponding to the *conserved* quantities

– can be used to define a convenient *basis* of states, e.g., we characterize our particle states by eigenvalues of P_μ, and Q:

$$|\mathbf{p}\rangle \equiv |\mathbf{p}, +1\rangle,\ |\bar{\mathbf{p}}\rangle \equiv |\mathbf{p}, -1\rangle \Rightarrow \hat{Q}|\mathbf{p}, q\rangle = q|\mathbf{p}, q\rangle, \hat{\mathbf{P}}|\mathbf{p}, q\rangle = \mathbf{p}|\mathbf{p}, q\rangle. \tag{75}$$

– act as *generators* of symmetries, e.g., for space-time translations we have a *unitary* operator $U(a)$

$$U(a) = \exp\left(i\hat{P}_\mu a_\mu\right), \tag{76}$$

which guarantees that *transition* probabilities between states do not change upon translations. In addition, classical relations between initial and transformed fields become *constraints* on quantum fields, e.g.,

$$\phi'(x+a) = \phi(x) \Rightarrow \hat{\phi}(x+a) = U(a)\hat{\phi}(x)U^\dagger(a). \tag{77}$$

It is worth mentioning that some symmetries can mix fields, e.g.,

$$\phi_i'(x') = S_{ij}(a)\phi_j(x) \Rightarrow \phi_i(x') = S_{ij}(a)U(a)\phi_j(x)U^\dagger(a), \quad x' = x'(x,a). \tag{78}$$

Typical examples are fields with non-zero spin, e.g., vectors and fermions that we discussed in Section 2.

3.2 Local (gauge) symmetries

In this section we revise local symmetries, which play essential role in the construction of *interacting* models. Let us consider the free Dirac Lagrangian

$$\mathcal{L}_0 = \bar{\psi}\left(i\hat{\partial} - m\right)\psi \tag{79}$$

and make the *global* $U(1)$-symmetry of $\mathcal{L}_0$

$$\psi \to \psi' = e^{ie\omega}\psi \tag{80}$$

local, i.e., $\omega \to \omega(x)$. In this case, the Lagrangian ceases to be invariant[14]:

$$\delta\mathcal{L}_0 = \partial_\mu\omega \cdot J_\mu, \qquad J_\mu = -e\bar{\psi}\gamma_\mu\psi. \tag{81}$$

To compensate this term, we *introduce the interaction* of the current J_μ with the photon field A_μ:

$$\mathcal{L}_0 \to \mathcal{L} = \mathcal{L}_0 + A_\mu J_\mu = \bar{\psi}\left[i(\hat{\partial} + ie\hat{A}) - m\right]\psi, \qquad A_\mu \to A'_\mu = A_\mu - \partial_\mu\omega. \tag{82}$$

The photon A_μ is an example of *gauge* field. To get the full QED Lagrangian, we should also add a kinetic term for the photon:

$$\mathcal{L}_{QED} = \bar{\psi}\left(i\hat{D} - m\right)\psi - \frac{1}{4}F_{\mu\nu}^2 \tag{83}$$

$$D_\mu = \partial_\mu + ieA_\mu, \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{84}$$

[14] Note that one can use this fact to get an expression for the Noether current J_μ.

Here we introduce a *covariant* derivative D_μ and a *field-strength* tensor $F_{\mu\nu}$. One can check that Eq. (83) is invariant under

$$\begin{aligned} \psi \to \psi' &= e^{ie\omega(x)}\psi \\ A_\mu \to A'_\mu &= A_\mu - \partial_\mu \omega \\ D_\mu\psi \to D'_\mu\psi' &= e^{ie\omega(x)} D_\mu \psi. \end{aligned}$$

As a consequence, *gauge principle* forces us to add interactions. But there is price to pay. The *second* Noether theorem [21] states that theories possessing local or *gauge* symmetries are *redundant*, i.e., some degrees of freedom are not physical. This makes quantization non-trivial. To deal with this problem in QED, one usually adds a *gauge-fixing term* to the free vector-field Lagrangian:

$$\mathcal{L}_0(A) = -\frac{1}{4}F_{\mu\nu}^2 - \frac{1}{2\xi}\left(\partial_\mu A_\mu\right)^2 \equiv -\frac{1}{2}A_\mu K_{\mu\nu} A_\nu. \tag{85}$$

This term allows one to obtain the photon propagator by inverting $K_{\mu\nu}$:

$$\langle 0|TA_\mu(x)A_\nu(y)|0\rangle = \int \frac{d^4p}{(2\pi)^4} \frac{-i\left[g_{\mu\nu} - (1-\xi)p_\mu p_\nu/p^2\right]}{p^2 + i\epsilon} e^{-ip(x-y)} \tag{86}$$

The propagator now involves an auxiliary parameter ξ, and Eq. (26) corresponds to $\xi = 1$ (Feynman gauge). The parameter controls the propagation of *unphysical* longitudinal polarization $\epsilon^L_\mu \propto p_\mu$. The polarization turns out to be harmless in QED since the corresponding terms *drop out* of physical quantities, e.g., due to current conservation

$$e^L_\mu J_\mu \propto p_\mu J_\mu = 0 \quad \text{[we have no source for unphysical } \gamma\text{]}. \tag{87}$$

One can see that the propagator has good UV behaviour and falls down as $1/p^2$ for large p. The gauge symmetry of QED is $U(1)$. It is *Abelian* since the order of two transformations is irrelevant (see Fig. 6).

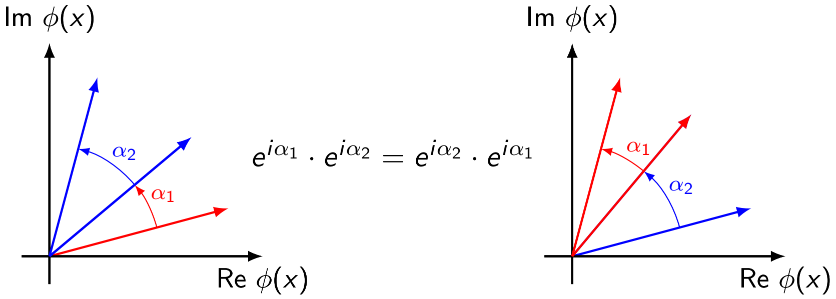


Fig. 6: $U(1)$ transformations commute with each other.

We can generalize $U(1)$ to the *Non-Abelian* case, which is relevant for the EW and QCD interactions. Let us consider the $SU(n)$ group, i.e., unitary $n \times n$ matrices U_{ij} depending on $n^2 - 1$ parameters ω^a and having det $U = 1$:

$$\psi_i \to \psi'_i = U_{ij}(\omega)\psi_j, \quad U(\omega) = e^{igt^a\omega^a}. \tag{88}$$

In general, different transformations do not commute in the non-Abelian case. This fact is reflected in commutation relations for the group *generators* t^a, which obey the $su(n)$-algebra:

$$[t^a, t^b] = if^{abc}t^c, \qquad f^{abc} - \text{ structure constants }. \tag{89}$$

For *constant* ω^a the transformation, Eq. (88), is a symmetry of the Lagrangian

$$\mathcal{L}_0 = \bar{\psi}_i \left(i\hat{\partial} - m \right) \psi_i, \qquad i = 1, ..., n \tag{90}$$

describing n free fermions in the *fundamental* representation of $SU(n)$.

In order to make the symmetry local, we introduce a (matrix) *covariant derivative* depending on $n^2 - 1$ gauge fields W_μ^a:

$$(D_\mu)_{ij} = \partial_\mu \delta_{ij} - igt^a_{ij} W^a_\mu. \tag{91}$$

The transformation properties of W^a_μ should guarantee that for space-time dependent $\omega^a(x)$ the covariant derivative of ψ transforms in the same way as the field itself:

$$D'_\mu \psi' = U(\omega)(D_\mu \psi), \quad U(\omega) = e^{igt^a\omega^a}. \tag{92}$$

One can find that

$$W^a_\mu \to W'^a_\mu = W^a_\mu + \partial_\mu \omega^a + gf^{abc} W^b_\mu \omega^c \tag{93}$$

$$= W^a_\mu + (D_\mu)^{ab}\omega_b, \qquad (D_\mu)^{ab} \equiv \partial_\mu \delta^{ab} - ig(-if^{abc})W^c_\mu, \tag{94}$$

where we introduce the covariant derivative, Eq. (91), D^{ab}_μ with generators $(t^c)^{ab} = -if^{cab}$ in the *adjoint* representation. The field-strength tensor for each component of W^a_μ is given by the commutator

$$[D_\mu, D_\nu] = -igt^a \mathcal{F}^a_{\mu\nu}, \quad \mathcal{F}^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\mu W^a_\nu + gf^{abc} W^a_\mu W^b_\nu. \tag{95}$$

Contrary to the $U(1)$ case, $\mathcal{F}^a_{\mu\nu}$ contains an additional term quadratic in W^a_μ. Due to this, the gauge symmetry predicts not only interactions between fermions ψ (or fields in the fundamental representation of the gauge group) and W^a_μ but also *self-interactions* of the latter (the gauge fields are *"charged"* under the group).

Combining all the ingredients, we can write down the following Lagrangian for an $SU(n)$ gauge (Yang-Mills) theory :

$$\mathcal{L} = \bar{\psi} \left(i\hat{D} - m \right) \psi - \frac{1}{4} \mathcal{F}^a_{\mu\nu} \mathcal{F}^a_{\mu\nu} = \mathcal{L}_0 + \mathcal{L}_I, \tag{96}$$

$$\mathcal{L}_0 = \bar{\psi} \left(i\hat{\partial} - m \right) - \frac{1}{4} F^a_{\mu\nu} F^a_{\mu\nu}, \quad F^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu, \tag{97}$$

$$\mathcal{L}_I = g\bar{\psi}^i_\alpha \gamma^\mu_{\alpha\beta} t^a_{ij} \psi^j_\beta W^a_\mu - \frac{g}{2} f^{abc} W^b_\mu W^c_\nu F^a_{\mu\nu} - \frac{g^2}{4} f^{abc} f^{ade} W^a_\mu W^b_\nu W^d_\mu W^e_\nu. \tag{98}$$

For illustration purposes we explicitly specify all the indices in the first term of interaction Lagrangian $\mathcal{L}_I$: the Greek ones correspond to Dirac (α, β) and Lorentz (μ) indices, while the Latin ones belong to different representations of $SU(n)$: i, j – fundamental, a – adjoint. One can also see that the strength of all interactions in $\mathcal{L}_I$ is governed by the single dimensionless coupling g.

To quantize a Yang-Mills theory, we generalize the QED gauge-fixing term and write, e.g.,

$$\mathcal{L}_{gf} = -\frac{1}{2\xi} \left(F^a \right)^2, \qquad F^a = \partial_\mu W^a_\mu \tag{99}$$

with F^a being a gauge-fixing function. This again introduces unphysical states in the W^a_μ propagator. However, contrary to the case of QED, the *fermionic* current $J^a_\mu = g\bar{\psi} t^a \gamma_\mu \psi$ is not conserved and can produce longitudinal W^a_μ. Nevertheless, the *structure* of vector-boson self-interactions guarantees that at *tree* level amplitudes, in which *one* of W^a_μ has an unphysical polarization, *vanish* (see, e.g., Fig. 7).

Unfortunately, this is not sufficient to get rid of unphysical states completely. For example, a virtual gauge boson can produce *a pair* of unphysical polarizations. At tree level we, in principle, can avoid

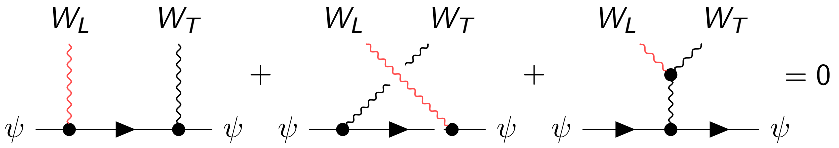


Fig. 7: Gauge symmetry at work: tree-level amplitudes with unphysical polarization (L) vanish.

them by restricting ourselves to physical external states. However, it is hard to control their appearance in loops. To deal with the problem in a *covariant* way, one introduces the so-called *Fadeev-Popov ghosts* $\bar{c}_a$ and c_a. They are *anticommuting* "scalars" and precisely cancel the annoying contribution[15]. The Lagrangian for the fictitious particles is related to the gauge-fixing function $F_a(W_\mu) = \partial_\mu W^a_\mu$ via

$$\begin{aligned}\mathcal{L}_{ghosts} &= -\bar{c}^a \frac{\partial F_a(W^\omega)}{\partial \omega_b} c^b = -\bar{c}^a \partial_\mu D^{ab}_\mu c^b \\ &= -\bar{c}^a \partial^2 c^a - g f^{abc} (\partial_\mu \bar{c}^a) c^b A^b_\mu. \end{aligned} \tag{100}$$

The ghosts are charged under $SU(n)$ and interact with gauge fields in the same way as the unphysical modes. However, there is an additional minus sign for the loops involving anticommuting ghosts (see, e.g., Fig. 8) that leads to the above-mentioned cancellations.

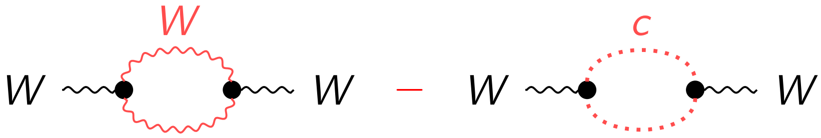


Fig. 8: Ghosts cancel contributions due to virtual unphysical states.

4 Gauge theory of electroweak interactions

4.1 From Fermi theory to the electroweak model

In 1957 R. Marshak and G. Sudarshan, R. Feynman and M. Gell-Mann modified the original Fermi theory of beta-decay to incorporate 100 % violation of parity discovered by C.S. Wu in 1956 :

$$-\mathcal{L}_{\text{Fermi}} = \frac{G_F}{2\sqrt{2}}(J^+_\mu J^-_\mu + \text{h.c.}). \tag{101}$$

Here the current

$$J^-_\rho = (V-A)^{\text{nucleons}}_\rho + \overline{\Psi}_e \gamma_\rho (1-\gamma_5)\, \Psi_{\nu_e} + \overline{\Psi}_\mu \gamma_\rho (1-\gamma_5)\, \Psi_{\nu_\mu} + ... \tag{102}$$

is the difference between Vector (V) and Axial (A) parts. It is worth mentioning that under parity

$$\begin{aligned} V^0 &\overset{P}{\to} \;\; V^0, & \mathbf{V} &\overset{P}{\to} -\mathbf{V}, \\ A^0 &\overset{P}{\to} -A^0, & \mathbf{A} &\overset{P}{\to} \;\; \mathbf{A}. \end{aligned}$$

[15] In a sense, ghosts also fix the unitary issue in non-Abelian theories: *optical* theorem applied to Feynman diagrams relates imaginary parts of loop integrals to the *squared* matrix elements, which can be obtained by "cutting" loop propagators (see, e.g., Ref. [14] for details).

As a consequence, parity P is conserved for *pure* vector $V_\mu V_\mu$ and axial $A_\mu A_\mu$ interactions, while it is the mixed $A_\mu V_\mu$ terms play a role in parity violation. One can also convince oneself that the charge-conjugation symmetry C is also not respected in this case (see, e.g., a cartoon in Fig. 9). Nevertheless, Eq. (101) conserves combined CP-parity, and it is better not to use the Wu experiment to set the notion of left and right in a phone call with aliens made of antimatter [22].

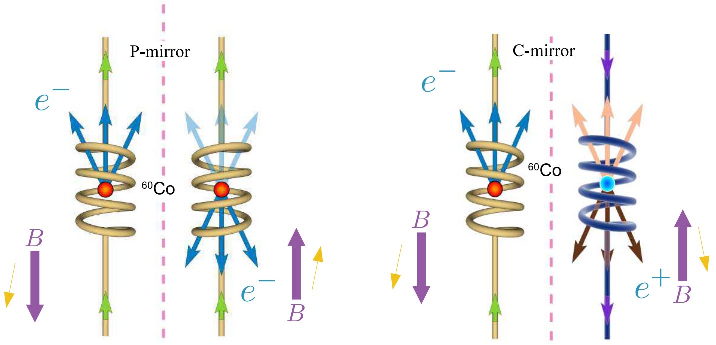


Fig. 9: A cartoon of the Wu experiment and its "distorted" images in P- and C-mirrors. One can see a correlation between the direction of the magnetic field (axial vector) and the direction of motion (polar vector) of the emitted electrons (positrons). The yellow arrows indicate the spin of the (anti) nuclei. The direction of the latter is correlated with that of emitted fermions. Adopted from Wikipedia.

The *current-current* interactions given in Eq. (101) can describe not only the proton beta-decay but also the muon decay $\mu \to e\nu_\mu\bar{\nu}_e$ or the process of $\nu_e e$ - scattering. Since the *Fermi* constant $G_F \simeq 10^{-5}$ GeV^{-1}, from simple *dimensional* grounds we have

$$\sigma(\nu_e e \to \nu_e e) \propto G_F^2 s, \qquad s = (p_e + p_\nu)^2. \tag{103}$$

With such a dependence on energy we eventually *violate unitarity*. This is another manifestation of the fact that non-renormalizable interactions are not self-consistent.

However, a modern view on the Fermi model treats it as an *effective* field theory [19] with certain *limits of applicability*. It perfectly describes low-energy experiments and one can fit the value of G_F very precisely (see Ref. [23]). The *magnitude* of G_F tells us something about a *more fundamental* theory (the SM in our case): around $G_F^{-1/2} \sim 10^2 - 10^3$ GeV there should be some "new physics" (NP) to cure the above-mentioned shortcomings. Indeed, by analogy with (renormalizable) QED we can introduce *mediators* of the weak interactions – electrically charged *vector* fields $W_\mu^\pm$:

$$\mathcal{L}_{\text{Fermi}} = -\frac{G_F}{2\sqrt{2}}(J_\mu^+ J_\mu^- + \text{h.c.}) \to \mathcal{L}_I = \frac{g}{2\sqrt{2}}(W_\mu^+ J_\mu^- + \text{h.c.}) \tag{104}$$

with a *dimensionless* coupling g. Since we know that weak interactions are *short-range*, the W-bosons should be *massive*. Given $\mathcal{L}_I$ we can calculate the tree-level scattering amplitude due to the exchange of $W^\pm$ between two fermionic currents:

$$T = i(2\pi)^4 \frac{g^2}{8} J_\alpha^+ \left[\frac{g_{\alpha\beta} - p_\alpha p_\beta/M_W^2}{p^2 - M_W^2}\right] J_\beta^-. \tag{105}$$

In the limit $|p| \ll M_W$, Eq. (105) reproduces the prediction of the effective theory (Fermi model) if we identify ("match")

$$\text{(effective theory)} \quad \frac{G_F}{\sqrt{2}} = \frac{g^2}{8{M_W}^2} \quad \text{(more fundamental theory)}. \tag{106}$$

At this point, it is good idea to compare the *chirality* structure of the W-coupling to fermions with that of the photon γ. In QED, the γ-fermion-fermion vertex conserves *chirality* and treats ψ_L and ψ_R on equal footing:

$$\mathcal{L}_I \ni -eA_\mu \cdot \bar{\psi}\gamma_\mu\psi = -eA_\mu\left[\bar{\psi}_L\gamma_\mu\psi_L + \bar{\psi}_R\gamma_\mu\psi_R + \cancel{\bar{\psi}_L\gamma_\mu\psi_R} + \cancel{\bar{\psi}_R\gamma_\mu\psi_L}\right].$$

As a consequence, in the high-energy limit ($m \to 0$) we have two helicity combinations, both for electrons and positrons, that give a non-zero amplitude. The weak vertex with W also conserves chirality but, due to postulated parity violation, involves only ψ_L

$$\mathcal{L}_I \ni -\frac{g}{2\sqrt{2}}W_\mu^+ \cdot \bar{\psi}_e\gamma_\mu(1-\gamma_5)\psi_{\nu_e} + \text{h.c.} = -\frac{g}{\sqrt{2}}W_\mu^+\left[\bar{\psi}_{eL}\gamma_\mu\psi_{\nu L}\right] + \text{h.c.},$$

and, thus, only one *helicity* combination take part in the ultra-relativistic processes involving W-bosons (see Fig. 10).

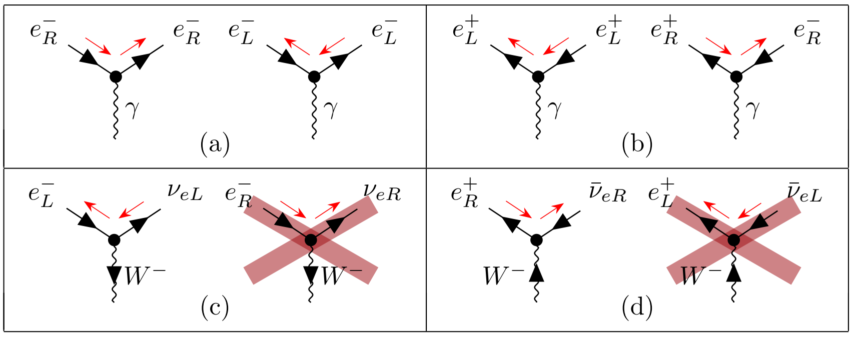


Fig. 10: Non-zero *helicity* combinations for electrons (a) and positrons (b) coupled to the photon in *massless* QED. In the case of W boson, only left-handed electrons (c) and right-handed positrons (d) contribute, if masses of the fermions are neglected. The red arrows represent helicities.

One interesting phenomenological consequence of the peculiar nature of the weak vertices is that it can be used to probe the (anti)quark content of the proton in Deep Inelastic Scattering (DIS) of (anti)neutrino. Indeed, let us consider a high-energy ($30\,\text{GeV} \lesssim E_\nu \lesssim 350\,\text{GeV}$) *muon* antineutrino produced in an accelerator-based beam. It can give an antimuon in the *charged-current* scattering either over the u quark, or over the $\bar{d}$ antiquark. Moreover, in the considered limit, the u quark should be *left-handed*, while $\bar{d}$ should be *right-handed*. The outgoing antimuon is also *right-handed* and to conserve helicity, the antineutrino cross-sections have the following form (we neglect the momentum transfer in the W-propagator, or, equivalently, use effective, Eq. (101), theory):

$$\frac{d\sigma_{\bar{\nu}q}}{d\Omega^*} = \frac{G_F s}{4\pi^2}\left(\frac{1+\cos\theta^*}{2}\right)^2, \qquad \sigma^{\bar{\nu}q} = \frac{G_F s}{3\pi}, \tag{107}$$

$$\frac{d\sigma_{\bar{\nu}\bar{q}}}{d\Omega^*} = \frac{G_F s}{4\pi^2}, \qquad \sigma^{\bar{\nu}\bar{q}} = \frac{G_F s}{\pi}. \tag{108}$$

Here θ^* is the scattering angle in the center-of-mass frame. Analogously, for the *left-handed* neutrino (see, Fig. 11), only incoming *left-handed* d or *right-handed* $\bar{u}$ can give a non-zero cross-section in the ultra-relativistic limit, so

$$\sigma^{\nu\bar{q}} = \frac{G_F s}{3\pi}, \qquad \sigma^{\nu q} = \frac{G_F s}{\pi}. \tag{109}$$

In the parton model the neutrino DIS over proton and neutron can be described by

$$\sigma_{\nu p} = \frac{G_F^2 s}{\pi}\left[f_d + \frac{1}{3}f_{\bar{u}}\right], \quad \sigma_{\nu n} = \frac{G_F^2 s}{\pi}\left[f_u + \frac{1}{3}f_{\bar{d}}\right], \tag{110}$$

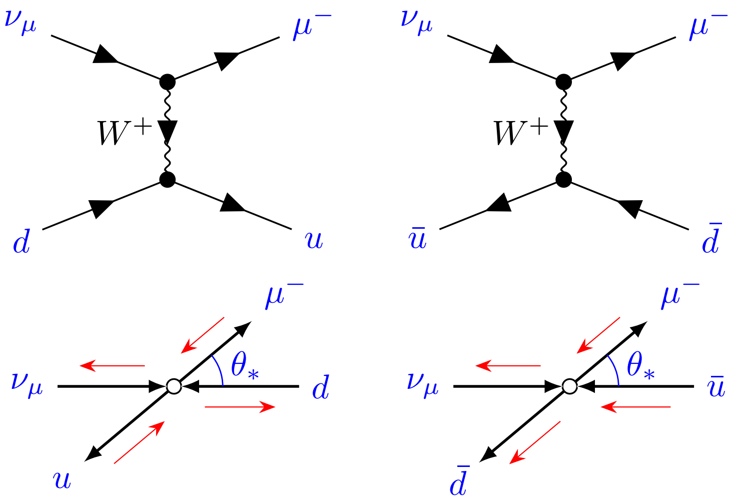


Fig. 11: Neutrino scattering on quarks and antiquarks. Red arrows represent helicity of the particles.

where $f_q = \int_0^1 xq(x)dx$ corresponds to the fraction of *proton* momentum carried by the quark q, and we assumed that $f_u(\text{proton}) = f_d(\text{neutron})$, etc. For an isoscalar target that have equal number of protons and neutrons, there is an equal probability to scatter either on p or n, so averaging over these possibilities gives

$$\sigma_{\nu N} = \frac{1}{2}\left[\sigma_{\nu p} + \sigma_{\nu n}\right] = \frac{G_F^2 s}{2\pi}\left[f_q + \frac{1}{3}f_{\bar{q}}\right], f_q = f_d + f_u,$$
$$\sigma_{\bar{\nu} N} = \frac{G_F^2 s}{2\pi}\left[f_{\bar{q}} + \frac{1}{3}f_q\right]. \tag{111}$$

One consequence of Eq. (111) is that the experimentally measured ratio

$$\frac{\sigma_{\nu N}}{\sigma_{\bar{\nu} N}} = \frac{3f_q + f_{\bar{q}}}{f_q + 3f_{\bar{q}}} = 1.984 \pm 0.012 \tag{112}$$

probes the *antiquark* $\bar{q}$ content of the proton, and indicates that antiquarks carry a non-zero fraction of the proton momentum $f_{\bar{q}} \simeq 0.08$.

4.2 The electroweak gauge bosons in the Standard Model

One can see that by construction $W^\pm$ is electrically charged, and interact with fermions and photons. Due to this, we can consider the W-pair production process ($e^+e^- \to W^+W^-$) at a lepton-antilepton collider (e.g., Large Electron-Positron (LEP) collider at CERN). We pretend to know nothing about the Z boson, so only two diagrams contribute in our theory (see first two graphs in Fig. 16). It turns out that in this case the predicted cross-section for the *longitudinal* W-bosons *increases* with center-of-mass energy s and, again, eventually violates unitarity.

In addition, the W-boson propagator, Eq. (24), behaves rather badly in the UV region (due to the $p_\mu p_\nu/m^2$ term in the numerator) and in loops can lead to severe UV divergencies. To deal with these issues in the SM, we associate a *gauge* symmetry with $W^\pm$, much like we do with photon. It turns out that to introduce EW interactions we need to utilize the

$$SU(2)_L \otimes U(1)_Y \tag{113}$$

gauge group that has four generators or, equivalently, four gauge bosons. Three of them, W_μ, belong to *weak-isospin* $SU(2)_L$, while the photon-like B_μ mediates *weak-hypercharge* $U(1)_Y$ interactions. The

SM fermions are charged under the group described in Eq. (113). To account for the $(V - A)$ pattern only *left* fermions interact with W_μ and form $SU(2)_L$ doublets:

$$L = \begin{pmatrix} \nu_l \\ l^- \end{pmatrix}_L, Q = \begin{pmatrix} q_u \\ q_d \end{pmatrix}_L, \qquad q_u = u, c, t; \; q_d = d, s, b; \; l = e, \mu, \tau. \tag{114}$$

Since the generators of $SU(2)$ are just the Pauli matrices, we immediately write the following expression for the corresponding covariant derivative

$$D^L_\mu = \begin{pmatrix} \partial_\mu - \frac{i}{2}\left(gW^3_\mu + g'Y^f_L B_\mu\right) & -i\frac{g}{\sqrt{2}}W^+_\mu \\ -i\frac{g}{\sqrt{2}}W^-_\mu & \partial_\mu + \frac{i}{2}\left(gW^3_\mu - g'Y^f_L B_\mu\right) \end{pmatrix}. \tag{115}$$

The *right* fermions[16] are $SU(2)_L$ singlets and do not couple to W_μ:

$$D^R_\mu = \partial_\mu - ig'\frac{Y^f_R}{2}B_\mu. \tag{116}$$

The covariant derivatives involve two gauge couplings g, g' corresponding to $SU(2)_L$ and $U(1)_Y$, respectively. Different $Y^f_{L/R}$ denote weak hypercharges of the fermions and up to now the values are not fixed. Let us put some constraints on $Y^f_{L/R}$. The first restriction comes from the $SU(2)_L$ symmetry, i.e., $Y^u_L = Y^d_L \equiv Y^Q_L$, and $Y^\nu_L = Y^e_L \equiv Y^L_l$.

One can see that the EW interaction Lagrangian

$$\mathcal{L}_W = \mathcal{L}_{NC} + \mathcal{L}_{CC}, \tag{117}$$

in addition to the *charged-current* interactions of the form

$$\mathcal{L}^l_{CC} = \frac{g}{\sqrt{2}}\bar{\nu}^e_L\gamma_\mu W^+_\mu e_L + \text{h.c.} = \frac{g}{2\sqrt{2}}\bar{\nu}_e\gamma_\mu W^+_\mu\left(1 - \gamma_5\right)e + \text{h.c.} \tag{118}$$

also involves *neutral-current* interactions

$$\mathcal{L}^l_{NC} = \bar{\nu}^e_L\gamma_\mu\left(\frac{1}{2}gW^3_\mu + \frac{Y^l_L}{2}g'B_\mu\right)\nu^e_L + \bar{e}_L\gamma_\mu\left(-\frac{1}{2}gW^3_\mu + \frac{Y^l_L}{2}g'B_\mu\right)e_L + g'\bar{e}_R\gamma_\mu\frac{Y^e_R}{2}B_\mu e_R. \tag{119}$$

It is obvious that we have to account for QED in the SM and should predict a photon field that couples to fermions with the correct values of the electric charges. Since both W^3_μ and B_μ are *electrically neutral*, they can mix

$$\begin{aligned} W^3_\mu &= Z_\mu\cos\theta_W + A_\mu\sin\theta_W \\ B_\mu &= -Z_\mu\sin\theta_W + A_\mu\cos\theta_W. \end{aligned} \tag{120}$$

Here we introduce the *Weinberg* angle θ_W. One can try to fix $\sin\theta_W$ and various $Y^f_{L/R}$ from the requirement that, e.g., A_μ has the same interactions as the photon in QED. Indeed, given fermion *electric* charges Q_f (see Table 3) in the units of the elementary charge e, one can derive the following relations:

$$\begin{aligned} g\sin\theta_W &= e(Q_\nu - Q_e) = e(Q_u - Q_d), \\ g'Y^l_L\cos\theta_W &= e(Q_\nu + Q_e) = -e, \\ g'Y^Q_L\cos\theta_W &= e(Q_u + Q_d) = \frac{1}{3}e, \end{aligned}$$

[16] In what follows we do not consider right-handed neutrino and refer again to Ref. [20].

$$g' Y_R^f \cos\theta_W = 2eQ_f, \qquad f = e,\, u,\, d. \tag{121}$$

As a consequence, $e = g\sin\theta_W$ and, e.g., $e = 3g' Y_L^Q \cos\theta_W$, so that

$$Y_L^l = -3Y_L^Q, \quad Y_R^e = -6Y_L^Q, \quad Y_R^u = 4Y_L^Q, \quad Y_R^d = -2Y_L^Q \tag{122}$$

are fixed in terms of one (arbitrary chosen) Y_L^Q. It is convenient to normalize the $U(1)_Y$ coupling g' so that $e = g'\cos\theta_W$, so $Y_L^Q = 1/3$. As a consequence, the photon field couples to the electric charge Q_f of a fermion f. The latter is related to the weak hypercharge and the third component of weak isospin T_3^f via the Gell-Mann–Nishijima formula:

$$\mathcal{L}_{NC} \ni \bar{f}\left[\left(gT_3^f \sin\theta_W + g'\frac{Y_f^L}{2}\cos\theta_W\right)P_L + \left(g'\frac{Y_f^R}{2}\cos\theta_W\right)P_R\right]\gamma_\mu f A_\mu \tag{123}$$

$$= e\bar{f}\left(T_3 + \frac{Y}{2}\right)\gamma_\mu f A_\mu = eQ_f \bar{f}\gamma_\mu f A_\mu, \tag{124}$$

where in Eq. (124) we assume that T_3 and Y are operators, which give T_3^f and Y_L^f, when acting on left components, and $T_3^f = 0$ and $Y_R^f = 2Q_f$ for right fermions.

The relations, Eq. (122), allow one to rewrite the neutral-current Lagrangian as

$$\mathcal{L}_{NC} = eJ_\mu^A A^\mu + g_Z J_\mu^Z Z_\mu, \qquad g_Z = \frac{g}{\cos\theta_W}, \tag{125}$$

where the photon A_μ and a new Z-boson couple to the currents of the form

$$J_\mu^A = \sum_f Q_f \bar{f}\gamma_\mu f, \qquad J_\mu^Z = \sum_f \bar{f}\left(c_L^f P_L + c_R^f P_R\right) f = \frac{1}{4}\sum_f \bar{f}\gamma_\mu\left(v_f - a_f\gamma_5\right) f \tag{126}$$

$$c_L^f = T_3^f - Q_f\sin^2\theta_W, \quad c_R^f = -Q_f\sin^2\theta_W, \quad v_f = 2T_3^f - 4Q_f\sin^2\theta_W, \quad a_f = 2T_3^f. \tag{127}$$

Here $T_3^f = \pm\frac{1}{2}$ for left up-type/down-type fermions. For example, in the case of u-quarks, $Q_u = 2/3$, $T_3^u = 1/2$, so

$$v_u = 1 - \frac{8}{3}\sin^2\theta_W, \qquad a_u = 1. \tag{128}$$

In Table 3, we summarize the SM fermion charges and the Z-boson couplings for $\sin^2\theta_W \simeq 0.23$.

Table 3: The values of the electric charge Q_f, the weak isospin T_3^f (for left particles), and the hypercharge for left Y_L^f and right Y_R^f SM fermions f. The Z-bosons coupling parameters c_L^f and c_R^f from Eq. (127) are also provided for $\sin^2\theta_W \simeq 0.23$.

fermion	Q_f	T_3^f	Y_L^f	Y_R^f	c_L^f	c_R^f
ν_l	0	$+\frac{1}{2}$	-1	0	$+\frac{1}{2}$	0
l^-	-1	$-\frac{1}{2}$	-1	-2	-0.27	$+0.23$
u	$+\frac{2}{3}$	$+\frac{1}{2}$	$+\frac{1}{3}$	$+\frac{4}{3}$	$+0.35$	-0.15
d	$-\frac{1}{3}$	$-\frac{1}{2}$	$+\frac{1}{3}$	$-\frac{2}{3}$	-0.42	$+0.08$

For completeness, let us give the expression for the charged-current interactions in the EW model

$$\mathcal{L}_{CC} = \frac{g}{\sqrt{2}}\left(J_\mu^+ W^{+\mu} + J_\mu^- W^{-\mu}\right), \quad J_\mu^+ = \frac{1}{2}\sum_f \bar{f}_u\gamma_\mu\left(1-\gamma_5\right) f_d, \tag{129}$$

where $f_u(f_d)$ is the up-type (down-type) component of an $SU(2)_L$ doublet f. The corresponding interaction vertices are given in Fig. 12. It is worth emphasizing that in the SM the couplings between fermions and gauge bosons exhibit *universality*.

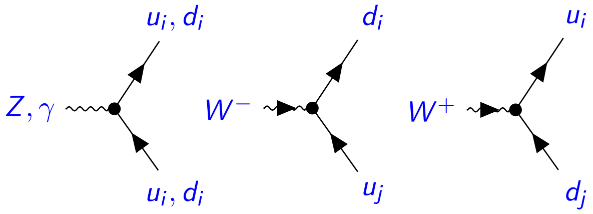


Fig. 12: Gauge-boson–quark vertices. Leptons interact with the EW bosons in the same way.

It turns out that it was *a prediction* of the electroweak SM that there should be an additional neutral gauge boson Z_μ. Contrary to the photon, the Z-boson also interacts with neutrinos. This crucial property was used in the experiment called *Gargamelle* at CERN, which presented the discovery in 1973 (Fig. 13). About ten years later both W and Z were directly produced at Super Proton Synchrotron (SPS) at CERN. Finally, in the early 90s a comprehensive analysis of the $e^+e^- \to f\bar{f}$ process, which was carried out at LEP, CERN, and at the Standford Linear Collider (SLC), SLAC, confirmed the SM predictions for the Z couplings to fermions, see Eg. (127).

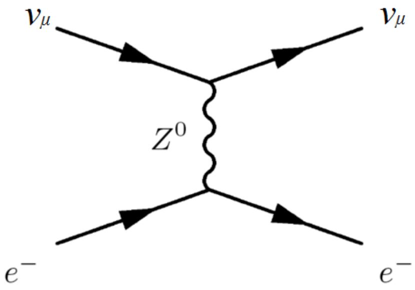


Fig. 13: The chamber of Gargamelle at CERN (left), ν_μ scattering due to Z-boson (right). From Wikipedia.

It is also worth mentioning the fact that the (hyper)-charge assignment, Eq. (122), satisfies very non-trivial constraints related to cancellation of *gauge anomalies*. Anomalies correspond to situations when a symmetry of the classical Lagrangian is violated at the quantum level. A well-known example is *Axial or Chiral or Adler–Bell–Jackiw(ABJ)* anomaly when the classical conservation law for the axial current J^A_μ is modified due to quantum effects:

$$J^A_\mu = \bar{\Psi}\gamma_\mu\gamma_5\Psi, \qquad \partial_\mu J^A_\mu = 2im\Psi\gamma_5\Psi + \underbrace{\frac{\alpha}{2\pi}F_{\mu\nu}\tilde{F}_{\mu\nu}}_{\text{anomaly}}, \qquad \tilde{F}_{\mu\nu} = 1/2\epsilon_{\mu\nu\rho\sigma}F_{\rho\sigma}. \tag{130}$$

The $F\tilde{F}$-term appears due to loop diagrams presented in Fig. 14.

There is nothing wrong when the anomalous current J^A_μ corresponds to a global symmetry and does not enter into $\mathcal{L}$. It just implies that a classically forbidden processes may actually occur in the quantum theory. For example, it is the anomaly in the *global* axial *flavour* symmetry that is responsible for the decay $\pi \to \gamma\gamma$. On the contrary, if an axial current couples to a gauge field, anomalies break gauge invariance, thus rendering the corresponding QFT inconsistent. In the SM left and right fermions (eigenvectors of γ_5) have different $SU(2)_L \times U(1)_Y$ quantum numbers, leaving space for potential

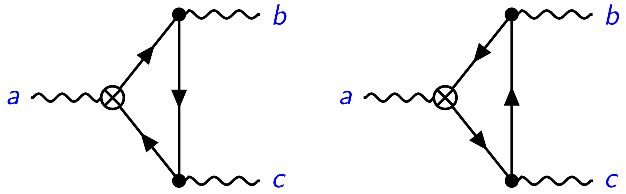


Fig. 14: Diagrams contributing to the anomaly of an axial current (crossed vertex).

anomalies. However, since we have to take into account all fermions which couple to a gauge field, there is a possibility that contributions from different species cancel each other due to a special assignment of fermion charges. Indeed, in the case of chiral[17] theories, anomalies are proportional to ($\gamma_5 = P_R - P_L$)

$$\text{Anom} \propto \text{Tr}[t^a, \{t^b, t^c\}]_L - \text{Tr}[t^a, \{t^b, t^c\}]_R, \tag{131}$$

where t^a are generators of the considered symmetries and the traces are over left (L) or right (R) fields. In the SM the requirement that all anomalies should be zero imposes the following conditions on fermion hypercharges:

$$0 = 2Y_L^Q - Y_R^u - Y_R^d, \qquad U(1)_Y - SU(3)_c - SU(3)_c, \tag{132a}$$

$$0 = N_c Y_L^Q + Y_L^l, \qquad U(1)_Y - SU(2)_L - SU(2)_L, \tag{132b}$$

$$0 = N_c \left[2(Y_L^Q)^3 - (Y_R^u)^3 - (Y_R^d)^3\right] + \left[2(Y_L^l)^3 - (Y_R^e)^3\right], \qquad U(1)_Y - U(1)_Y - U(1)_Y, \tag{132c}$$

$$0 = N_c \left[2Y_L^Q - Y_R^u - Y_R^d\right] + \left[2Y_L^l - Y_R^e\right], \qquad U(1)_Y - grav. - grav., \tag{132d}$$

where, in addition to the EW gauge group, we also consider strong interactions of quarks that have $N_c = 3$ colours. While the first three conditions come from the SM interactions, the last one, Eq. (132d) is due to the coupling to gravity. Other anomalies are trivially zero. One can see that the hypercharges introduced in Eq. (122) do satisfy the equations. It is interesting to note that contributions due to colour quarks miraculously cancel those of leptons and the cancellation works within a single generation. This put a rather strong restriction on possible new fermions that can couple to the SM gauge bosons: new particles should appear in a complete generation (quarks + leptons) in order not to spoil anomaly cancellation within the SM. Moreover, the anomaly cancellation condition can select viable models that go beyond the SM (BSM).

Due to the non-Abelian nature of the $SU(2)_L$ group, the gauge fields W_i have triple and quartic self-interactions (see Eq. (98)). Since W_3 is a linear combination of the Z-boson and photon, the same is true for Z and γ. In Fig. 15, self-interaction vertices for the EW gauge bosons are depicted. The triple vertex WWZ predicted by the SM allows one to cure the bad behavior of the $e^+e^- \to W^+W^-$

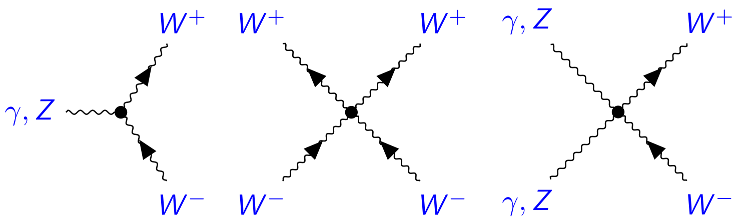


Fig. 15: Gauge-boson self-interaction vertices.

[17] that distinguish left and right fermions

cross-section, which we discussed in the beginning of the Section. Moreover, the coupling was tested experimentally at LEP2 (Fig. 16) and agreement with the SM predictions was found. Subsequent studies at hadron colliders (Tevatron and LHC) aimed at both quartic and triple gauge couplings (QGC and TGC, respectively) also show consistency with the SM and put limits on possible deviations (so-called anomalous TGC and QGC).

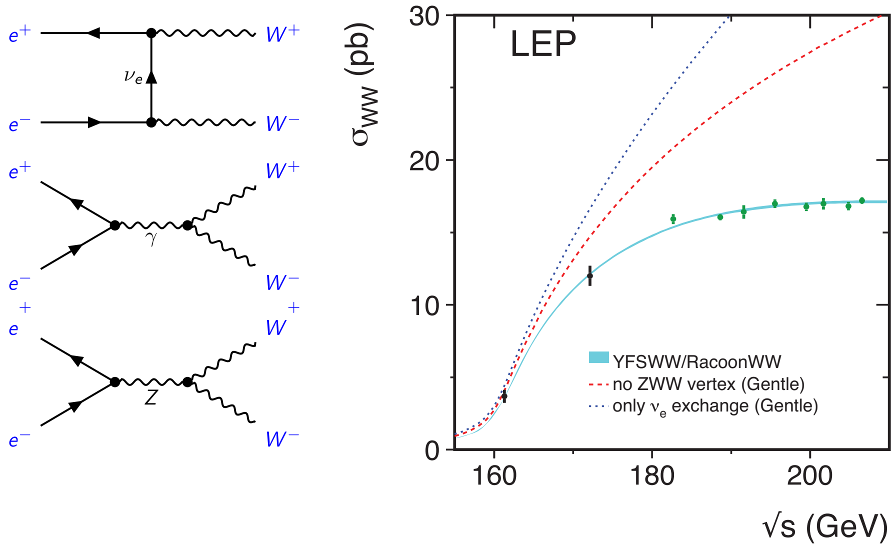


Fig. 16: $e^+e^- \to W^+W^-$.

Since we do not observe Z-bosons flying around like photons, Z_μ should have a non-zero mass M_Z and similar to $W^\pm$ give rise to Fermi-like interactions between *neutral* currents J_Z^μ at low energies. The relative strength of the *charged* and *neutral* current-current interactions $(J_\mu^Z J_Z^\mu)/(J^{+\mu} J_\mu^+)$ can be measured by the parameter ρ:

$$\rho \equiv \frac{M_W^2}{M_Z^2 \cos^2\theta_W}. \tag{133}$$

Up to now, we do not specify any relations between M_Z and M_W. Due to this, the value of ρ can, in principle, be arbitrary. However, it is a prediction of the full SM that $\rho \simeq 1$ (see below).

The fact that both W and Z should be massive poses a serious problem for theoretical description of the EW interactions. The naive introduction of the corresponding mass terms breaks the *gauge* symmetry, see Eq. (113). For example, $m_W^2 W_\mu^+ W_\mu^-$ is forbidden due to $W_\mu \to W_\mu + \partial_\mu \omega + ...$ One can also mention an issue with unitarity, which arises in the scattering of longitudinal EW bosons due to gauge self-interactions in Fig. 15.

In addition, the symmetry also forbids *explicit* mass terms for fermions, since e.g., $m_\mu(\bar\mu_L \mu_R + \mathrm{h.c.})$, which accounts for muon mass, mixes left and right fields that transform differently under the electroweak group, see Eq. (113). In the next section, we discuss how these problems can be solved by coupling the SM fermions and gauge bosons to the scalar (Higgs) sector (see also Ref. [24]).

4.3 Spontaneous symmetry breaking and hidden symmetry

We need to *generate* masses for $W_\mu^\pm$ and Z_μ (but not for A_μ) without *explicit* breaking of the gauge symmetry. Let us consider for simplicity *scalar* electrodynamics:

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial_\mu \phi - V(\phi^\dagger\phi) - \frac{1}{4}F_{\mu\nu}^2 + ie\left(\phi^\dagger \partial_\mu \phi - \phi \partial_\mu \phi^\dagger\right) A_\mu + e^2 A_\mu A_\mu \phi^\dagger \phi \equiv \mathcal{L}_1, \tag{134}$$

which is invariant under $U(1)$

$$\phi \to e^{ie\omega(x)}\phi, \quad A_\mu \to A_\mu + \partial_\mu \omega. \tag{135}$$

In Eq. (134) a *complex* scalar ϕ interacts with the photon A_μ. We can use *polar* coordinates to rewrite the Lagrangian in terms of new variables

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \rho)^2 + \frac{e^2\rho^2}{2}\left(A_\mu - \frac{1}{e}\partial_\mu\theta\right)\left(A_\mu - \frac{1}{e}\partial_\mu\theta\right) - V(\rho^2/2) - \frac{1}{4}F^2_{\mu\nu}, \tag{136}$$

$$= \frac{1}{2}(\partial_\mu \rho)^2 + \frac{e^2\rho^2}{2}B_\mu B_\mu - V(\rho^2/2) - \frac{1}{4}F^2_{\mu\nu}(B), \tag{137}$$

where ρ is gauge invariant, while the $U(1)$ transformation (135) gives rise to a *shift* in θ:

$$\phi = \frac{1}{\sqrt{2}}\rho(x)e^{i\theta(x)}, \quad \rho \to \rho, \quad \theta \to \theta + e\omega. \tag{138}$$

One can also notice that $B_\mu \equiv A_\mu - \frac{1}{e}\partial_\mu\theta$ is also invariant! Moreover, since $F_{\mu\nu}(A) = F_{\mu\nu}(B)$, we can completely get rid of θ. As a consequence, the gauge symmetry becomes "hidden" when the system is described by the variables $B_\mu(x)$ and $\rho(x)$.

If in Eq. (134) we replace our *dynamical* field $\rho(x)$ by a constant $\rho \to v = \text{const}$, we get the mass term for B_μ. This can be achieved by considering the potential $V(\phi)$ of the form (written in terms of initial variables)

$$V = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2. \tag{139}$$

One can distinguish two different situations (see Fig. 17):

- $\mu^2 > 0$ — a *single* minimum with $\phi = 0$;
- $\mu^2 < 0$ — a valley of *degenerate* minima with $\phi \neq 0$.

In both cases we solve EOM for the homogeneous (in space and time) field. When $\mu^2 > 0$ the *solution* is unique and symmetric, i.e., it does not transform under $U(1)$. In the second case, in which we are interested here, the potential has non-trivial minima

$$\left.\frac{\partial V}{\partial \phi^\dagger}\right|_{\phi=\phi_0} = 0 \Rightarrow \phi_0^\dagger\phi_0 = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2} > 0 \Rightarrow \phi_0 = \frac{v}{\sqrt{2}}e^{i\beta}, \tag{140}$$

which are *related* by *global* $U(1)$ transformations, Eq. (135), that change $\beta \to \beta + e\omega$. So, in spite of the fact that we do not break the symmetry *explicitly*, it is *spontaneously broken* (SSB) due to a particular choice of our solution (β).

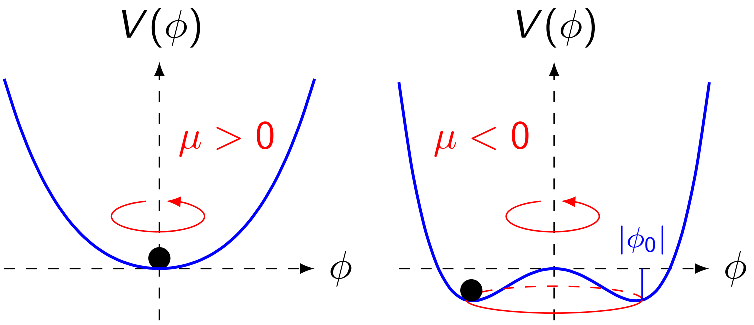


Fig. 17: A symmetric vacuum (left) and degenerate vacua (right).

In QFT we interpret ϕ_0 as a characteristic of our *vacuum* state, i.e., as a *vacuum expectation value* (VEV) or *condensate* of the quantum field:

$$\phi_0 = \langle 0|\phi(x)|0\rangle \overset{\beta=0}{=} \frac{v}{\sqrt{2}}. \tag{141}$$

Since we want to introduce particles as *excitations* above the vacuum, we have to shift the field:

$$\phi(x) = \frac{v + h(x)}{\sqrt{2}} e^{i\zeta(x)/v}, \qquad \langle 0|h(x)|0\rangle = 0, \quad \langle 0|\zeta(x)|0\rangle = 0. \tag{142}$$

As a consequence, Eq. (137) can be rewritten as

$$\mathcal{L} = \frac{1}{2}(\partial_\mu h)^2 + \frac{e^2v^2}{2}\left(1 + \frac{h}{v}\right)^2 B_\mu B_\mu - V(h) - \frac{1}{4}F^2_{\mu\nu}(B) \equiv \mathcal{L}_2, \tag{143}$$

$$V(h) = -\frac{|\mu|^2}{2}(v+h)^2 + \frac{\lambda}{4}(v+h)^4 = \frac{2\lambda v^2}{2}h^2 + \lambda v h^3 + \frac{\lambda}{4}h^4 - \frac{\lambda}{4}v^4. \tag{144}$$

The Lagrangian, Eq. (144), describes a massive vector field B_μ with $m_B^2 = e^2v^2$ and a massive scalar h with $m_h^2 = 2\lambda v^2$. We do not break the symmetry explicitly. It is again *hidden* in the relations between couplings and masses. This is the essence of the *Brout-Englert-Higgs-Hagen-Guralnik-Kibble* mechanism [25–27].

The Lagrangians $\mathcal{L}_1$, Eq. (134), and $\mathcal{L}_2$, Eq. (144), describe the same physics but written in terms of different quantities (variables). Eq. (134) involves a *complex* scalar ϕ with 2 (real) degrees of freedom (DOFs) and a *massless* gauge field (A_μ) also having 2 DOFs. It is manifestly gauge invariant but not suitable for perturbative expansion (ϕ has imaginary mass).

On the contrary, in $\mathcal{L}_2$ the gauge symmetry is hidden[18] and it is written in terms of *physical* DOFs, i.e., a *real* scalar h (1 DOF) and a *massive* vector B_μ (3 DOFs). In a sense, one *scalar* DOF (ζ) is "eaten" by the gauge field to become massive. It is important to note that the postulated *gauge* symmetry allows us to avoid the consequences of the *Goldstone* theorem, which states that if the vacuum breaks a *global* continuous symmetry there is a *massless* boson (Nambu-Goldstone) in the spectrum[19]. This boson is associated with 'oscillations" along the valley, i.e., in the *broken* direction (see Fig. 17). However, due to the local character of symmetry, χ is not physical anymore, its disappearance (or appearance, see below) reflects the *redundancy*, which was mentioned above.

In Section 2.6, we indicated that the massive-vector propagator has rather bad UV behavior and is not very convenient for doing calculations in PT. It looks like we gain nothing from the gauge principle. But it is not true. We can write the model Lagrangian in the *Cartesian* coordinates $\phi = \frac{1}{\sqrt{2}}(v + \eta + i\chi)$:

$$\mathcal{L}_3 = -\frac{1}{4}F_{\mu\nu}F_{\mu\nu} + \frac{e^2v^2}{2}A_\mu A_\mu + \frac{1}{2}\partial_\mu\chi\,\partial_\mu\chi - \underline{evA_\mu\partial_\mu\chi} + \frac{1}{2}\partial_\mu\eta\,\partial_\mu\eta - \frac{2v^2\lambda}{2}\eta^2 + \frac{v^4\lambda}{4} \tag{145}$$

$$+ eA_\mu\chi\partial_\mu\eta - eA_\mu\eta\partial_\mu\chi - v\lambda\eta(\eta^2+\chi^2) - \frac{\lambda}{4}(\eta^2+\chi^2)^2 + \frac{e^2}{2}A_\mu A_\mu(2v\eta + \eta^2 + \chi^2). \tag{146}$$

The "free" part, Eq. (145), of $\mathcal{L}_3$ seems to describe 5 real DOFs: a massive scalar η, a *massless* (would-be *Nambu-Goldstone*) boson χ and a massive A_μ. However, there is a mixing between the *longitudinal* component of A_μ and χ that spoils this naive counting (unphysical χ is "partially eaten" by A_μ) .

In spite of this subtlety, $\mathcal{L}_3$ is more convenient for calculations in PT. To quantize the model, one can utilize the gauge-fixing freedom and add the following expression to $\mathcal{L}_3$

$$\delta\mathcal{L}_{g.f.} = -\frac{1}{2\xi}(\partial_\mu A_\mu + ev\xi\chi)^2 = -\frac{1}{2\xi}(\partial_\mu A_\mu)^2 - \underline{ev\chi\partial_\mu A_\mu} - \frac{e^2v^2\xi}{2}\chi^2. \tag{147}$$

[18] One can also say that $\mathcal{L}_2$ corresponds to the *unitary* gauge, i.e., no unphysical "states" in the particle spectrum.

[19] Any non-derivative interactions violate the shift symmetry $\zeta \to \zeta + ev\omega$ for $\omega = \text{const}$

It removes the mixing from Eq. (145) and introduces a mass for χ, $m_\chi^2 = (e^2v^2)\xi$. In addition, the vector-boson propagator in this case looks like

$$\langle 0|TA_\mu(x)A_\nu(y)|0\rangle = \int \frac{d^4p}{(2\pi)^4}\frac{-i\left[g_{\mu\nu}-(1-\xi)\frac{p_\mu p_\nu}{p^2-\xi m_A^2}\right]}{p^2-m_A^2+i\epsilon}e^{-ip(x-y)}, \quad m_A = ev. \tag{148}$$

One can see that for $\xi \to \infty$ we reproduce Eq. (24), while for finite ξ the propagator behaves like $1/p^2$ as $p \to \infty$, thus making it convenient for PT calculations.

It should be mentioned that contrary to $\mathcal{L}_2$ the full Lagrangian corresponding to $\mathcal{L}_3$ involves also unphysical *ghosts*, which do not decouple in the considered case. Nevertheless, it is a relatively small price to pay for the ability to perform high-order calculations required to obtain high-precision predictions.

Let us switch back to the SM. We have three gauge bosons that should become massive. According to our reasoning, three symmetries should be broken by the SM vacuum to feed hungry $W_\mu^\pm$ and Z_μ with (would-be) Goldstone bosons

$$SU(2)_L \times U(1)_Y \to U(1)_{em}. \tag{149}$$

The photon should remain massless and correspond to the unbroken electromagnetic $U(1)_{em}$. This can be achieved by considering an $SU(2)_L$ doublet of scalar fields:

$$\Phi = \frac{1}{\sqrt{2}}\exp\left(i\frac{\zeta_j(x)\sigma^j}{2v}\right)\begin{pmatrix}0\\ v+h(x)\end{pmatrix}, \qquad \Phi_0 \equiv \langle 0|\Phi|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix}, \tag{150}$$

where we decompose $\Phi(x)$ in terms of three (would-be) Goldstone bosons ζ_j and a Higgs h. The Pauli matrices σ_j represent broken generators of $SU(2)_L$. Let Φ also be charged under $U(1)_Y$:

$$\Phi \to \exp\left(ig\frac{\sigma^i}{2}\omega_a + ig'\frac{Y_H}{2}\omega'\right)\Phi. \tag{151}$$

We do not want to break $U(1)_{em}$ spontaneously so the vacuum characterized by the VEV Φ_0 should be invariant under $U(1)_{em}$, i.e., has no electric charge Q

$$e^{ieQ\theta}\Phi_0 = \Phi_0 \to Q\Phi_0 = 0. \tag{152}$$

The operator Q is a linear combination of diagonal generators of $SU(2)_L \times U(1)_Y$, $T_3 = \sigma_3/2$ and $Y/2$:

$$Q\Phi_0 = \left(T_3 + \frac{Y}{2}\right)\Phi_0 = \frac{1}{2}\begin{pmatrix}1+Y_H & 0\\ 0 & -1+Y_H\end{pmatrix}\begin{pmatrix}0\\ \frac{v}{\sqrt{2}}\end{pmatrix} \overset{?}{=} 0. \tag{153}$$

As a consequence, to keep $U(1)_{em}$ unbroken, we should set $Y_H = 1$. Since Φ transforms under the EW group, we introduce gauge interactions for the Higgs doublet to make sure that the scalar sector respects the corresponding local symmetry:

$$\mathcal{L}_\Phi = (D_\mu\Phi)^\dagger(D_\mu\Phi) - V(\Phi), \quad \text{with} \quad V(\Phi) = m_\Phi^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2. \tag{154}$$

For $m_\Phi^2 < 0$ the symmetry is spontaneously broken. In the *unitary* gauge (Goldstone bosons are gauged away: in Eq. (150) we put $\zeta_j = 0$) the first term in Eq. (154) can be cast into

$$\begin{aligned}|D_\mu\Phi|^2 &= \frac{1}{2}(\partial_\mu h)^2 + \frac{g^2}{8}(v+h)^2|W_\mu^1 + iW_\mu^2|^2 + \frac{1}{8}(v+h)^2(gW_\mu^3 - g'Y_HB_\mu)^2 \\ &= \frac{1}{2}(\partial_\mu h)^2 + \frac{g^2}{4}(v+h)^2W^+W^- \quad \left[\sqrt{2}W^\pm = W_\mu^1 \mp iW_\mu^2\right]\end{aligned} \tag{155}$$

$$+\frac{1}{8}(v+h)^2\left[Z_\mu(g\cos\theta_W+g'\sin\theta_W)+A_\mu(g\sin\theta_W-g'\cos\theta_W)\right]^2 \tag{156}$$

$$=\frac{1}{2}(\partial_\mu h)^2+M_W^2\left(1+\frac{h}{v}\right)^2W^+W^-+\frac{M_Z^2}{2}\left(1+\frac{h}{v}\right)^2Z_\mu Z_\mu, \tag{157}$$

where we *require* the photon to be massless after SSB, i.e.,

$$g\sin\theta_W-g'\cos\theta_W=0 \quad\Rightarrow\quad \sin\theta_W=\frac{g'}{\sqrt{g^2+g'^2}},\cos\theta_W=\frac{g}{\sqrt{g^2+g'^2}} \tag{158}$$

and, consequently,

$$g\cos\theta_W+g'\sin\theta_W=\sqrt{g^2+g'^2}, \qquad e=g\sin\theta_W=g'\cos\theta_W=\frac{gg'}{\sqrt{g^2+g'^2}}. \tag{159}$$

The masses of the Z and W-bosons are proportional to the EW gauge couplings

$$M_W^2=\frac{g^2v^2}{4}, \quad M_Z^2=\frac{(g^2+g'^2)v^2}{4}. \tag{160}$$

One can see that the Higgs-gauge boson vertices (Fig. 18) are related to the masses M_W and M_Z.

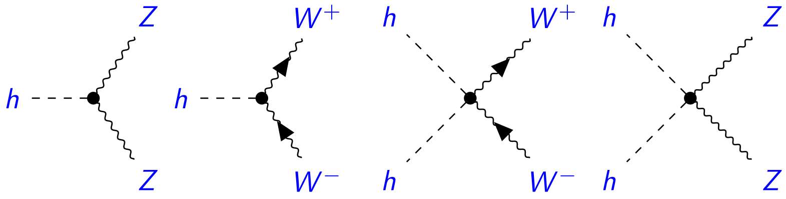


Fig. 18: Gauge-boson–Higgs interactions.

Earlier we emphasized that the introduction of the Z-boson cures the problem with unitarity in the process $e^-e^+\to W^-W^+$ for longitudinal W-bosons. Another important consequence of the SM gauge symmetry and the *existence* of the Higgs boson is the *unitarization* of massive vector-boson scattering. By means of simple power counting, one can easily convince oneself that the amplitude for (longitudinal) W-boson scattering originating from the quartic vertex in Fig. 15 scales with energy as E^4/M_W^4. However, in the SM, thanks to gauge symmetry, QGC and TGC are related. This results in E^2/M_W^2 behavior when Z/γ exchange is taken into account. Moreover, since the gauge bosons couple also to Higgs, we need to include the corresponding contribution to the total amplitude. It turns out that it is this contribution that cancels the E^2 terms and saves unitarity in the WW-scattering, as shown in Fig. 19. Obviously, this pattern is a consequence of the EW symmetry breaking in the SM and can be modified by the presence of new physics. Due to this, experimental studies of vector boson scattering (VBS) play a role in proving overall consistency of the SM.

Having in mind Eq. (106), one can derive the relation

$$G_F=\frac{1}{\sqrt{2}v^2}\Rightarrow v\simeq 246\text{ GeV}, \tag{161}$$

which gives a numerical estimate of v. One can also see that due to Eq. (160) we have (at the tree level)

$$\rho=\frac{M_W^2}{M_Z^2\cos^2\theta_W}=1. \tag{162}$$

Let us emphasize that it is a consequence of the fact that the SM Higgs is a weak *doublet* with *unit* hypercharge. Due to this, $\rho\simeq 1$ imposes important constraints on possible extensions of the SM Higgs

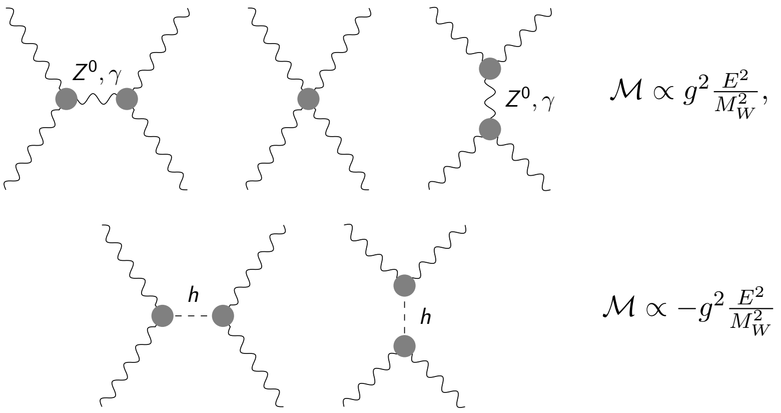

Z^0, γ
Z^0, γ
$\mathcal{M} \propto g^2 \frac{E^2}{M_W^2}$,
h
h
$\mathcal{M} \propto -g^2 \frac{E^2}{M_W^2}$

Fig. 19: WW-scattering and unitarity.

sector. For example, we can generalize Eq. (162) to account for n scalar $(2I_i + 1)$-plets $(i = 1, ..., n)$ that transform under $SU(2)_L$ and have hypercharges Y_i. In case they acquire VEVs v_i, which break the EW group, we have

$$\rho = \frac{\sum_i (I_i(I_i+1) - Y_i^2) v_i^2}{\sum_i 2 Y_i^2 v_i^2}. \tag{163}$$

Consequently, any non-doublet (with total weak isospin $I_i \neq 1/2$) VEV leads to a deviation from $\rho = 1$.

4.4 Fermion-Higgs interactions and masses of quarks and leptons

Since we fixed all the gauge quantum numbers of the SM fields, it is possible to construct the following *gauge-invariant* Lagrangian:

$$\mathcal{L}_Y = -y_e(\underset{+1}{\bar{L}}\ \underset{+1}{\Phi})\ \underset{-2}{e_R} - y_d(\underset{-\frac{1}{3}}{\bar{Q}}\ \underset{+1}{\Phi})\ \underset{-\frac{2}{3}}{d_R} - y_u(\underset{-\frac{1}{3}}{\bar{Q}}\ \underset{-1}{\Phi^c})\ \underset{\frac{4}{3}}{u_R} + \text{h.c.}, \tag{164}$$

which involves *dimensionless* Yukawa couplings y_f. It describes interactions between the Higgs field Φ, left fermion doublets, Eq. (114), and right singlets. In Eq. (164) we also indicate weak hypercharges of the corresponding fields. One can see that combinations of two doublets, $(\bar{Q}\Phi)$ etc., are invariant under $SU_L(2)$ but have a non-zero charge under $U(1)_Y$. The latter is compensated by hypercharges of right fermions. In addition, $U(1)_Y$ symmetry forces us to use a charge-conjugated Higgs doublet $\Phi^c = i\sigma_2\Phi^*$ with $Y = -1$ to account for Yukawa interactions involving u_R.

In the spontaneously broken phase with non-zero Higgs VEV, the Lagrangian $\mathcal{L}_Y$ can be written in the following simple form:

$$-\mathcal{L}_Y = \sum_f \frac{y_f v}{\sqrt{2}} \left(1 + \frac{h}{v}\right) \bar{f} f = \sum_f m_f \left(1 + \frac{h}{v}\right) \bar{f} f, \quad f = u,\ d,\ e, \tag{165}$$

where unitary gauge is utilized. One can see that SSB generates fermion masses m_f and, similarly to Eq. (157), *relates* them to the corresponding couplings of the Higgs boson h (see Fig 20a).

It is worth noting that Eq. (164) is not the most general renormalizable Lagrangian involving the SM scalars and fermions. One can introduce *flavour* indices and non-diagonal *complex* Yukawa matrices y_f^{ij} to account for a possible mixing between the SM fermions, i.e.,

$$\mathcal{L}_{\text{Yukawa}} = -y_l^{ij} (\bar{L}_i \Phi) l_{jR} - y_d^{ij} (\bar{Q}_i \Phi) d_{jR} - y_u^{ij} (\bar{Q}_i \Phi^c) u_{jR} + \text{h.c.}. \tag{166}$$

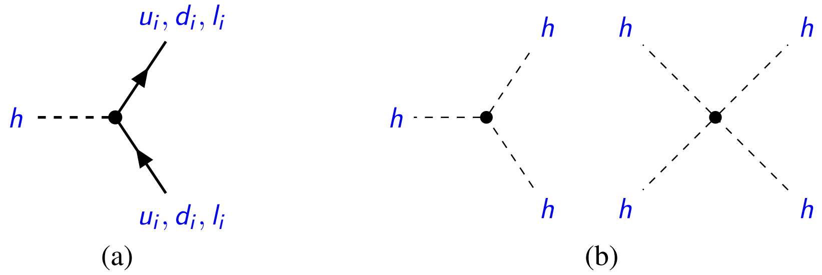


Fig. 20: Higgs–fermion couplings (a) and self-interactions of the Higgs boson (b).

Substituting $\Phi \to \Phi_0$ we derive the expression for fermion mass matrices $m_f^{ij} = y_f^{ij}\frac{v}{\sqrt{2}}$, which can be diagonalized by suitable unitary rotations of left and right fields. In the SM the Yukawa matrices, Eq. (166), are also diagonalized by the *same* transformations. This leads again (in the unitary gauge) to Eq. (165) but with the fields corresponding to the *mass* eigenstates. The latter *do not* coincide with *weak* states, which enter into $\mathcal{L}_W$, Eq. (117). However, one can rewrite $\mathcal{L}_W$ in terms of mass eigenstates. Due to large *flavour symmetry* of weak interactions, this introduces a single mixing matrix (the Cabibbo–Kobayashi–Maskawa matrix, or CKM), which manifests itself in the charged-current interactions $\mathcal{L}_{CC}$. A remarkable fact is that three generations are *required* to have $\mathcal{CP}$ violation in the quark sector. Moreover, a single CKM with only four physical parameters (angles and one phase) proves to be very successful in accounting for plethora of phenomena involving transitions between different flavours. We will not discuss further details but refer to the dedicated lectures on flavor physics [28].

5 The Standard Model: theory vs experiment

5.1 The Standard Model input parameters

Let us summarize and write down the full SM Lagrangian as

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{Gauge}}(g_s, g, g') + \mathcal{L}_{\text{Yukawa}}(y_u, y_d, y_l) + \mathcal{L}_{\text{Higgs}}(\lambda, m_\Phi^2) + \mathcal{L}_{\text{Gauge-fixing}} + \mathcal{L}_{\text{Ghosts}}. \tag{167}$$

The Yukawa part $\mathcal{L}_{\text{Yukawa}}$ is given in Eq. (166), while $\mathcal{L}_{\text{Higgs}} = -V(\Phi)$ is the Higgs potential from Eq. (154). After SSB the corresponding terms give rise to the Higgs couplings to the SM fermions (Fig. 20a) and Higgs self-interactions (Fig. 20b). The former are diagonal in the *mass* basis. The kinetic term for the Higgs field is included in

$$\mathcal{L}_{\text{Gauge}} = -\frac{1}{4}\underbrace{G^a_{\mu\nu}G^a_{\mu\nu}}_{SU(3)_c} - \frac{1}{4}\underbrace{W^i_{\mu\nu}W^i_{\mu\nu}}_{SU(2)_L} - \frac{1}{4}\underbrace{B_{\mu\nu}B_{\mu\nu}}_{U(1)_Y} + (D_\mu\Phi)^\dagger(D_\mu\Phi) \tag{168}$$

$$+ \underbrace{\bar{L}_i\, i\hat{D}\, L_i + \bar{Q}_i\, i\hat{D}\, Q_i}_{SU(2)_L \text{ doublets}} + \underbrace{\bar{l}_{Ri}\, i\hat{D}\, l_{R_i} + \bar{u}_{Ri}\, i\hat{D} u_{R_i} + \bar{d}_{Ri}\, i\hat{D}\, d_{R_i}}_{SU(2)_L \text{ singlets}}, \tag{169}$$

where for completeness we also add the colour group $SU(3)_c$ responsible for the strong force. The first three terms in Eq. (168) introduce gauge bosons for the SM gauge groups and in the non-Abelian case account for self-interactions of the latter (Fig. 15). The fourth term in Eq. (168) written in the form shown in Eq. (157) accounts for gauge interactions of the Higgs field (Fig. 18). Finally, Eq. (169) gives rise to interactions between gauge bosons and the SM fermions (see, e.g., Fig. 12).

The SM Lagrangian, Eq. (167), depends on 18 physical[20] parameters — 17 dimensionless couplings (gauge, Yukawa, and scalar self-interactions) and only 1 mass parameter m_Φ^2 (see Table 4). It is

[20] We do not count unphysical gauge-fixing parameters entering into $\mathcal{L}_{\text{Gauge-fixing}}$ and $\mathcal{L}_{\text{Ghosts}}$.

worth emphasizing here that there is certain freedom in the definition of *input* parameters. In principle, one can write down the SM predictions for a set of 18 observables (e.g., physical particle masses or cross-sections at fixed kinematics) that can be measured in experiments. With the account of loop corrections the predictions become non-trivial functions of *all* the Lagrangian parameters. By means of PT it is possible to invert these relations and express these primary parameters in terms of the chosen measured quantities. This allows us to *predict* other *observables in terms of* a finite set of measured *observables*[21].

However, it is not always practical to strictly follow this procedure. For example, due to confinement we are not able to directly probe the strong coupling g_s and usually treat it as a scale-dependent parameter $(4\pi)\alpha_s = g_s^2$ defined in the modified minimal-subtraction ($\overline{\text{MS}}$) scheme. It is customary to use the value of $\alpha_s^{(5)}(M_Z) = 0.1181 \pm 0.011$ at the Z-mass scale as an input for theoretical predictions. A convenient choice of other input parameters is presented in Table 4. It is mostly dictated by the fact that the parameters from the "practical" set are measured with better precision than the others.

Let us discuss some of the so-called Z pole observables that, after being measured with high precision at LEP and SLC, serve as an important input for the determination of the SM parameters.

Table 4: Parameters of the SM.

18=	1	1	1	1	1	9	4
primary:	g_s	g	g'	λ	m_Φ^2	y_f	y_{ij}
practical:	α_s	M_Z^2	α	M_H^2	G_F	m_f	V_{CKM}

5.2 *Z* pole observables

The electroweak model provides precise predictions for the properties of the Z boson, which can be tested in the process $e^+e^- \to Z \to f\bar{f}$ (Fig. 21). The latter dominate the cross-section $e^+e^- \to f\bar{f}$, if the center-of-mass energy is tuned to be $\sqrt{s} \simeq m_Z$. As we known, the heaviest SM fermion, that can be produced at such energies, is the b-quark with mass $m_b \simeq 4\,\text{GeV}$. Due to this, we will neglect all m_f in the following considerations.

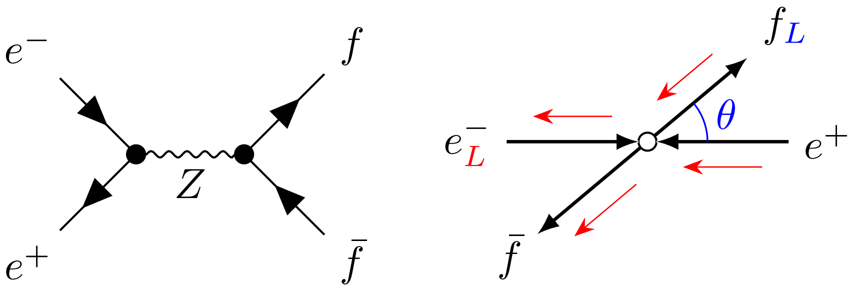


Fig. 21: The e^+e^- annihilation at $s \simeq m_Z^2$ (left), and one of the helicity combinations (right), which corresponds to the amplitude M_{LL} given in Eq.(170). Fermions are assumed to be massless.

Let us start by calculating the *tree-level* matrix elements for the processes involving fermions with certain helicity (= chirality). Since the Zff vertex conserve chirality, we will label the amplitudes by the *helicities* of the incoming e^- and outgoing f. For example, the *squared* amplitude

$$|M_{LL}|^2 = g_Z^4|P(z)|^2[c_L^e]^2[c_L^f]^2\,(1+\cos\theta)^2 \tag{170}$$

corresponds to the process $e_L^- e_R^+ \to f_L\bar{f}_R$, in which left-handed electron and right-handed positron annihilate to produce left-handed f and right-handed $\bar{f}$. In Eq. (170) the (Breit-Wigner) factor $|P(z)|^2 =$

[21] One can even avoid the introduction of *renormalizable* parameters and use *bare* quantities at the intermediate step.

$1/[(s-m_Z)^2+m_Z^2\Gamma_Z^2]$ originates from the Z-boson propagator, and the dependence on the scattering angle θ in the center-of-mass frame can again be understood from simple arguments based on helicity conservation (see Fig. 21). In the same way, one can obtain

$$|M_{RR}|^2 = g_Z^4|P(z)|^2[c_R^e]^2[c_R^f]^2\,(1-\cos\theta)^2\,, \tag{171}$$
$$|M_{LR}|^2 = g_Z^4|P(z)|^2[c_L^e]^2[c_R^f]^2\,(1-\cos\theta)^2\,, \tag{172}$$
$$|M_{RL}|^2 = g_Z^4|P(z)|^2[c_R^e]^2[c_L^f]^2\,(1-\cos\theta)^2\,. \tag{173}$$

For *unpolarised* $e^\pm$ beams we have to average over all possible initial helicity combinations. Integration over the scattering angle θ gives the total cross-section, which (in the narrow-width) approximation can be rewritten as

$$\sigma(ee\to Z\to ff) = \frac{12\pi s}{m_Z^2}\frac{\Gamma_{ee}\Gamma_{ff}}{(s-m_Z)^2+m_Z^2\Gamma_Z^2} \overset{\sqrt{s}=m_Z}{\Longrightarrow} \sigma^0_{ff} \equiv \frac{12\pi}{m_Z^2}\frac{\Gamma_{ee}\Gamma_{ff}}{\Gamma_Z^2}. \tag{174}$$

Here the partial width $\Gamma_{ff}\equiv\Gamma(Z\to ff)$ is given (at the tree-level) in terms of $c^f_{L,R}$ as

$$\Gamma(Z\to f\bar{f}) = \frac{g_Z^2 m_Z}{24\pi}\left([c_L^f]^2+[c_R^f]^2\right), \tag{175}$$

and Γ_Z represents the total Z width. From Eq. (174) one can see that the maximal value of the cross-section corresponds to $\sqrt{s}=m_Z$, so the position of the peak allows us to measure the mass of the Z boson. In addition, the fact that (full-width-at-half-maximum - FWHM)

$$\sigma(\sqrt{s}=m_Z\pm\Gamma_Z/2) = \sigma^0_{ff}/2,$$

allows us to extract Γ_Z directly from the energy dependence of the cross-section. Moreover, assuming *lepton universality*, we can experimentally determine $\Gamma_{ee}\simeq\Gamma_{\mu\mu}\simeq\Gamma_{\tau\tau}$ from $\sigma^0_{\mu\mu}$:

$$(12\pi)\Gamma_{ee}^2 = \sigma^0_{\mu\mu}\Gamma_Z^2 m_Z^2. \tag{176}$$

In the same way, by considering $e^+e^-\to$ hadrons one can extract the partial width for Z decaying into hadrons

$$(12\pi)\Gamma_{\text{hadrons}} = \sigma^0_{\text{hadrons}}\Gamma_Z^2 m_Z^2/\Gamma_{ee}. \tag{177}$$

Finally, assuming that $\Gamma_{\nu\nu}=\Gamma_{\nu\nu}^{SM}$ is *calculated* by means of Eq. (175) and Table 3, the number of neutrino (with $m_\nu < m_Z/2$) can be determined via

$$N_\nu = (\Gamma_Z - 3\Gamma_{ee} - \Gamma_{hadrons})/\Gamma^{SM}_{\nu\nu} \simeq 2.98. \tag{178}$$

Clearly, this is consistent with three fermion generations predicted by the SM.

We can have additional constraints on the SM parameters from measurements of various Z-pole *asymmetries* (see also Fig. 26). One example of such kind of observables is the Forward-Backward asymmetry A^f_{FB}, e.g.,

$$A^\mu_{FB} = \frac{\sigma_F-\sigma_B}{\sigma_F+\sigma_B} = \frac{3}{4}\mathcal{A}_e\mathcal{A}_\mu,\quad \mathcal{A}_f = \frac{(c_L^f)^2-(c_R^f)^2}{(c_L^f)^2+(c_R^f)^2} = \frac{v_f/a_f}{1+(v_f/a_f)^2}, \tag{179}$$

$$\sigma_F = 2\pi\int\limits_0^1 \frac{d\sigma}{d\Omega}d(\cos\theta),\quad \sigma_B = 2\pi\int\limits_{-1}^0 \frac{d\sigma}{d\Omega}d(\cos\theta). \tag{180}$$

Measurements of the asymmetry parameters $\mathcal{A}_f$ for leptons at LEP and SLC indicate that albeit being slightly different they are consistent with *universality* hypothesis $\mathcal{A}_e \simeq \mathcal{A}_\mu \simeq \mathcal{A}_\tau \simeq 0.15$. In addition, we can directly measure $\sin^2\theta_W$, since for leptons $v_l/a_l = 1 - 4\sin\theta_W^2$.

It is worth pointing here that our reasoning in this section was based on the *tree-level* amplitudes. Of course, to confront theory with high-precision experiment we have to take into account various quantum corrections. Moreover, one should perform various re-summations, e.g., to account for initial state radiation (ISR), which distorts the Breit-Wigner form of the distribution. Since the topic is quite involved, we will not go into further detail (see, e.g., Ref. [9]) here but give some other arguments regarding the importance of the quantum corrections in the SM.

5.3 On the importance of radiative corrections

At the *tree* level one can write the following relations between the parameters given in Table 4:

$$\begin{aligned} \alpha_s &= \tfrac{g_s^2}{4\pi}, \quad & (4\pi)\alpha &= g^2 g'^2/(g^2+g'^2), \quad & M_Z^2 &= \tfrac{(g^2+g'^2)v^2}{4}, \\ G_F &= \tfrac{1}{\sqrt{2}v^2}, \quad & M_h^2 &= 2\lambda v^2 = 2|m_\Phi|^2, \quad & m_f &= y_f v/\sqrt{2} \quad . \end{aligned} \tag{181}$$

At higher orders in PT, the relations are modified and perturbative corrections turn out to be mandatory if one wants to confront theory predictions [29–31] with high-precision experiments. A simple example to demonstrate this fact comes from the tree-level "prediction" for the W-mass M_W. From Eq. (160) and Eq. (181) we can derive

$$\frac{G_F}{\sqrt{2}} = \frac{\pi\alpha}{2M_W^2(1 - M_W^2/M_Z^2)}. \tag{182}$$

Plugging the Particle Data Group (PDG) [23] values

$$\alpha^{-1} = 137.035999139(31), \ M_Z = 91.1876(21)\ \text{GeV}, \ G_F = 1.1663787(6)\times 10^{-5}\ \text{GeV}^{-2}, \tag{183}$$

in Eq. (182), one predicts

$$M_W^{tree} = 80.9387(25)\ \text{GeV}, \tag{184}$$

where only uncertainties due to the input parameters, Eq. (183), are taken into account. Comparing M_W^{tree} with the measured value $M_W^{exp} = 80.379(12)$ GeV, one sees that our naive prediction is off by about 47σ! Of course, this is not the reason to abandon the SM. We just need to take quantum corrections into account (see, e.g., Fig. 22).

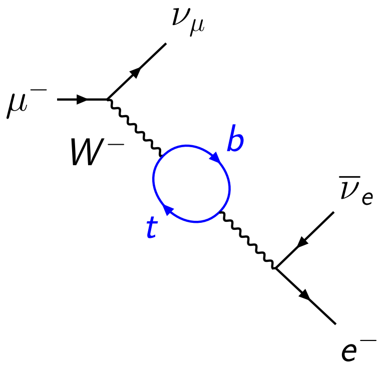


Fig. 22: An example of loop corrections to the muon decay, which give rise to the modification of the tree-level relation in Eq. (182).

The radiative corrections allows one to *relate* phenomena at different scales in the context of a single model. For example, we can study scale dependence of primary parameters, e.g., gauge couplings, and calculate high-order contribution to the corresponding beta-functions. At the one-loop order, we have the following general expression for the gauge-coupling RGE

$$\frac{d\alpha}{d\log\mu^2} = \beta\alpha^2 + \mathcal{O}(\alpha^3), \quad \alpha = \frac{g^2}{4\pi}, \quad \beta = -\frac{1}{4\pi}\left[\frac{11}{3}C_2 - \frac{2}{3}\sum_F T_F - \frac{1}{3}\sum_S T_S\right]. \tag{185}$$

Here $C_2 = N$ for the $SU(N)$ group, the sum goes over (Weyl) fermions (F) and scalars (S) coupled to the gauge field, and $T_F = T_S = \frac{1}{2}$. Figure 23 illustrates the scale dependence of the gauge couplings g, g' and g_s in the SM. It is worth pointing out that it was obtained by taking into account three-loop contributions to Eq. (185) and other SM couplings, which are also depicted for convenience. One can see that the gauge couplings tend to converge to a single value at about 10^{13-15} GeV, thus providing a hint for grand unification.

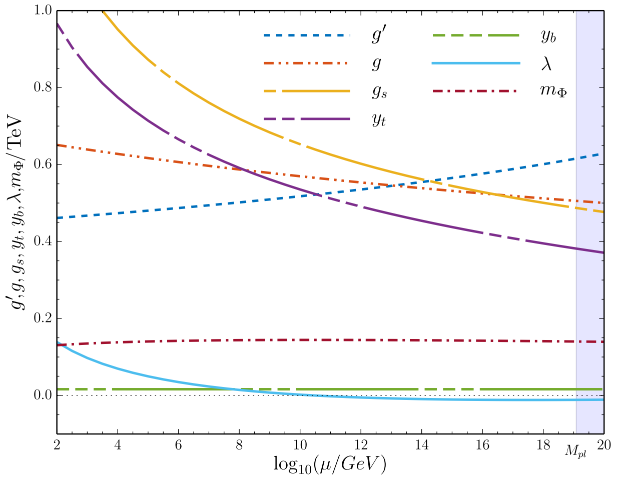


Fig. 23: Scale dependence of the SM parameters obtained by means of `mr` package [32].

Another important consequence of this kind of studies is related to the EW vacuum (meta)stability (see, e.g., Ref. [33]). In Fig. 23, it manifests itself at the scale $\mu \simeq 10^{10}$ GeV, at which the self-coupling λ becomes negative, making the tree-level potential unbounded from below. The two key parameters here are the top-quark mass M_t and the mass of the Higgs bosons M_h. According to Eq. (181) they can be related to the (boundary) values of λ and the top Yukawa y_t at the EW scale. The latter significantly influence self-coupling running, since (*cf.* Fig. 4)

$$(4\pi)^2\frac{d\lambda}{d\log\mu^2} = 12\lambda - 3{y_t}^4 + \ldots \quad . \tag{186}$$

A more elaborated analysis of the vacuum stability problem is based on the effective potential and gives rise to the well-known phase diagram in the $M_t - M_h$ plane (see, e.g., Fig. 24). One can see that the measured values of M_t and M_h lie just near the boundary between absolute stability (the EW vacuum is the true vacuum) and metastability (there exists a deeper minimum, but the tunneling time is much larger then the age of the Universe). This fact triggered many discussions about the fate of the EW vacuum in theoretical community. Without going into details, we just want to indicate again the importance of high-order corrections in the analysis: it is the next-to-next-to-leading (NNLO) effects (two-loop corrections to Eq. (181) and three-loop RGE) that "move" the absolute stability boundary just below the point corresponding to the experimentally measured values.

A modern way to obtain the values of the SM parameters is to perform a global fit to confront state-of-the-art SM predictions with high-precision experimental data. Due to quantum effects, we can

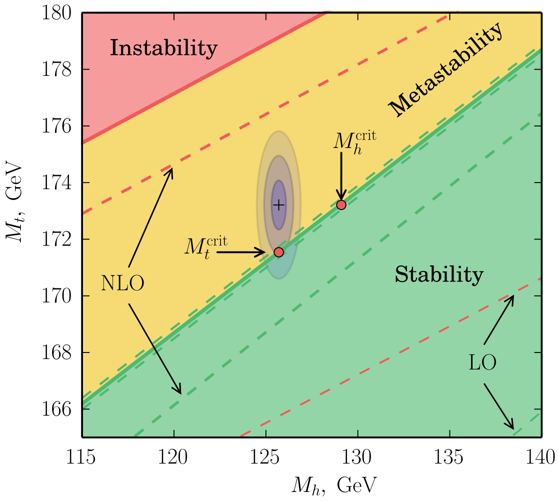


Fig. 24: Vacuum-stability phase diagram at three loops (NNLO). One can notice how the phase boundaries move upon switching from the leading-order (LO) one-loop evolution to NLO and NNLO running.

even probe new physics that can contribute to the SM processes at low energies via virtual states. Indeed, LEP precision measurements interpreted in the context of the SM were used in a multidimensional parameter fits to predict the mass of the top quark M_t ("new physics"), prior to its discovery at the Tevatron. After M_t was measured it was included in the fit as an additional constraint, and the same approach led to the prediction of a *light* Higgs boson. In Fig. 25, the famous *blue-band* plot by the LEP Electroweak Working Group (LEPEWWG) is presented [34]. It was prepared a couple of months before the official announcement of the Higgs-boson discovery. One can see that the best-fit value corresponding to $\Delta\chi^2_{min} = 0$ lies just about 1σ below the region *not* excluded by LEP and LHC.

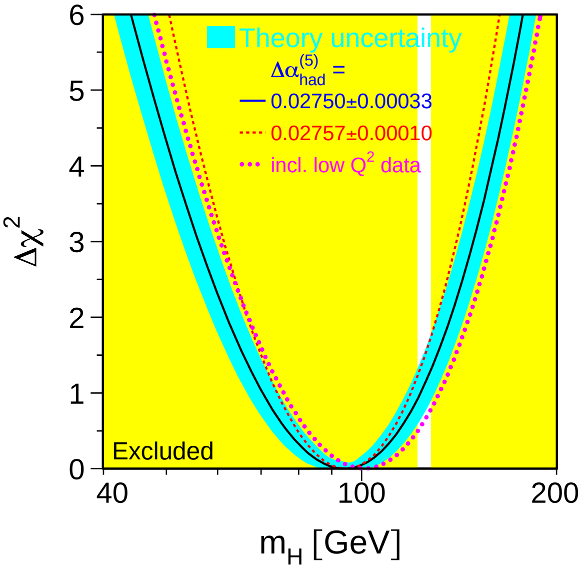
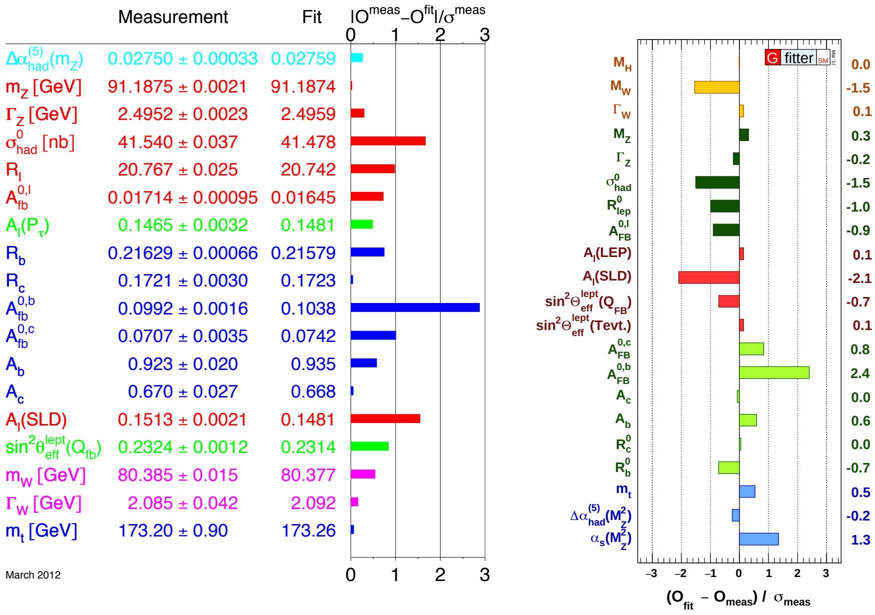


Fig. 25: The dependence of $\Delta\chi^2_{\rm min}(M_H^2) = \chi^2_{\rm min}(M_H^2) - \chi^2_{\rm min}$ on the value of M_H. The width of the shaded band around the curve shows the theoretical uncertainty. Exclusion regions due to LEP and LHC are also presented.

Obviously, at the moment the global EW fit is *over constrained* and can be used to test overall consistency of the SM. In Fig. 26 we present the comparison between measurements of different (pseudo)observables $O^{\rm meas}$ and the SM predictions $O^{\rm fit}$ corresponding to the best-fit values of fitted parameters. Although there are several quantities where *pulls*, i.e., deviations between the theory and experiment, reach more than two standard deviations, the average situation should be considered as extremely good. A similar conclusion can be drawn from the recent Figs. 27 and 28, in which experimental results for various cross-sections measured by ATLAS and CMS are compared with the SM predictions.



(a) (b)

Fig. 26: Pulls of various (pseudo)observables due to (a) LEPEWWG [34] and (b) Gfitter [35].

6 Conclusions

Let us summarize and discuss briefly the pros and cons of the SM. The model has many nice features:

- it is based on symmetry principles: Lorentz + $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry;
- it is renormalizable and unitary;
- the structure of all interactions is fixed (but not all couplings are tested experimentally);
- it is an anomaly-free theory;
- it can account for rich flavour physics (see Ref. [28]);
- three generations allow $\mathcal{CP}$-violation (see Ref. [28]);
- it can be extended to incorporate neutrino masses and mixing (see Ref. [20]);
- it allows making systematic predictions for a wide range of phenomena at different scales;
- all predicted particles have been discovered experimentally;
- it survives stringent experimental tests.

Due to this, the SM is enormously successful (*Absolutely Amazing Theory of Almost Everything*). Since it works so well, *any* new physics should reproduce it in the low-energy limit. Unfortunately, contrary to the Fermi-like non-renormalizable theories, the values of the SM parameters do not give us obvious

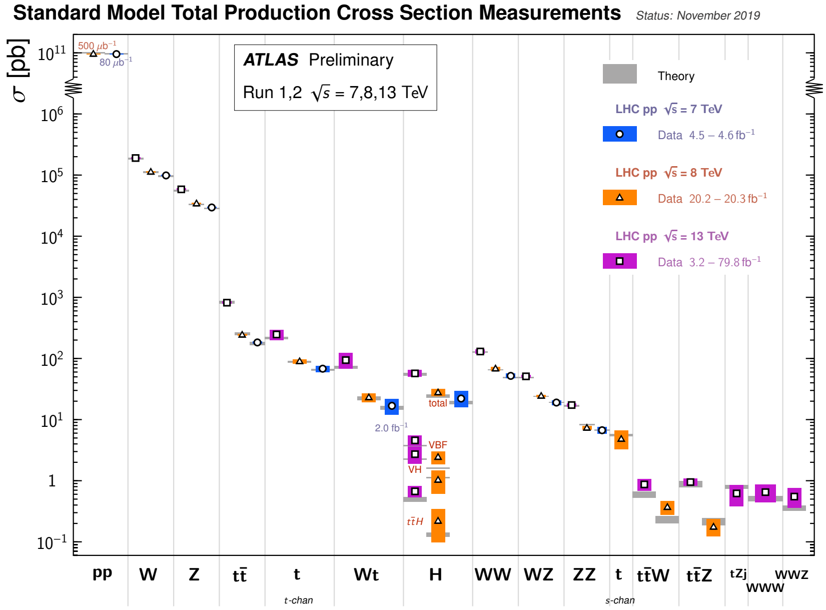


Fig. 27: ATLAS results of the SM cross-section measurements.

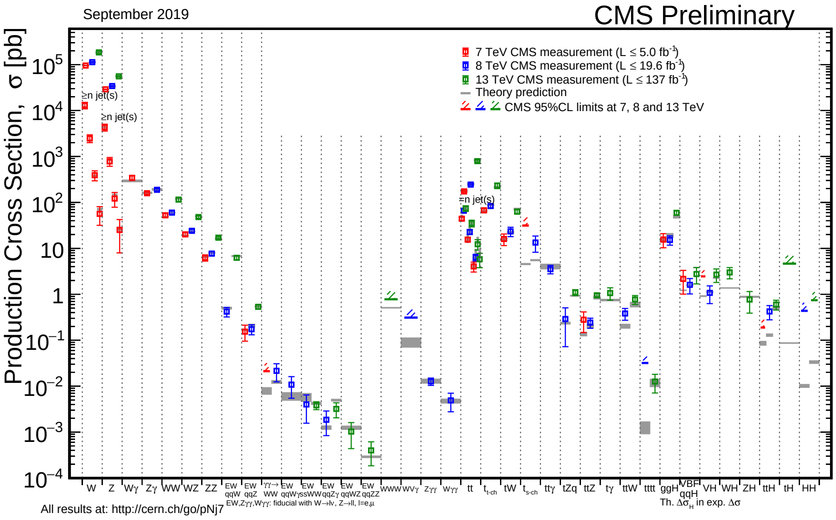


Fig. 28: SM processes at CMS.

hints for a new physics scale. But why do we need new physics if the model is so perfect? It turns out that we do not *understand*, why the SM works so well. For example, one needs to clarify the following:

- What explains the pattern behind flavour physics (hierarchy in masses and mixing, 3 generations)?
- Is there a symmetry behind the SM (electric) charge assignment?
- What is the origin of the Higgs potential?
- What is the origin of accidental baryon and lepton number symmetries?
- Why is there no CP-violation in the strong interactions (strong CP problem)?[22]
- Why is the Higgs-boson mass so low? (Hierarchy/naturalness problem, see Ref. [24])
- Is it possible to unify all the interactions, including gravity?

[22] The SM Gauge group allows such a term in the SM Lagrangian, $\mathcal{L} \ni \theta_{CP} \frac{1}{16\pi^2} F^a_{\mu\nu} \tilde{F}^a_{\mu\nu}$, but it turns out that $\theta_{CP} = 0$.

In addition, there are phenomenological problems that are waiting for solutions and probably require introduction of some new physics:

- Origin of neutrino masses (see Ref. [20]);
- Baryon asymmetry (see Ref. [36]);
- Dark matter, dark energy, inflation (see Ref. [36]);
- Tension in $(g-2)_\mu$, $b \to s\mu\mu$, $b \to cl\nu$;
- Possible problems with lepton universality of EW interactions (see Refs. [28, 37]).

In view of the above-mentioned issues we believe that the SM is not an ultimate theory (see Ref. [37]) and enormous work is ongoing to prove the existence of some new physics. In the absence of a direct signal a key role is played by *precision* measurements, which can reveal tiny, yet significant, deviations from the SM predictions. The latter should be accurate enough (see, e.g., Ref. [38]) to compete with modern and future experimental precision [39].

To conclude, one of the most important *tasks* in modern high-energy physics is to find the scale at which the SM breaks down. There is a big chance that some new physical phenomena will eventually manifest themselves in the ongoing or future experiments, thus allowing us to single out viable model(s) in the enormous pool of existing NP scenarios.

Acknowledgements

I would like to thank the organizers for the invitation to participate in the School. I am also grateful to students, discussion leaders, and other lecturers for interesting questions and discussions. Special thanks go to those, who spotted misprints in my slides. I hope that most, if not all of them, are exorcised from these lecture notes.

References

[1] S.L. Glashow, *Nucl. Phys.* **22** (1961) 579, doi:10.1016/0029-5582(61)90469-2.

[2] S. Weinberg, *Phys. Rev. Lett.* **19** (1967) 1264, doi:10.1103/PhysRevLett.19.1264.

[3] A. Salam, Weak and electromagnetic interactions, Proc. 8th Nobel Symposium, Ed. N. Svartholm (Almqvist & Wiksell, Stockholm, 1968), pp. 367–377, reprinted in Selected Papers of Abdus Salam, Eds. A. Ali *et al.*, (World Scientific, Singapore, 1994), pp. 244–254, doi:10.1142/9789812795915_0034.

[4] G. Aad *et al.* [ATLAS Collaboration], *Phys. Lett.* **B716** (2012) 1, doi:10.1016/j.physletb.2012.08.020.

[5] S. Chatrchyan *et al.* [CMS Collaboration], *Phys. Lett.* **B716** (2012) 30, doi:10.1016/j.physletb.2012.08.021.

[6] R. Oerter, *The theory of almost everything: The Standard Model, the unsung triumph of modern physics*, (Pi Press, New York, NY, 2006), CERN Library.

[7] R. Kleiss, Quantum field theory for the electroweak Standard Model, in Proc. European School of High-Energy Physics, Trest, Czech Republic, 2007, Eds. N. Ellis and R. Fleischer, pp. 1–137, doi:10.5170/CERN-2008-007.1.

[8] E. Boos, Quantum field theory and the electroweak Standard Model, in Proc. European School of High-Energy Physics, Paradfurdo, Hungary, 2013, Eds. M. Mulders and G. Perez, pp. 1–64, doi:10.5170/CERN-2015-004.1.

[9] M.E. Peskin, Lectures on the theory of the weak interaction, in Proc. European School of High-Energy Physics, Skeikampen, Norway, 2016, Eds. M. Mulders and G. Zanderighi, pp. 1–70, doi:10.23730/CYRSP-2017-005.1.

[10] A.B. Arbuzov, Quantum field theory and the electroweak Standard Model, in Proc. European School of High-Energy Physics, Bansko, Bulgaria, 2015, Eds. M. Mulders and G. Zanderighi, pp. 1–34, doi:10.23730/CYRSP-2017-004.1.

[11] J. Iliopoulos, Introduction to the Standard Model of the electro-weak interactions, in Proc. European School of High-Energy Physics, La Pommeraye, Anjou, 2012, Eds. C. Grojean and M. Mulders, pp. 1–42, doi:10.5170/CERN-2014-008.1.

[12] L.B. Okun, *Leptons and quarks*, (World Scientific, Singapore, 2014), doi:10.1142/9162.

[13] M. Thomson, *Modern particle physics*, (Cambridge University Press, Cambridge, 2013), doi:10.1017/CBO9781139525367.

[14] M.E. Peskin, D.V. Schroeder, *An Introduction to quantum field theory*, (Westview Press, Boulder, CO, 1995), CERN Library.

[15] N.N. Bogoliubov, D.V. Shirkov, *Introduction to the theory of quantized fields*, (Interscience Publishers, New York, NY, 1959), CERN Library.

[16] L.H. Ryder, *Quantum field theory*, 2nd ed., Cambridge University Press, Cambridge, 1996, doi:10.1017/CBO9780511813900.

[17] S. Weinberg, *The quantum theory of fields*, (Cambridge University Press, Cambridge, 1995–96), vol.1, 1995, doi: 10.1017/CBO9781139644167, vol.2, 1996, doi:10.1017/CBO9781139644174.

[18] A. Zee, *Quantum field theory in a nutshell*, 2nd ed. (Princeton Univ. Press, Princeton, NJ, 2010), CERN Library.

[19] H. Georgi, *Ann. Rev. Nucl. Part. Sci.* **43** (1993) 209. doi:10.1146/annurev.ns.43.120193.001233.

[20] M.C. Gonzalez-Garcia, Neutrino physics, these proceedings, pp. 85–128, doi:10.23730/CYRSP-2021-005.85.

[21] E. Noether, *Nachr. d. König. Gesellsch. d. Wiss. zu Göttingen, Math-phys. Klasse* (1918) 235–257, http://www.digizeitschriften.de/dms/resolveppn/?PID=GDZPPN00250510X, translation publ. in *Transp.Theory Statist.Phys.* **1** (1971), 186–207, doi:10.1080/00411457108231446.

[22] R.P. Feynman, R.B. Leighton and M. Sands, *The Feynman lectures on physics, vols. I–III*, (Addison-Wesley, Reading, MA, 1963–65), http://www.feynmanlectures.info, last accessed 23 March 2022.

[23] M. Tanabashi *et al.* [Particle Data Group], *Phys. Rev.* **D98** (2018) 030001, doi:10.1103/PhysRevD.98.030001.

[24] J. Ellis, Higgs physics, lecture at this school (not part of the proceedings).

[25] F. Englert and R. Brout, *Phys. Rev. Lett.* **13** (1964) 321, doi:10.1103/PhysRevLett.13.321.

[26] P.W. Higgs, *Phys. Rev. Lett.* **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.

[27] Scientific background on the Nobel Prize in Physics 2013: The BEH-mechanism, interactions with short range forces and scalar particles, compiled by the class for Physics of the Royal Swedish Academy of Sciences, (The Nobel Foundation, Stockholm, Sweden, 2013), https://www.nobelprize.org/uploads/2018/06/advanced-physicsprize2013.pdf.

[28] M. Vysotsky, Flavour physics and CP violation, these proceedings, pp. 47–84, doi:10.23730/CYRSP-2021-005.47.

[29] D.Y. Bardin and G. Passarino, *The Standard Model in the making: Precision study of the electroweak interactions*, (Clarendon Press, Oxford, 1999), CERN Library.

[30] A.B.Ȧrbuzov *et al.*, S. Riemann and T. Riemann, *Comput. Phys. Commun.* **174** (2006) 728, doi:10.1016/j.cpc.2005.12.009.

[31] G. Montagna *et al.*, *Comput. Phys. Commun.* **117** (1999) 278, doi:10.1016/10.1016/S0010-4655(98)00080-0.

[32] B.A. Kniehl, A.F. Pikelner and O.L. Veretin, *Comput. Phys. Commun.* **206** (2016) 84, doi:10.1016/j.cpc.2016.04.017.

[33] A.V. Bednyakov *et al.*, *Phys. Rev. Lett.* **115** (2015) 201802, doi:10.1103/PhysRevLett.115.201802.

[34] The LEP Electroweak Working Group, http://lepewwg.web.cern.ch/LEPEWWG/, last accessed 23 March 2022.

[35] M. Baak *et al.* [Gfitter Group], *Eur. Phys. J.* **C74** (2014) 3046, doi:10.1140/epjc/s10052-014-3046-5.

[36] V. Rubakov, Cosmology and dark matter, these proceedings, pp. 129–194, doi:10.23730/CYRSP-2021-005.129.

[37] V. Sanz, Physics beyond the Standard Model, lecture at this school (not part of the proceedings).

[38] A. Blondel *et al.*, Standard model theory for the FCC-ee Tera-Z stage, CERN-2019-003 (2019), doi:10.23731/CYRM-2019-003.

[39] M. Dam, *PoS* **(EPS-HEP2015)** 334, doi:10.22323/1.234.0334.

Proceedings of the 2019 European School of High-Energy Physics, St Petersburg, Russia, 14–27 September 2019, edited by C. Duhr and M. Mulders, CERN-2021-005 (CERN, Geneva, 2021)

Flavor physics and CP violation

M.I. Vysotsky
I.E. Tamm Department of Theoretical Physics, Lebedev Physical Institute, Moscow, Russia

Abstract
These notes contain a general introduction to the principles of flavor physics and CP violation. The material is based on the corresponding lectures given at the 2019 European School of High-Energy Physics that took place in St. Petersburg, Russia.

Keywords
Flavor physics; Heavy quarks; Meson mixing; CP; CP violation; Lectures.

1 Introduction

1.1 Fundamental particles and the periodic table

All known fundamental elementary particles are shown in Fig. 1.

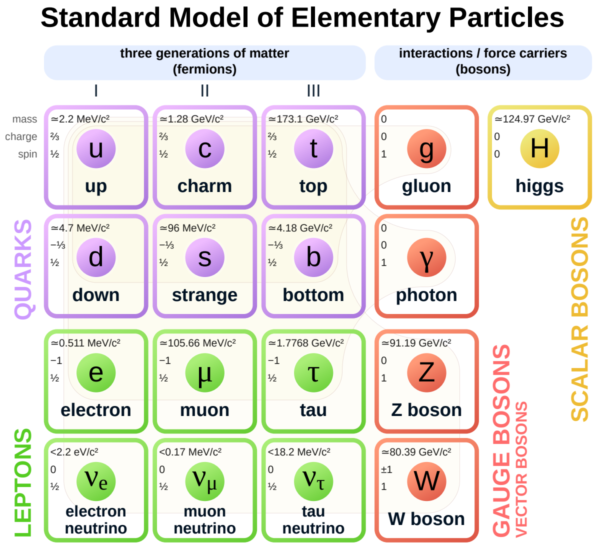
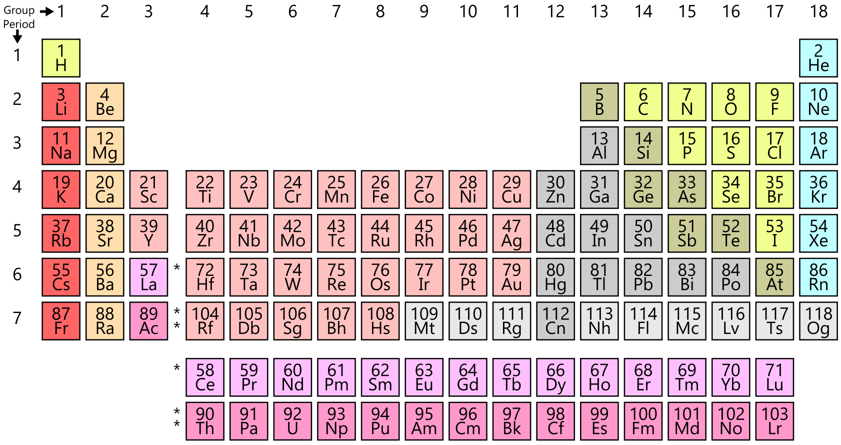


Fig. 1: Fundamental particles

One of the main problems for particle physics in the 21st century is why there are 3 quark-lepton generations and what explains fermion properties. This is a modern version of I.Rabi's question which he asked in response to the news that a recently discovered muon is not a hadron: "*Who ordered that?*".

Dmitry Mendeleev, professor of St. Petersburg University, discovered his Periodic table (modern version shown in Fig. 2) in 1869, just 150 years ago. He put there 63 existing elements and predicted 4 new elements. This 19th century discovery was explained by quantum mechanics in the beginning of the 20th century. Let us hope that an explanation of the table of elementary particles in general and the solution of a flavor problem (why there are 3 quark-lepton families and what is the physics which determines the values of quark and lepton masses and mixing parameters) in particular will be found in this century. There is much in common with the periodic table: the existence of W, Z and H was predicted as well. The central question is: what is an analog of Quantum Mechanics which explains so nicely the structure of the periodic table?

http://doi.org/10.23730/CYRSP-2021-005.47



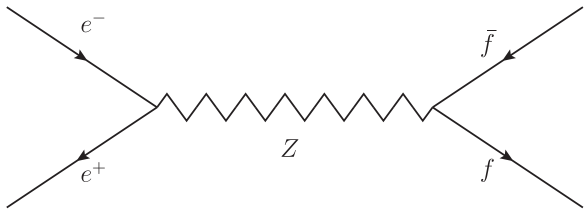
Group → Period ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H																	2 He
2	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	57 La *	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	89 Ac **	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Nh	114 Fl	115 Mc	116 Lv	117 Ts	118 Og
			*	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu	
			**	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr	

Fig. 2: Mendeleev's table

1.2 More generations?

After the discovery of the third generation the speculations on the 4th generation were very popular. Why only 3?

However for the invisible width of the Z boson we have:



$$\Gamma_{Z\to f\bar{f}} = \frac{G_F M_Z^3}{6\sqrt{2}\pi}[(g_V^f)^2 + (g_A^f)^2] = 332[(g_V^f)^2 + (g_A^f)^2] \text{ MeV}. \tag{1}$$

And taking into account Z decays into $\nu_e\bar{\nu}_e, \nu_\mu\bar{\nu}_\mu$ and $\nu_\tau\bar{\nu}_\tau$ we obtain:

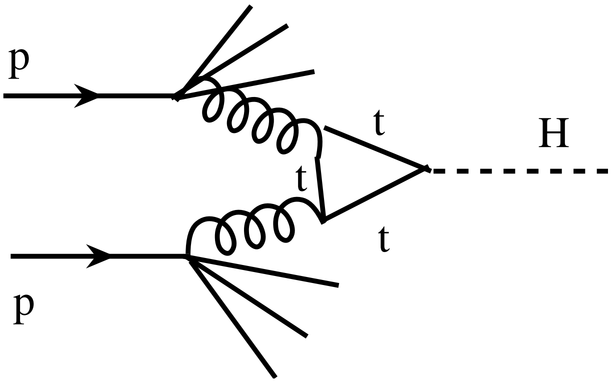
$$\Gamma^{\text{theor}}_{Z\to\nu\nu} = 3\cdot 332[\frac{1}{4}+\frac{1}{4}] = 498 \text{ MeV} \quad . \tag{2}$$

The invisible width of the Z-boson equals the difference between its total width and the sum of its decay width to hadrons and charged leptons. In this way the following result was obtained:

$$\Gamma^{\text{exp}}_{inv} = 499 \pm 1.5 \text{ MeV} \quad . \tag{3}$$

Comparing the last two equations we see that there is no space for Z decay into $\nu_4\bar{\nu}_4$ - so, there is no 4th generation. This statement is valid only for $m(\nu_4) < M_Z/2$. BUT: what if $m(\nu_4) > M_Z/2$?

In H production at LHC the following diagram dominates:



and for $2m_t >> M_H$ the corresponding amplitude does not depend on m_t.

In the case of a 4th generation $T-$ and $B-$ quarks would contribute as well, so the amplitude triples and the cross section of H production at LHC becomes 9 times larger than in the SM, which is definitely excluded by experimental data.

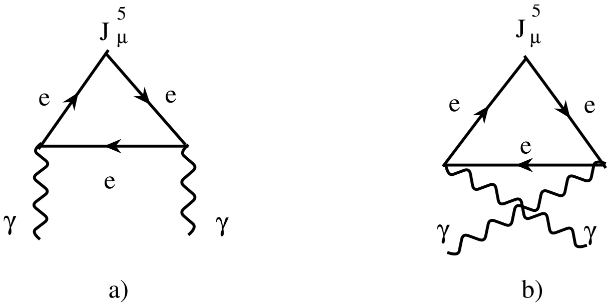
Problem 1

At LHC the values of signal strength $\mu_f \equiv \sigma(pp \longrightarrow H+X)*Br(H \longrightarrow f)/()_{SM}$ are measured. What will the change in μ_f be in case of a fourth generation?

1.3 Why $N_q = N_l$?

The equality $N_q = N_l$ must hold in order to compensate chiral anomalies, which would violate the conservation of gauge axial currents, making the theory non-renormalizable.

The following two diagrams lead to the axial current non-conservation in case of QED with massless electrons:



Fortunately photons couple to electrons by vector current which is conserved. Unlike QED with Dirac fermions (electrons), $SU(2)_L \times U(1)$ gauge invariant Standard Model (SM) [1] deals with Weyl fermions - states with definite chirality. Thus the gauge bosons A_i and B interact not only with vector currents, but with axial currents as well. In each generation the quarkonic and leptonic $A_i^2 B$ and B^3 triangles compensate each other, that is why N_q should be equal to N_l.

Problem 2

Prove that the quarkonic triangles cancel the leptonic ones when $Q_e = -Q_p$ (so hydrogen atoms are neutral) and $Q_n = Q_\nu = 0$ (thus neutrino and neutron are neutral).

2 Cabibbo-Kobayashi-Maskawa (CKM) matrix, unitarity triangles

2.1 The CKM matrix - where from?

In constructing the Standard Model Lagrangian the basic ingredients are:

1. gauge group,
2. particle content,
3. renormalizability of the theory.

The CKM matrix in charged current quark interactions appears automatically - one should not consider it as the Standard Model building block. Let us demonstrate where it comes from.

This is the SM Lagrangian:

$$\begin{aligned}\mathcal{L}_{\rm SM} = -\frac{1}{2}{\rm tr}G^2_{\mu\nu} - \frac{1}{2}{\rm tr}A^2_{\mu\nu} - \frac{1}{4}B^2_{\mu\nu} + |D_\mu H|^2 - \frac{\lambda^2}{2}[H^+H - \eta^2/2]^2 + \\ + \bar{Q}^i_L \hat{D} Q^i_L + \bar{u}^i_R \hat{D} u^i_R + \bar{d}^i_R \hat{D} d^i_R + \bar{L}^i_L \hat{D} L^i_L + \bar{l}^i_R \hat{D} l^i_R + \bar{N}^i_R \hat{\partial} N^i_R + \\ + \left[f^{(u)}_{ik} \bar{Q}^i_L u^k_R H + f^{(d)}_{ik} \bar{Q}^i_L d^k_R \tilde{H} + f^{(\nu)}_{ik} \bar{L}^i_L N^k_R H + f^{(l)}_{ik} \bar{L}^i_L l^k_R \tilde{H} + M_{ik} N^i_R C^+ N^k_R + c.c. \right] ,\end{aligned} \tag{4}$$

$$\hat{D} \equiv D_\mu \gamma_\mu \ , \ \ D_\mu = \partial_\mu - i g_s G^i_\mu \lambda_i/2 - i g A^i_\mu \sigma_i/2 - i g' B_\mu Y/2. \tag{5}$$

The CKM matrix originates from Higgs field interactions with quarks.

Quark fields in this lagrangian do not have definite masses. That is why it is convenient to write them with prime, changing fields in the lagrangian accordingly: $Q_L \to Q^{'}_L, u_R \to u^{'}_R, ...)$

2.2 The CKM matrix originates from Higgs field interactions with quarks.

The piece of the Lagrangian from which the up quarks get their masses looks like:

$$\Delta\mathcal{L}_{\rm up} = f^{(u)}_{ik} \bar{Q}^{i'}_L u^{k'}_R H + {\rm c.c.} \ , \ \ i,k = 1,2,3 \ \ , \tag{6}$$

where

$$Q^{1'}_L = \left(\begin{array}{c} u' \\ d' \end{array} \right)_L \ , \ \ Q^{2'}_L = \left(\begin{array}{c} c' \\ s' \end{array} \right)_L \ , \ \ Q^{3'}_L = \left(\begin{array}{c} t' \\ b' \end{array} \right)_L \ \ ; \tag{7}$$

$$u^{1'}_R = u'_R \ , \ \ u^{2'}_R = c'_R \ , \ \ u^{3'}_R = t'_R \tag{8}$$

and H is the higgs doublet:

$$H = \left(\begin{array}{c} H^0 \\ H^- \end{array} \right) . \tag{9}$$

The piece of the Lagrangian which is responsible for the down quark masses looks the same way:

$$\Delta\mathcal{L}_{\rm down} = f^{(d)}_{ik} \bar{Q}^{i'}_L d^{k'}_R \tilde{H} + {\rm c.c.} \ , \tag{10}$$

where

$$d^{1'}_R = d'_R \ , \ \ d^{2'}_R = s'_R \ , \ \ d^{3'}_R = b'_R \ \text{ and } \ \tilde{H}_a = \varepsilon_{ab} H^*_b \ , \tag{11}$$

$$\varepsilon_{ab} = \left(\begin{array}{cc} 0 & 1 \\ -1 & 0 \end{array} \right) . \tag{12}$$

After $SU(2) \times U(1)$ symmetry breaking by the Higgs field expectation value $< H^0 >= v$, two mass matrices emerge:

$$M^{ik}_{\rm up} \bar{u}^{i'}_L u^{k'}_R + M^{ik}_{\rm down} \bar{d}^{i'}_L d^{k'}_R + c.c. \tag{13}$$

The matrices $M_{\rm up}$ and $M_{\rm down}$ are arbitrary 3×3 matrices; their matrix elements are complex numbers. According to the very useful theorem, an arbitrary matrix can be written as a product of the hermitian and unitary matrices:

$$M = UH \ , \ \ \text{where} \ \ H = H^+ \ , \ \ \text{and} \ \ UU^+ = 1 \ , \tag{14}$$

(do not mix the hermitian matrix H with the Higgs field!) which is analogous to the following representation of an arbitrary complex number:

$$a = e^{i\phi}|a| \ . \tag{15}$$

Matrix M can be diagonalized by 2 different unitary matrices acting from left and right:

$$U_L M U_R^+ = M_{\rm diag} = \begin{pmatrix} m_u & & 0 \\ & m_c & \\ 0 & & m_t \end{pmatrix} \ , \tag{16}$$

where m_i are the real numbers (if matrix M is hermitian ($M = M^+$) then we will get $U_L = U_R$, the case of Hamiltonian in QM). Having these formulas in mind, let us rewrite the up-quarks mass term:

$$\bar{u}_L^{i'} M_{ik} u_R^{k'} + c.c. \equiv \bar{u}_L' U_L^+ U_L M U_R^+ U_R u_R' + c.c. = \bar{u}_L M_{\rm diag} u_R + c.c. = \bar{u} M_{\rm diag} u \ , \tag{17}$$

where we introduce the fields u_L and u_R according to the following formulas:

$$u_L = U_L u_L' \ , \ \ u_R = U_R u_R' \ . \tag{18}$$

Applying the same procedure to matrix $M_{\rm down}$ we observe that it becomes diagonal as well in the rotated basis:

$$d_L = D_L d_L' \ , \ \ d_R = D_R d_R' \ . \tag{19}$$

Thus we start from the primed quark fields and get that they should be rotated by 4 unitary matrices U_L, U_R, D_L and D_R in order to obtain unprimed fields with diagonal masses.

Since kinetic energies and interactions with the vector fields A_μ^3, B_μ and gluons are proportional to the unit matrix, these terms remain diagonal in a new unprimed basis. The only term in the SM Lagrangian where matrices U and D show up is charged current interactions with the emission of W-boson:

$$\Delta\mathcal{L} = gW_\mu^+ \bar{u}_L' \gamma_\mu d_L' = gW_\mu^+ \bar{u}_L \gamma_\mu U_L D_L^+ d_L \ , \tag{20}$$

and the unitary matrix $V \equiv U_L D_L^+$ is called Cabibbo-Kobayashi-Maskawa (CKM) quark mixing matrix.

2.3 Parametrization of the CKM matrix: angles, phases, unitarity triangles

$n \times n$ unitary matrix has $n^2/2$ complex or n^2 real parameters. The orthogonal $n \times n$ matrix is specified by $n(n-1)/2$ angles (3 Euler angles in case of $O(3)$). That is why the parameters of the unitary matrix are divided between phases and angles according to the following relation:

$$n^2 = \underset{\text{angles}}{\frac{n(n-1)}{2}} + \underset{\text{phases}}{\frac{n(n+1)}{2}} \ . \tag{21}$$

Are all these phases physical observables or, in other words, can they be measured experimentally?

The answer is “no” since we can perform phase rotations of quark fields ($u_L \to e^{i\zeta}u_L$, $d_L \to e^{i\xi}d_L$ ...) removing in this way $2n-1$ phases of the CKM matrix. The number of unphysical phases equals the number of up and down quark fields minus one. The simultaneous rotation of all up-quarks on one and the same phase multiplies all the matrix elements of matrix V by (minus) this phase. The rotation of all down-quark fields on one and the same phase acts on V in the same way. That is why the number of the “unremovable” phases of matrix V is decreased by the number of possible rotations of up and down quarks minus one.

Finally for the number of observable phases we get:

$$\frac{n(n+1)}{2} - (2n-1) = \frac{(n-1)(n-2)}{2} \quad . \tag{22}$$

As you see, for the first time one observable phase arrives in the case of 3 quark-lepton generations.

2.4 A bit of history

Introduced in 1963 by Cabibbo, the angle θ_c [2] in a modern language mixes d- and s-quarks in the expression for the charged quark current:

$$J^+_\mu = \bar{u}\gamma_\mu(1+\gamma_5)[d\cos\theta_c + s\sin\theta_c] \quad . \tag{23}$$

In this way he related the suppression of the strange particles weak decays to the smallness of angle θ_c, $\sin^2\theta_c \approx 0.05$.[1] In order to explain the suppression of $K^0 - \bar{K}^0$ transition the GIM mechanism (and c-quark) was suggested in 1970 [4]. After the discovery of a J/Ψ-meson made from ($c\bar{c}$) quarks in 1974 it was confirmed that 2 quark-lepton generations exist. The mixing of two quark generations is described by the unitary 2×2 matrix parametrized by one angle and zero observable phases. This angle is Cabibbo angle.

However, even before the c-quark discovery in 1973 Kobayashi and Maskawa noticed that one of the several ways to implement CP-violation in the Standard Model is to postulate the existence of 3 quark-lepton generations since for the first time the observable phase shows up for $n = 3$ [5]. At that time CPV was known only in neutral K-meson decays and to test KM mechanism one needed other systems. Almost 30 years after KM model had been suggested it was confirmed in B-meson decays.

Here is the CKM matrix

$$\overline{(uct)}_L \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L \quad , \tag{24}$$

and it’s standard parametrization looks like:

$$V = R_{23} \times R_{13} \times R_{12} \quad , \tag{25}$$

$$R_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} , \; R_{13} = \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} , \; R_{12} = \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} , \tag{26}$$

and, finally:

$$V = \begin{pmatrix} c_{13}c_{12} & c_{13}s_{12} & s_{13}e^{-i\delta} \\ -c_{23}s_{12} - s_{23}s_{13}c_{12}e^{i\delta} & c_{23}c_{12} - s_{12}s_{13}s_{23}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -s_{23}c_{12} - c_{23}s_{13}s_{12}e^{i\delta} & c_{23}c_{13} \end{pmatrix} . \tag{27}$$

[1] Earlier in the framework of "eightfold way“ such a suppression of the charged strange current was discussed by Gell-Mann [3].

2.5 Wolfenstein parametrization

Let us introduce new parameters λ, A, ρ and η according to the following definitions:

$$\lambda \equiv s_{12}, \quad A \equiv \frac{s_{23}}{s_{12}^2}, \quad \rho = \frac{s_{13}}{s_{12}s_{23}} \cos\delta, \quad \eta = \frac{s_{13}}{s_{12}s_{23}} \sin\delta \ , \tag{28}$$

and get the expressions for V_{ik} through λ, A, ρ and η:

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \approx \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda - iA^2\lambda^5\eta & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 - iA\lambda^4\eta & 1 \end{pmatrix} . \tag{29}$$

In the last expression the expansion in powers of λ is made.

The last form of CKM matrix is very convenient for qualitative estimates [6]. Approximately we have: $\lambda \approx 0.225, A \approx 0.83, \eta \approx 0.36, \rho \approx 0.15$.

2.6 Unitarity triangles; FCNC

The unitarity of the matrix V ($V^+V = 1$) leads to the following six equations that can be drawn as triangles on a complex plane (under each term in these equations the power of λ entering it, is shown):

$$\underset{\sim\lambda}{V^*_{ud}V_{us}} + \underset{\sim\lambda}{V^*_{cd}V_{cs}} + \underset{\sim\lambda^5}{V^*_{td}V_{ts}} = 0 \qquad s \to d \tag{30}$$

$$\underset{\sim\lambda^3}{V^*_{ud}V_{ub}} + \underset{\sim\lambda^3}{V^*_{cd}V_{cb}} + \underset{\sim\lambda^3}{V^*_{td}V_{tb}} = 0 \qquad b \to d \tag{31}$$

$$\underset{\sim\lambda^4}{V^*_{us}V_{ub}} + \underset{\sim\lambda^2}{V^*_{cs}V_{cb}} + \underset{\sim\lambda^2}{V^*_{ts}V_{tb}} = 0 \qquad b \to s \tag{32}$$

$$\underset{\sim\lambda}{V_{ud}V^*_{cd}} + \underset{\sim\lambda}{V_{us}V^*_{cs}} + \underset{\sim\lambda^5}{V_{ub}V^*_{cb}} = 0 \qquad c \to u \tag{33}$$

$$\underset{\sim\lambda^3}{V_{ud}V^*_{td}} + \underset{\sim\lambda^3}{V_{us}V^*_{ts}} + \underset{\sim\lambda^3}{V_{ub}V^*_{tb}} = 0 \tag{34}$$

$$\underset{\sim\lambda^4}{V_{cd}V^*_{td}} + \underset{\sim\lambda^2}{V_{cs}V^*_{ts}} + \underset{\sim\lambda^2}{V_{cb}V^*_{tb}} = 0 \tag{35}$$

Among these triangles four are almost degenerate: one side is much shorter than two others, and two triangles have all three sides of more or less equal lengths, of the order of λ^3. These two non-degenerate triangles have almost equal elements.

So, as a result we have only one non-degenerate unitarity triangle; it is usually defined by a complex conjugate of our equation:

$$V_{ud}V^*_{ub} + V_{cd}V^*_{cb} + V_{td}V^*_{tb} = 0 \tag{36}$$

and it is shown in Fig. 3. It has the angles which are called β, α and γ. They are determined from CPV asymmetries in B-mesons decays.

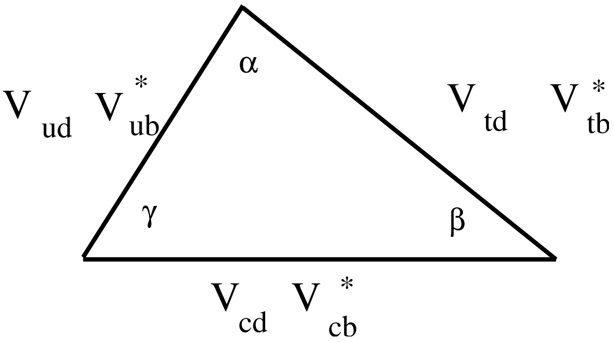


Fig. 3: Unitarity triangle

Looking at Fig. 3 one can easily obtain the following formulas:

$$\beta = \pi - \arg \frac{V_{tb}^* V_{td}}{V_{cb}^* V_{cd}} = \phi_1 \tag{37}$$

$$\alpha = \arg \frac{V_{tb}^* V_{td}}{-V_{ub}^* V_{ud}} = \phi_2 \tag{38}$$

$$\gamma = \arg \frac{V_{ub}^* V_{ud}}{-V_{cb}^* V_{cd}} = \phi_3 \tag{39}$$

- Angle β was measured through time dependent CPV asymmetry in $B_d \to \mathrm{charmonium}\ K^0$ decays,
- Angle α was measured from CPV asymmetries in $B_d \to \pi\pi$, $\rho\rho$ and $\pi\rho$ decays,
- $B^\pm$ decays are used to determine angle γ.

Multiplying any quark field by an arbitrary phase and absorbing it by CKM matrix elements we do not change some unitarity triangles, while the others are rotating as a whole, preserving their shapes and areas. For the area of any of unitarity triangle we get:

$$A = 1/2\mathrm{Im}(a \cdot b^*) = 1/2|a| \cdot |b| \cdot \sin\alpha \ \ , \tag{40}$$

where a and b are the sides of the triangle.

Problem 3

Prove that the areas of all unitarity triangles are the same. *Hint:* Use equations which define unitarity triangles.

2.7 Cecilia Jarlskog's invariant

The area of unitarity triangles contains an important information about the properties of CKM matrix.

CPV in the SM is proportional to this area, which equals $1/2$ of the Jarlskog invariant J [7].

Writing $J = Im(V_{ud}V_{ub}^*V_{cd}^*V_{cb})$ we see, that J is not changed when quark fields are multiplied by arbitrary phases.

The source of CPV in the SM is the phase δ - this is a correct statement; BUT it is like a phantom. If somebody says that the source of CPV is the phase of V_{td}, then another one can rotate d-quark, or t-quark, or both making V_{td} real.

However, there is an invariant quantity, which is not a phantom - J.

3 CP, CP violation

3.1 CP: history

Landau thought that space-time symmetries of a Lagrangian should be that of an empty space. Indeed, from a shift symmetry we deduce energy and momentum conservation, from rotation symmetry - angular momentum conservation. In 1956 Lee and Yang (in order to solve $\theta - \tau$ problem) suggested that P-parity is broken in weak interactions [8].

This was unacceptable for Landau: empty space has left-right interchange symmetry, so a Lagrangian should have it as well. Then Ioffe, Okun and Rudik noted that Lee and Yang's theory violates charge conjugation symmetry (C) as well, while CP is conserved explaining the difference of life times of K_L- and K_S- mesons [9] a-la Gell-Mann and Pais [10] but with CP replacing C. C-parity violation in weak interactions was discussed in [11] as well.

Just at this point Landau found the way to resurrect P-invariance stating that the theory should be invariant under the product of P reflection and C conjugation. He called this product the combined inversion and according to him it should substitute P-inversion broken in weak interactions. In this way the theory should be invariant when together with changing the sign of the coordinates, $\bar{r} \to -\bar{r}$, one changes an electron to positron, proton to antiproton and so on. Combined parity instead of parity.

It is clearly seen from 1957 Landau paper that CP-invariance should become a basic symmetry for physics in general and weak interactions in particular [12].

Nevertheless L.B. Okun considered the search for $K_L \to 2\pi$ decay to be one of the most important problems in weak interactions [13].

The notion of CP appears to be so important, that more than 60 years later you are listening to the lectures on CPV.

3.2 PV

Landau's answer to the question "Why is parity violated in weak interactions" was: because CP, not P is the fundamental symmetry of nature.

A modern answer to the same question is: because in P-invariant theory with the Dirac fermions the gauge invariant mass terms can be written for quarks and leptons which are not protected from being of the order of $M_{\rm GUT}$ or $M_{\rm Planck}$. So in order to have our world made from light particles P-parity should be violated, thus Weyl fermions should be used.

3.3 CPV

$K_L \to 2\pi$ decay discovered in 1964 by Christenson, Cronin, Fitch and Turlay [14] occurs due to CPV in the mixing of neutral kaons ($\tilde{\varepsilon} \neq 0$). Only thirty years later the second major step was done: direct CPV was observed in kaon decays [15]:

$$\frac{\Gamma(K_L \to \pi^+\pi^-)}{\Gamma(K_S \to \pi^+\pi^-)} \neq \frac{\Gamma(K_L \to \pi^0\pi^0)}{\Gamma(K_S \to \pi^0\pi^0)} \; , \quad \varepsilon' \neq 0 \; . \tag{41}$$

In the year 2001 CPV was for the first time observed beyond the decays of neutral kaons: the time dependent CP-violating asymmetry in B^0 decays was measured [16]:

$$a(t) = \frac{dN(B^0 \to J/\Psi K_{S(L)})/dt - dN(\bar{B}^0 \to J/\Psi K_{S(L)})/dt}{dN(B^0 \to J/\Psi K_{S(L)})/dt + dN(\bar{B}^0 \to J/\Psi K_{S(L)})/dt} \neq 0 \; . \tag{42}$$

Finally, in 2019 direct CPV was found in $D^0(\bar{D}^0)$ decays to $\pi^+\pi^-(K^+K^-)$ [17].

Since 1964 we have known that there is no symmetry between particles and antiparticles. In particular, the C-conjugated partial widths are different:

$$\Gamma(A \to BC) \neq \Gamma(\bar{A} \to \bar{B}\bar{C}) \ . \tag{43}$$

However, CPT (deduced from the invariance of the theory under 4-dimensional rotations) remains intact. That is why the total widths as well as the masses of particles and antiparticles are equal:

$$M_A = M_{\bar{A}} \ , \ \ \Gamma_A = \Gamma_{\bar{A}} \quad (\mathrm{CPT}) \ . \tag{44}$$

The consequences of CPV can be divided into macroscopic and microscopic. CPV is one of the three famous Sakharov's conditions to get a charge non-symmetric Universe as a result of evolution of a charge symmetric one [18]. In these lectures we will not discuss this very interesting branch of physics, but will deal with CPV in particle physics where the data obtained up to now confirm Kobayashi-Maskawa model of CPV. New data which should become available in coming years may as well disprove it clearly demonstrating the necessity of physics beyond the Standard Model.

3.4 CPV and complex couplings

The next question I would like to discuss is why the phases are relevant for CPV. In the SM charged currents are left-handed:

$$\Delta\mathcal{L} = g\bar{u}_L\gamma_\mu V d_L W_\mu + g\bar{d}_L\gamma_\mu V^+ u_L W_\mu^* \ . \tag{45}$$

Under space inversion (P) they become right-handed. Under charge conjugation (C) left-handed charged currents become right-handed as well and field operators become complex conjugate.

So, weak interactions are P- and C-odd.

However, CP transforms the left-handed current to left-handed, so the theory can be CP-even. If all coupling constants in the SM Lagrangian were real then, being hermitian, the Lagrangian would be CP invariant.

Since coupling constants of charged currents are complex (there is the CKM matrix V) CP invariance is violated. But when complex phases can be absorbed by a redefinition of field operators there is no CPV (the cases of one or two quark-lepton generations).

$$\mathcal{L}_W = \frac{g}{\sqrt{2}}\bar{u}\gamma_\mu\frac{1+\gamma_5}{2}VdW_\mu + \frac{g}{\sqrt{2}}\bar{d}\gamma_\mu\frac{1+\gamma_5}{2}V^+uW_\mu^* \tag{46}$$

$$P\psi = i\gamma_0\psi \ , \ \ P(W_0, W_i) = (W_0, -W_i) \tag{47}$$

$$\bar{u}(\gamma_0, \gamma_i)d \to \bar{u}(\gamma_0, -\gamma_i)d \tag{48}$$

$$\bar{u}(\gamma_0\gamma_5, \gamma_i\gamma_5)d \to \bar{u}(-\gamma_0\gamma_5, \gamma_i\gamma_5)d \tag{49}$$

$$\mathcal{L}_W^P = \frac{g}{\sqrt{2}}\bar{u}\gamma_\mu\frac{1-\gamma_5}{2}VdW_\mu + \frac{g}{\sqrt{2}}\bar{d}\gamma_\mu\frac{1-\gamma_5}{2}V^+uW_\mu^* \ , \tag{50}$$

$$C\psi = \gamma_2\gamma_0\bar{\psi} \ , \ \ C(W_0, W_i) = -(W_0^*, W_i^*) \tag{51}$$

$$\mathcal{L}_W^C = \frac{g}{\sqrt{2}}\bar{d}\gamma_\mu\frac{1-\gamma_5}{2}V^T uW_\mu^* + \frac{g}{\sqrt{2}}\bar{u}\gamma_\mu\frac{1-\gamma_5}{2}V^* dW_\mu \tag{52}$$

$$\mathcal{L}_W^{\mathrm{CP}} = \frac{g}{\sqrt{2}}\bar{d}\gamma_\mu\frac{1+\gamma_5}{2}V^T uW_\mu^* + \frac{g}{\sqrt{2}}\bar{u}\gamma_\mu\frac{1+\gamma_5}{2}V^* dW_\mu \tag{53}$$

Comparing (46) with (53) we see, that for real V $\mathcal{L}_W^{\mathrm{CP}} = \mathcal{L}_W$, and there is no CPV.

Complex V which cannot be made real by fields redefinition $u_i \to e^{i\alpha_i}u_i$, $d_j \to e^{i\beta_j}d_j$ (which is so when $N_{\mathrm{gen}} \geq 3$) – CP is violated.

4 $M^0 - \bar{M}^0$ mixing; CPV in mixing

In order to mix, a meson must be neutral and not coincide with its antiparticle. There are four such pairs:

$$K^0(\bar{s}d) - \bar{K}^0(s\bar{d}) \ , \quad D^0(c\bar{u}) - \bar{D}^0(\bar{c}u) \ \ ,$$
$$B^0_d(\bar{b}d) - \bar{B}^0_d(b\bar{d}) \ \text{ and } \ B^0_s(\bar{b}s) - \bar{B}^0_s(b\bar{s}) \ \ . \tag{54}$$

Mixing occurs in the second order in weak interactions through the box diagram which is shown in Fig. 4 for $K^0 - \bar{K}^0$ pair.

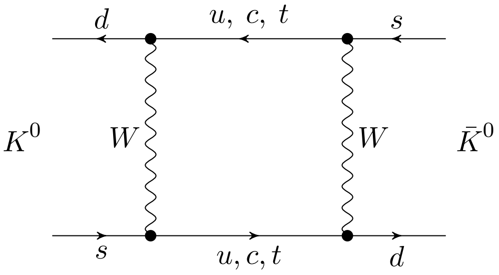


Fig. 4: $K^0 - \bar{K}^0$ transition.

The effective 2×2 Hamiltonian H is used to describe the meson-antimeson mixing. It is most easily written in the following basis:

$$M^0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ \bar{M}^0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{55}$$

The meson-antimeson system evolves according to the Shroedinger equation with this effective Hamiltonian which is not hermitian since it takes meson decays into account. So, $H = M - \frac{i}{2}\Gamma$, where both M and Γ are hermitian. M can be named a mass matrix, and Γ - a matrix of widths.

According to CPT invariance the diagonal elements of H are equal:

$$\langle M^0 \mid H \mid M^0 >=< \bar{M}^0 \mid H \mid \bar{M}^0\rangle \ \ . \tag{56}$$

Substituting into the Shroedinger equation

$$i\frac{\partial\psi}{\partial t} = H\psi \tag{57}$$

ψ – function in the following form:

$$\psi = \begin{pmatrix} p \\ q \end{pmatrix} e^{-i\lambda t} \tag{58}$$

we come to the following equation:

$$\begin{pmatrix} M - \frac{i}{2}\Gamma & M_{12} - \frac{i}{2}\Gamma_{12} \\ & \\ M^*_{12} - \frac{i}{2}\Gamma^*_{12} & M - \frac{i}{2}\Gamma \end{pmatrix} \begin{pmatrix} p \\ \\ q \end{pmatrix} = \lambda \begin{pmatrix} p \\ \\ q \end{pmatrix} \tag{59}$$

from which for eigenvalues ($\lambda_\pm$) and eigenvectors ($M_\pm$) we obtain:

$$\lambda_\pm = M - \frac{i}{2}\Gamma \pm \sqrt{(M_{12} - \frac{i}{2}\Gamma_{12})(M^*_{12} - \frac{i}{2}\Gamma^*_{12})} \ , \tag{60}$$

$$\left\{ \begin{array}{l} M_+ = pM^0 + q\bar{M}^0 \\ M_- = pM^0 - q\bar{M}^0 \end{array} \right. \ , \ \ \frac{q}{p} = \sqrt{\frac{M^*_{12} - \frac{i}{2}\Gamma^*_{12}}{M_{12} - \frac{i}{2}\Gamma_{12}}} \ \ . \tag{61}$$

If there is no CPV in mixing, then:

$$\langle M^0 \mid H \mid \bar{M}^0 \rangle = \langle \bar{M}^0 \mid H \mid M^0 \rangle \ ,$$
$$M_{12} - \frac{i}{2}\Gamma_{12} = M^*_{12} - \frac{i}{2}\Gamma^*_{12} \ , \tag{62}$$

and

$$\frac{q}{p} = 1 \ , \ < M_+ \mid M_- >= 0 \ \text{(in case of kaons } M_+ = K^0_1, \ M_- = K^0_2). \tag{63}$$

However, even if the phases of M_{12} and Γ_{12} are nonzero but equal (modulo π) we can eliminate this common phase rotating M^0.

We observe the one-to-one correspondence between CPV in mixing and non-orthogonality of the eigenstates M_+ and M_-. According to Quantum Mechanics if two hermitian matrices M and Γ commute, then they have a common orthonormal basis. Let us calculate the commutator of M and Γ:

$$[M, \Gamma] = \begin{pmatrix} M_{12}\Gamma^*_{12} - M^*_{12}\Gamma_{12} & 0 \\ 0 & M^*_{12}\Gamma_{12} - M_{12}\Gamma^*_{12} \end{pmatrix} . \tag{64}$$

It equals zero if the phases of M_{12} and Γ_{12} coincide (modulo π). So, for $[M\Gamma] = 0$ we get $\mid q/p \mid= 1$, $< M_+ \mid M_- >= 0$ and there is no CPV in the meson-antimeson mixing. And vice versa.

Problem 4

CPV in kaon mixing. According to the box diagram which describes $K^0 - \bar{K}^0$ mixing $\Gamma_{12} \sim (V^*_{ud}V_{us})^2$. Find an analogous expression for M_{12}. Use unitarity of the matrix V and eliminate $V^*_{cd}V_{cs}$ from M_{12}. Observe that the quantity $M_{12}\Gamma^*_{12} - M^*_{12}\Gamma_{12}$ is proportional to the Jarlskog invariant $J = Im(V^*_{ud}V_{us}V_{td}V^*_{ts})$.

Introducing quantity $\tilde{\varepsilon}$ according to the following definition:

$$\frac{q}{p} = \frac{1-\tilde{\varepsilon}}{1+\tilde{\varepsilon}} \ , \tag{65}$$

we see that if $Re \ \tilde{\varepsilon} \neq 0$, then CP is violated. For the eigenstates we obtain:

$$M_+ = \frac{1}{\sqrt{1+ \mid \tilde{\varepsilon} \mid^2}} \left[\frac{M^0 + \bar{M}^0}{\sqrt{2}} + \tilde{\varepsilon}\frac{M^0 - \bar{M}^0}{\sqrt{2}} \right] \ ,$$
$$M_- = \frac{1}{\sqrt{1+ \mid \tilde{\varepsilon} \mid^2}} \left[\frac{M^0 - \bar{M}^0}{\sqrt{2}} + \tilde{\varepsilon}\frac{M^0 + \bar{M}^0}{\sqrt{2}} \right] \ . \tag{66}$$

If CP is conserved, then $Re \ \tilde{\varepsilon} = 0$, M_+ is CP even and M_- is CP odd. If CP is violated in mixing, then $Re \ \tilde{\varepsilon} \neq 0$ and M_+ and M_- get admixtures of the opposite CP parities and become non-orthogonal.

5 Neutral kaons: mixing (Δm_{LS}) and CPV in mixing ($\tilde{\varepsilon}$)

5.1 $K^0 - \bar{K}^0$ mixing, Δm_{LS}

Γ_{12} for the $K^0 - \bar{K}^0$ system is given by the absorptive part of the diagram in Fig. 5. With our choice of CKM matrix V_{us} and V_{ud} are real, so Γ_{12} is real.

M_{12} is given by a dispersive part of the diagram in Fig. 6. Now all three up quarks should be taken into account.

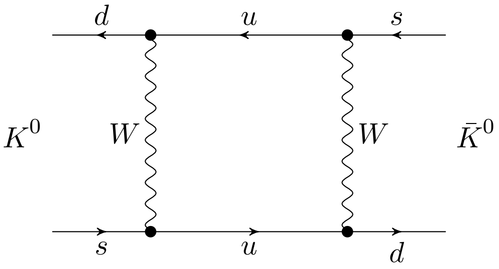


Fig. 5: The diagram which contributes to Γ_{12}.

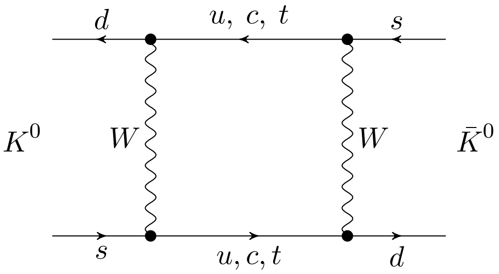


Fig. 6: The diagram which contributes to M_{12}.

To calculate this diagram it is convenient to implement GIM (Glashow-Illiopulos-Maiani) compensation mechanism [4] from the very beginning, subtracting zero from the sum of the fermion propagators:

$$\frac{V_{us}V^*_{ud}}{\hat{p}-m_u} + \frac{V_{cs}V^*_{cd}}{\hat{p}-m_c} + \frac{V_{ts}V^*_{td}}{\hat{p}-m_t} - \frac{\sum\limits_i V_{is}V^*_{id}}{\hat{p}} \quad . \tag{67}$$

Since u-quark is massless with good accuracy, $m_u \approx 0$, then its propagator drops out and we are left with the modified c- and t-quark propagators:

$$\frac{1}{\hat{p}-m_{c,t}} \longrightarrow \frac{m^2_{c,t}}{(p^2-m^2_{c,t})\hat{p}} \quad . \tag{68}$$

The modified fermion propagators decrease in ultraviolet so rapidly that one can calculate the box diagrams in the unitary gauge, where W-boson propagator is $(g_{\mu\nu} - k_\mu k_\nu/M^2_W)/(k^2 - M^2_W)$

We easily get the following estimates for three remaining diagram contributions to M_{12}:

$$\begin{array}{rc} (cc): & \lambda^2(1-2i\eta A^2\lambda^4)G^2_F m^2_c \ , \\ (ct): & \lambda^6(1-\rho+i\eta)G^2_F m^2_c \ln(\frac{m_t}{m_c})^2 \ , \\ (tt): & \lambda^{10}(1-\rho+i\eta)^2 G^2_F m^2_t \ . \end{array} \tag{69}$$

Since $m_c \approx 1.3$ GeV and $m_t \approx 175$ GeV we observe that the cc diagram dominates in ReM_{12} while ImM_{12} is dominated by (tt) diagram.

M_{12} is mostly real:

$$\frac{ImM_{12}}{ReM_{12}} \sim \lambda^8 \left(\frac{m_t}{m_c}\right)^2 \sim 0.1 \quad . \tag{70}$$

The explicit calculation of the cc exchange diagram gives:

$$\mathcal{L}^{\rm eff}_{\Delta s=2} = -\frac{g^4}{2^9\pi^2 M^4_W}(\bar{s}\gamma_\alpha(1+\gamma_s)d)^2\eta_1 m^2_c V^2_{cs}V^{*^2}_{cd} \quad , \tag{71}$$

where g is SU(2) gauge coupling constant, $g^2/8M_W^2 = G_F/\sqrt{2}$, and factor η_1 takes into account the hard gluon exchanges. Since

$$M_{12} - \frac{i}{2}\Gamma_{12} =< K^0 \mid H^{eff} \mid \bar{K}^0 > /(2m_K) \tag{72}$$

(here $H^{eff} = -\mathcal{L}^{eff}_{\Delta s=2}$) we should calculate the matrix element of the product of two $V-A$ quark currents between $\bar{K}^0$ and K^0 states. Using the vacuum insertion we obtain:

$$\begin{gathered}\langle K^0 \mid \bar{s}\gamma_\alpha(1+\gamma_5)d\bar{s}\gamma_\alpha(1+\gamma_5)d \mid \bar{K}^0\rangle = \\ = \frac{8}{3}B_K\langle K^0 \mid \bar{s}\gamma_\alpha(1+\gamma_s)d \mid 0\rangle \cdot \langle 0 \mid \bar{s}\gamma_\alpha(1+\gamma_5)d \mid \bar{K}^0\rangle = -\frac{8}{3}B_K f_K^2 m_K^2 \ ,\end{gathered} \tag{73}$$

where $B_K = 1$ would hold if the vacuum insertion would saturate this matrix element.

From Eq. (60) we obtain:

$$m_S - m_L - \frac{i}{2}(\Gamma_S - \Gamma_L) = 2[ReM_{12} - \frac{i}{2}\Gamma_{12}] \ , \tag{74}$$

where S and L are the abbreviations for K_S and K_L, short and long-lived neutral K-mesons respectively. For the difference of masses we get:

$$m_L - m_S \equiv \Delta m_{LS} = \frac{G_F^2 B_K f_K^2 m_K}{6\pi^2}\eta_1 m_c^2 |V_{cs}^2 V_{cd}^{*2}| \ . \tag{75}$$

Constant f_K is known from $K \to l\nu$ decays, $f_K = 160$ MeV. Gluon dressing of the box diagrams in 4 quark model in the leading logarithmic (LO) approximation gives $\eta_1^{LO} = 0.6$. It appears that the sub-leading logarithms are numerically very important, $\eta_1^{NLO} = 1.3 \pm 0.2$, the number which we will use in our estimates. We take $B_K = 0.8 \pm 0.1$ assuming that the vacuum insertion is good numerically, though the smaller values of B_K can be found in literature as well.

Experimentally the difference of masses is:

$$\Delta m_{LS}^{\text{exp}} = 0.5303(9) \cdot 10^{10}\ \text{sec}^{-1} \ . \tag{76}$$

Substituting the numbers we get:

$$\frac{\Delta m_{LS}^{\text{theor}}}{\Delta m_{LS}^{\text{exp}}} = 0.5 \pm 0.2 \ , \tag{77}$$

and we almost get an experimental number from the short-distance contribution described by the box diagram with c-quarks. Historically this was the first place from which the approximate value of c-quark mass was determined.

However, the very existence of a charm quark and its mass below 2 GeV were predicted *before* 1974 November revolution ($J/\Psi(c\bar{c})$ discovery, $M_{J/\Psi}$ = 3.1 GeV) from the value of Δm_{LS}.

Concerning the neutral kaon decays we have:

$$\Gamma_S - \Gamma_L = 2\Gamma_{12} \approx \Gamma_S = 1.1 \cdot 10^{10}\ \text{sec}^{-1} \ \ (\Delta m_{LS} \approx \Gamma_S/2) \ , \tag{78}$$

since $\Gamma_L \ll \Gamma_S$, $\Gamma_L = 2 \cdot 10^7\ \text{sec}^{-1}$. K_S rapidly decays to two pions which have CP$= +1$.

$D^0 - \bar{D}^0$ mixing is established but it is very small: $\Delta m/\Gamma, \Delta\Gamma/\Gamma \sim 10^{-3}$. One of the reasons is the absence of Cabbibo suppression of c-quark decay, while $D^0 - \bar{D}^0$ transition amplitude is proportional to $\sin^2\theta_c$.

5.2 CPV in $K^0 - \bar{K}^0$: $K_L \to 2\pi$, ε_K-hyperbola

CPV in $K^0 - \bar{K}^0$ mixing is proportional to the deviation of $\mid q/p \mid$ from one; so let us calculate this ratio taking into account that Γ_{12} is real, while M_{12} is mostly real:

$$\frac{q}{p} = 1 - \frac{iImM_{12}}{M_{12} - \frac{i}{2}\Gamma_{12}} = 1 + \frac{2iImM_{12}}{m_L - m_S + \frac{i}{2}\Gamma_S} \quad . \tag{79}$$

In this way for quantity $\tilde{\varepsilon}$ we obtain:

$$\tilde{\varepsilon} = -\frac{iImM_{12}}{\Delta m_{LS} + \frac{i}{2}\Gamma_S} \quad . \tag{80}$$

Branching of CP-violating $K_L \to 2\pi$ decay equals:

$$\begin{aligned} Br(K_L \to 2\pi^0) + Br(K_L \to \pi^+\pi^-) = \frac{\Gamma(K_L \to 2\pi)}{\Gamma_{K_L}} = \frac{\Gamma_{K_L \to 2\pi}}{\Gamma_{K_S \to 2\pi}} \frac{\Gamma(K_S)}{\Gamma(K_L)} = \\ = \frac{\mid \eta_{00} \mid^2 \Gamma(K_S \to 2\pi^0) + \mid \eta_{+-} \mid^2 \Gamma(K_S \to \pi^+\pi^-)}{\Gamma(K_S \to 2\pi^0) + \Gamma(K_S \to \pi^+\pi^-)} \frac{\Gamma(K_S)}{\Gamma(K_L)} \approx \\ \approx \mid \eta_{00} \mid^2 \frac{\Gamma(K_S)}{\Gamma(K_L)} \approx \mid \tilde{\varepsilon} \mid^2 \frac{\Gamma(K_S)}{\Gamma(K_L)} \approx \mid \tilde{\varepsilon} \mid^2 \frac{5.12(2) \cdot 10^{-8} \text{ sec}}{0.895(0.3) \cdot 10^{-10} \text{ sec}} \approx \\ \approx 572 \mid \tilde{\varepsilon} \mid^2 = 2.83(1) \cdot 10^{-3} \quad , \end{aligned} \tag{81}$$

where the last number is the sum of $K_L \to \pi^+\pi^-$ and $K_L \to \pi^0\pi^0$ branching ratios. In this way the experimental value of $\mid \tilde{\varepsilon} \mid$ is determined, and for a theoretical result we should have:

$$\mid \tilde{\varepsilon} \mid = \frac{\mid ImM_{12} \mid}{\sqrt{2}\Delta m_{LS}} = 2.22 \cdot 10^{-3}. \tag{82}$$

As we have already demonstrated, (tt) box gives the main contribution to ImM_{12}. It was calculated for the first time explicitly not supposing that $m_t \ll m_W$ in 1980 [19]:

$$\begin{aligned} ImM_{12} = -\frac{G_F^2 B_K f_K^2 m_K}{12\pi^2} m_t^2 \eta_2 Im(V_{ts}^2 V_{td}^{*^2}) \times I(\xi) \quad , \\ I(\xi) = \left\{ \frac{\xi^2 - 11\xi + 4}{4(\xi - 1)^2} - \frac{3\xi^2 \ln \xi}{2(1-\xi)^3} \right\} \quad , \quad \xi = \left(\frac{m_t}{m_W} \right)^2 \quad , \end{aligned} \tag{83}$$

where factor η_2 takes into account the gluon exchanges in the box diagram with (tt) quarks and in the leading logarithmic approximation it equals $\eta_2^{LO} = 0.6$. This factor is not changed substantially by sub-leading logs: $\eta_2^{NLO} = 0.57(1)$.

Let us present the numerical values for the expression in figure brackets for several values of the top quark mass:

$$\{ \ \} = \begin{array}{ll} 1 \ , & m_t = 0 \ , \ \xi = 0 \\ 0.55 \ , & \xi = 4.7 \ , \ \text{which corresponds to} \ m_t = 175 \text{ GeV} \\ 0.25 \ , & m_t = \xi = \infty \end{array} \tag{84}$$

It is clearly seen that the top contribution to the box diagram is not decoupled: it does not vanish in the limit $m_t \to \infty$. One can easily get where this enhanced at $m_t \to \infty$ behaviour originates by estimating the box diagram in 't Hooft-Feynman gauge. In the limit $m_t \gg m_W$ the diagram with two charged Higgs exchanges dominates (see Fig. 7), since each vertex of Higgs boson emission is proportional to m_t.

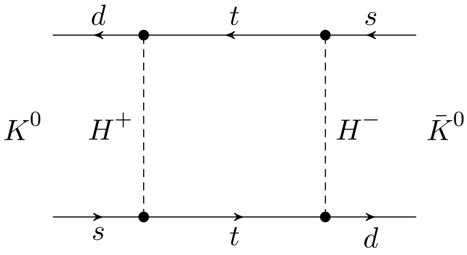


Fig. 7: The diagram which dominates in the limit $m_t \gg m_W$.

For the factor which multiplies the four-quark operator from this diagram we get:

$$\sim (\frac{m_t}{v})^4 \int \frac{d^4p}{(p^2 - M_W^2)^2} \left[\frac{\hat{p}}{p^2 - m_t^2}\right]^2 \sim (\frac{m_t}{v})^4 \frac{1}{m_t^2} = G_F^2 m_t^2 \ , \tag{85}$$

where v is the Higgs boson expectation value. No decoupling!

Substituting the numbers we obtain:

$$\eta(1 - \rho) = 0.47(5) \ , \tag{86}$$

where 10% uncertainty in the value of $B_K = 0.8 \pm 0.1$ dominates in the error. Taking into account (ct) and (cc) boxes we get the following equation:

$$\eta(1.4 - \rho) = 0.47(5) \tag{87}$$

which gives hyperbola on (ρ, η) plane.

Why is ε_K so small? We have the following estimate for ε_K:

$$\varepsilon_K \sim \frac{m_t^2 \lambda^{10} \eta(1 - \rho)}{m_c^2 \lambda^2} \ . \tag{88}$$

It means that ε_K is small not because CKM phase is small, but because 2×2 part of CKM matrix which describes the mixing of the first two generations is almost unitary and the third generation almost decouples. We are lucky that the top quark is so heavy; for $m_t \sim 10$ GeV CPV would not have been discovered in 1964.

6 Direct CPV

6.1 Direct CPV in K decays, $\varepsilon' \neq 0$ ($\mid \frac{\bar{A}}{A} \mid \neq 1$)

Let us consider the neutral kaon decays into two pions. It is convenient to deal with the amplitudes of the decays into the states with a definite isospin:

$$A(K^0 \to \pi^+\pi^-) = \frac{a_2}{\sqrt{3}} e^{i\xi_2} e^{i\delta_2} + \frac{a_0}{\sqrt{3}} \sqrt{2} e^{i\xi_0} e^{i\delta_0} \ , \tag{89}$$

$$A(\bar{K}^0 \to \pi^+\pi^-) = \frac{a_2}{\sqrt{3}} e^{-i\xi_2} e^{i\delta_2} + \frac{a_0}{\sqrt{3}} \sqrt{2} e^{-i\xi_0} e^{i\delta_0} \ , \tag{90}$$

$$A(K^0 \to \pi^0\pi^0) = \sqrt{\frac{2}{3}} a_2 e^{i\xi_2} e^{i\delta_2} - \frac{a_0}{\sqrt{3}} e^{i\xi_0} e^{i\delta_0} \ , \tag{91}$$

$$A(\bar{K}^0 \to \pi^0\pi^0) = \sqrt{\frac{2}{3}}a_2 e^{-i\xi_2}e^{i\delta_2} - \frac{a_0}{\sqrt{3}}e^{-i\xi_0}e^{i\delta_0} \quad , \tag{92}$$

where “2” and “0” are the values of $(\pi\pi)$ isospin, $\xi_{2,0}$ are the weak phases which originate from CKM matrix and $\delta_{2,0}$ are the strong phases of $\pi\pi$-rescattering. If the only quark diagram responsible for $K \to 2\pi$ decays were the charged current tree diagram which describes $s \to u\bar{u}d$ transition through W-boson exchange, then the weak phases would be zero and it would be no CPV in the decay amplitudes (the so-called direct CPV). All CPV would originate from $K^0 - \bar{K}^0$ mixing. Such indirect CPV was called superweak (L.Wolfenstein, 1964).

However, in Standard Model the CKM phase penetrates into the amplitudes of $K \to 2\pi$ decays through the so-called “penguin” diagrams shown in Fig. 8 and ξ_0 and ξ_2 are nonzero leading to direct CPV as well.

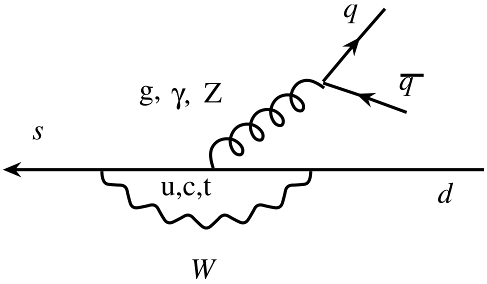


Fig. 8: The penguin diagrams contributing to kaon decays.

From Eqs. (89) and (90) we get:

$$\Gamma(K^0 \to \pi^+\pi^-) - \Gamma(\bar{K}^0 \to \pi^+\pi^-) = -4\frac{\sqrt{2}}{3}a_0 a_2 \sin(\xi_2 - \xi_0)\sin(\delta_2 - \delta_0) \quad , \tag{93}$$

so for direct CPV to occur through the difference of K^0 and $\bar{K}^0$ widths at least two decay amplitudes with different CKM and strong phases should exist.

In the decays of K_L and K_S mesons the violation of CP occurs due to that in mixing (indirect CPV) and in decay amplitudes of K^0 and $\bar{K}^0$ (direct CPV). The first effect is taken into account in the expression for K_L and K_S eigenvectors through K^0 and $\bar{K}^0$:

$$K_S = \frac{K^0 + \bar{K}^0}{\sqrt{2}} + \tilde{\varepsilon}\frac{K^0 - \bar{K}^0}{\sqrt{2}} \quad , \tag{94}$$

$$K_L = \frac{K^0 - \bar{K}^0}{\sqrt{2}} + \tilde{\varepsilon}\frac{K^0 + \bar{K}^0}{\sqrt{2}} \quad , \tag{95}$$

where we neglect $\sim \tilde{\varepsilon}^2$ terms. For the amplitudes of K_L and K_S decays into $\pi^+\pi^-$ we obtain:

$$\begin{aligned} A(K_L \to \pi^+\pi^-) = \frac{1}{\sqrt{2}}\left[\frac{a_2}{\sqrt{3}}e^{i\delta_2}2i\sin\xi_2 + \frac{a_0}{\sqrt{3}}\sqrt{2}e^{i\delta_0}2i\sin\xi_0\right] + \\ + \frac{\tilde{\varepsilon}}{\sqrt{2}}\left[\frac{a_2}{\sqrt{3}}e^{i\delta_2}2\cos\xi_2 + \frac{a_0}{\sqrt{3}}\sqrt{2}e^{i\delta_0}2\cos\xi_0\right] \quad , \end{aligned} \tag{96}$$

$$A(K_S \to \pi^+\pi^-) = \frac{1}{\sqrt{2}}\left[\frac{a_2}{\sqrt{3}}e^{i\delta_2}2\cos\xi_2 + \frac{a_0}{\sqrt{3}}\sqrt{2}e^{i\delta_0}2\cos\xi_0\right] \quad , \tag{97}$$

where in the last equation we omit the terms which are proportional to the product of two small factors, $\tilde{\varepsilon}$ and $\sin\xi_{0,2}$. For the ratio of these amplitudes we get:

$$\eta_{+-} \equiv \frac{A(K_L \to \pi^+\pi^-)}{A(K_S \to \pi^+\pi^-)} = \tilde{\varepsilon} + i\frac{\sin\xi_0}{\cos\xi_0} + \frac{ie^{i(\delta_2-\delta_0)}}{\sqrt{2}}\frac{a_2\cos\xi_2}{a_0\cos\xi_0}\left[\frac{\sin\xi_2}{\cos\xi_2} - \frac{\sin\xi_0}{\cos\xi_0}\right] \quad , \tag{98}$$

where we neglect the terms of the order of $(a_2/a_0)^2 \sin\xi_{0,2}$ because from the $\Delta I = 1/2$ rule in K-meson decays it is known that $a_2/a_0 \approx 1/22$.

The analogous treatment of $K_{L,S} \to \pi^0\pi^0$ decay amplitudes leads to:

$$\eta_{00} \equiv \frac{A(K_L \to \pi^0\pi^0)}{A(K_S \to \pi^0\pi^0)} = \tilde{\varepsilon} + i\frac{\sin\xi_0}{\cos\xi_0} - ie^{i(\delta_2-\delta_0)}\sqrt{2}\frac{a_2\cos\xi_2}{a_0\cos\xi_0}\left[\frac{\sin\xi_2}{\cos\xi_2} - \frac{\sin\xi_0}{\cos\xi_0}\right] \quad . \tag{99}$$

The difference of $\eta_{\pm}$ and η_{00} is proportional to ε':

$$\varepsilon' \equiv \frac{i}{\sqrt{2}}e^{i(\delta_2-\delta_0)}\frac{a_2\cos\xi_2}{a_0\cos\xi_0}\left[\frac{\sin\xi_2}{\cos\xi_2} - \frac{\sin\xi_0}{\cos\xi_0}\right] = \tag{100}$$

$$= \frac{i}{\sqrt{2}}e^{i(\delta_2-\delta_0)}\frac{ReA_2}{ReA_0}\left[\frac{ImA_2}{ReA_2} - \frac{ImA_0}{ReA_0}\right] = \frac{i}{\sqrt{2}}e^{i(\delta_2-\delta_0)}\frac{1}{ReA_0}\left[ImA_2 - \frac{1}{22}ImA_0\right] \quad ,$$

where $A_{2,0} \equiv e^{i\xi_{2,0}}a_{2,0}$.

Introducing quantity ε according to the standard definition

$$\varepsilon = \tilde{\varepsilon} + i\frac{ImA_0}{ReA_0} \quad , \tag{101}$$

we obtain:

$$\eta_{+-} = \varepsilon + \varepsilon' \ , \ \ \eta_{00} = \varepsilon - 2\varepsilon' \ . \tag{102}$$

The double ratio η_{+-}/η_{00} was measured in the experiment and its difference from 1 demonstrates direct CPV in kaon decays:

$$\left(\frac{\varepsilon'}{\varepsilon}\right)^{\exp} = (1.67 \pm 0.23) \cdot 10^{-3} \ . \tag{103}$$

The smallness of this ratio is due to (1) the smallness of the phases produced by the penguin diagrams and (2) smallness of the ratio $a_2/a_0 \approx ReA_2/ReA_0$.

Let us estimate the numerical value of ε'. The penguin diagram with the gluon exchange generates $K \to 2\pi$ transition with $\Delta I = 1/2$; those with γ- and Z-exchanges contribute to $\Delta I = 3/2$ transitions as well. The contribution of electroweak penguins being smaller by the ratio of squares of coupling constants is enhanced by the factor $ReA_0/ReA_2 = 22$, see the last part in equation for ε'. As a result the partial compensation of QCD and electroweak penguins occurs. In order to obtain an order of magnitude estimate let us take into account only QCD penguins. We obtain the following estimate for the sum of the loops with t- and c-quarks:

$$|\ \varepsilon'\ | \approx \frac{1}{22\sqrt{2}}\frac{\sin\xi_0}{\cos\xi_0} = \frac{1}{22\sqrt{2}}\frac{\alpha_s(m_c)}{12\pi}\ln(\frac{m_t}{m_c})^2 A^2\lambda^4\eta \approx 2 * 10^{-5}\frac{\alpha_s(m_c)}{12\pi}\ln(\frac{m_t}{m_c})^2 \ . \tag{104}$$

Taking into account that $|\ \varepsilon\ | \approx 2.4 \cdot 10^{-3}$ we see that the smallness of the ratio of ε'/ε can be readily understood.

In order to make an accurate calculation of ε'/ε one should know the matrix elements of the quark operators between K-meson and two π-mesons. Unfortunately at low energies our knowledge of QCD is not enough for such a calculation. That is why a horizontal strip to which an apex of the unitarity triangle should belong according to equation for ε'/ε has too large width and usually is not shown. Nevertheless we have discussed direct CPV since it will be important for B and D-mesons.

6.2 Direct CP asymmetries in $D^0(\bar{D}^0) \to \pi^+\pi^-, \quad K^+K^-$

The following result was reported by LHCb collaboration in 2019 [17]:

$$\Delta A_{CP} = A_{CP}(K^+K^-) - A_{CP}(\pi^+\pi^-) = (-15.4 \pm 2.9) \times 10^{-4}, \tag{105}$$

where CP asymmetry is defined as

$$A_{CP}(f) = \frac{\Gamma(D^0 \to f) - \Gamma(\bar{D}^0 \to f)}{\Gamma(D^0 \to f) + \Gamma(\bar{D}^0 \to f)}. \tag{106}$$

To distinguish D_0 from $\bar{D}_0$ the tagging by the charge of pions in $D^{*+} \to D^0\pi^+, D^{*-} \to \bar{D}^0\pi^-$ decays and by the charge of muon in semileptonic $\bar{B} \to D^0\mu^-\bar{\nu}_\mu X$ decays has been performed.

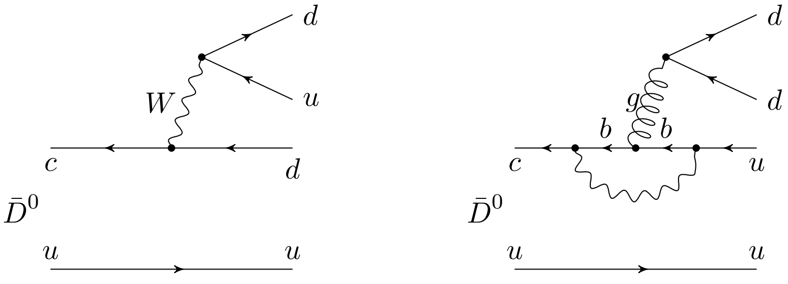


Fig. 9: The diagrams responsible for $\bar{D}^0 \to \pi^+\pi^-$ decay.

The interference of tree and penguin amplitudes shown in Fig. 9 leads to CP asymmetry:

$$A(\bar{D}) = e^{i\delta}TV_{cd}V_{ud}^* - PV_{cb}|V_{ub}|e^{i\gamma}, \tag{107}$$

$$A(D) = e^{i\delta}TV_{cd}^*V_{ud} - PV_{cb}^*|V_{ub}|e^{-i\gamma}, \tag{108}$$

$$A_{CP}(\pi^+\pi^-) = \frac{4TPV_{cd}V_{ud}^*|V_{ub}|V_{cb}^*\sin(\delta)\sin(\gamma)}{2T^2|V_{cd}V_{ud}|^2}. \tag{109}$$

In the limit of U-spin ($d \leftrightarrow s$) symmetry $A_{CP}(K^+K^-) = -A_{CP}(\pi^+\pi^-)$, and the sign “-" comes from $V_{cd} = -V_{us}$. Thus we get:

$$|\Delta A_{CP}| = 4|P/TA^2\lambda^4\sqrt{\rho^2+\eta^2}\sin(\delta)\sin(\gamma)| \approx |25\sin(\delta)P/T| \times 10^{-4}, \tag{110}$$

and to reproduce an experimental result the strong interactions phase δ should be big and the penguin amplitude should be of the order of the tree one.

The reason for the small value of CPV asymmetry in charm is the same as in K- mesons: the 2×2 part of the CKM matrix which describes the mixing of the first and second generations is almost unitary. The absence of $\Delta I = 1/2$ amplitude enhancement makes direct CPV asymmetry in the case of D decays larger than in kaon decays.

When the third generation is involved CPV can be big.

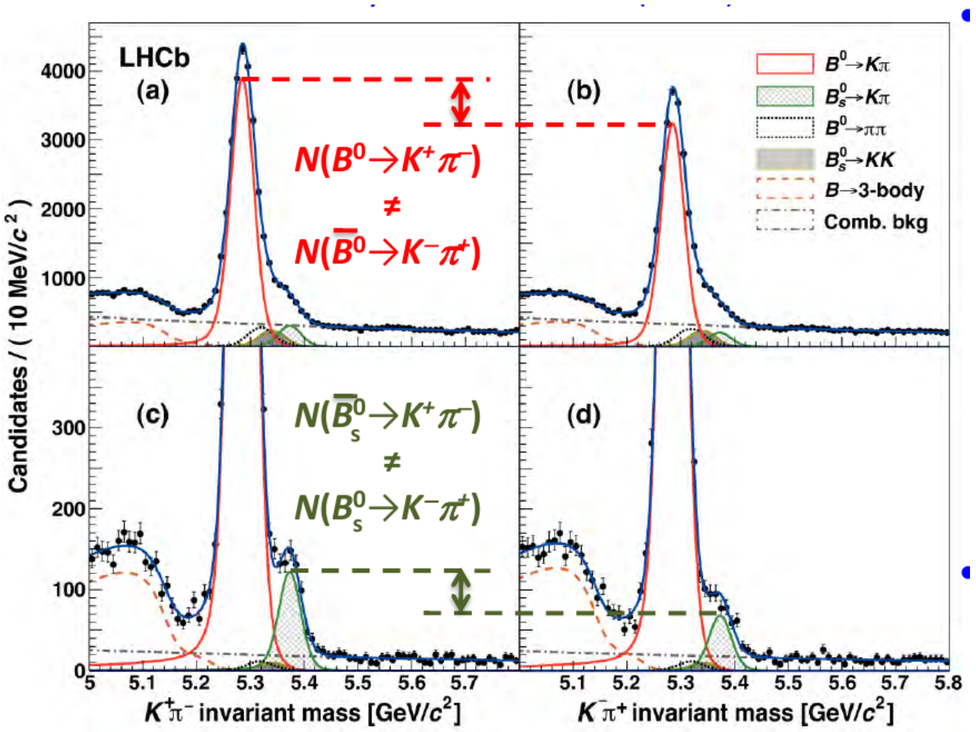


Fig. 10: Direct CPV in $B^0(B^0_s) \to K\pi$ decays.

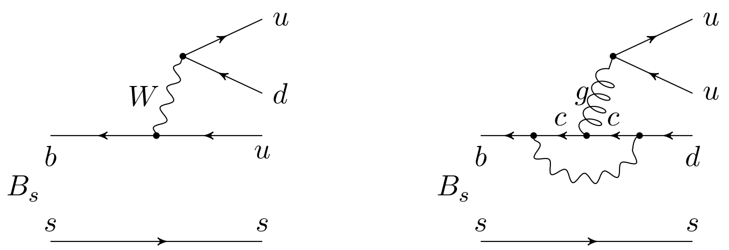


Fig. 11: $B_s \to K^-\pi^+$ decay.

6.3 25 % direct CP asymmetry in B_s decay

While direct CPV in kaons and D-mesons is very small it is sometimes huge in B-mesons, see Fig. 10 [20].

The diagrams shown in Fig. 11 describe $B_s \to K^-\pi^+$ decay.

$$A(B_s \longrightarrow K^-\pi^+) = T_s V^*_{ub} V_{ud} + P_s e^{i\delta} V^*_{cb} V_{cd}, \tag{111}$$

$$A(\bar{B}_s \longrightarrow K^+\pi^-) = T_s V_{ub} V^*_{ud} + P_s e^{i\delta} V_{cb} V^*_{cd}, \tag{112}$$

where δ is the strong phase; the CKM phase is contained in $V_{ub} = -e^{-i\gamma}|V_{ub}|$.

$$A_{CP}(B_s \longrightarrow K^-\pi^+) = \frac{|A(\bar{B}_s)|^2 - |A(B_s)|^2}{|A(\bar{B}_s)|^2 + |A(B_s)|^2} = \tag{113}$$

$$= \frac{4T_s P_s V^*_{ud} V_{cb} V^*_{cd} |V_{ub}| \sin(\delta)\sin(\gamma)}{2T_s^2 |V_{ub} V_{ud}|^2 + 2P_s^2 |V_{cb} V_{cd}|^2 - 4P_s T_s V^*_{ud} V_{cb} V^*_{cd} |V_{ub}| \cos(\delta)\cos(\gamma)}.$$

CKM factors in the nominator and denominator are of the order of λ^6 and there is no CKM suppression of $A_{CP}(B_s)$. Since the asymmetry is big, P_s/T_s is not that small.

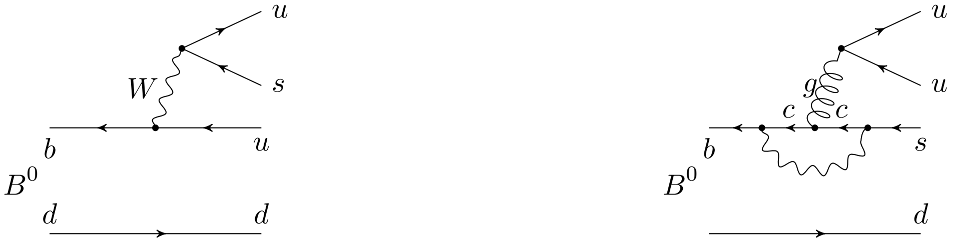


Fig. 12: $B^0 \to K^+\pi^-$ decay.

Though we cannot compute the diagrams in Figs. 11 and 12, we can relate them in the U spin invariance approximation.

Problem 5

Derive an expression for $A_{CP}(B^0 \longrightarrow K^+\pi^-)$ and get the following equality:

$$A_{CP}(B^0) \cdot \Gamma_{B^0 \to K\pi} = -A_{CP}(B_s) \cdot \Gamma_{B_s \to K\pi} \quad . \tag{114}$$

Substituting experimentally measured numbers from RPP (PDG) [21] for the asymmetries $A_{CP}(B^0) = -0.082(6)$, $A_{CP}(B_s) = 0.26(4)$ and branching ratios Br($B^0 \to K\pi) = 20 \cdot 10^{-6}$, Br($B_s \to K\pi) = 5.7 \cdot 10^{-6}$ check this equality.

The smallness of the branching ratio of any exclusive decay is the main problem in studying CPV in B-mesons.

6.4 CPV in neutrino oscillations

In order to have CPV we need not only a CP violating phase but a CP conserving phase as well ($i\Gamma_{12}$ in the case of mixing, $\delta_2 - \delta_0$ in the case of direct CPV in kaon decays).

Problem 6

In the case of leptons the flavor mixing is described by the PMNS matrix:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} V_{e1} & V_{e2} & V_{e3} \\ V_{\mu 1} & V_{\mu 2} & V_{\mu 3} \\ V_{\tau 1} & V_{\tau 2} & V_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \quad . \tag{115}$$

CPV means in particular that the probability of $\nu_\mu \longrightarrow \nu_e$ oscillation $P_{e\mu}$ does not coincide with the probability of $\bar{\nu}_\mu \longrightarrow \bar{\nu}_e$ oscillation $P_{\bar{e}\bar{\mu}}$.

Check that

$$P_{e\mu} - P_{\bar{e}\bar{\mu}} = 4Im(V_{\mu 1}^* V_{e1} V_{\mu 2} V_{e2}^*) * [\sin(\frac{\Delta m_{12}^2}{2E}x) + \sin(\frac{\Delta m_{31}^2}{2E}x) + \sin(\frac{\Delta m_{23}^2}{2E}x)]. \tag{116}$$

Just like in kaons CPV is proportional to the Jarlskog invariant.

When two neutrinos have equal masses there is no CPV.

Where is the CP conserving phase in the case of CPV in neutrino oscillations?

By the way, the driving force for Bruno Pontecorvo to consider neutrino oscillations was the observation of oscillations of neutral kaons [22].

6.5 CPV - absolute notion of a particle

$$\delta_L = \frac{\Gamma(K_L \to \pi^- e^+ \nu) - \Gamma(K_L \to \pi^+ e^- \bar{\nu})}{\Gamma(K_L \to \pi^- e^+ \nu) + \Gamma(K_L \to \pi^+ e^- \bar{\nu})} = 2Re\tilde{\varepsilon} \approx 3.3 * 10^{-3}. \quad (117)$$

Pions of low energies mostly produce K^0 on the Earth, while $\bar{K}^0$ on the "antiEarth" ($\pi N \to K^0(\Lambda, \Sigma)$; $\pi \bar{N} \to \bar{K}^0(\bar{\Lambda}, \bar{\Sigma})$). However, in both cases K_L decay (a little bit) more often into positrons than into electrons.

"The atoms on the Earth contain antipositrons (electrons) - and what about your planet?"

In this way the measurements of the probabilities of semileptonic K_L decays allow to decide if the other planet is made from antimatter.

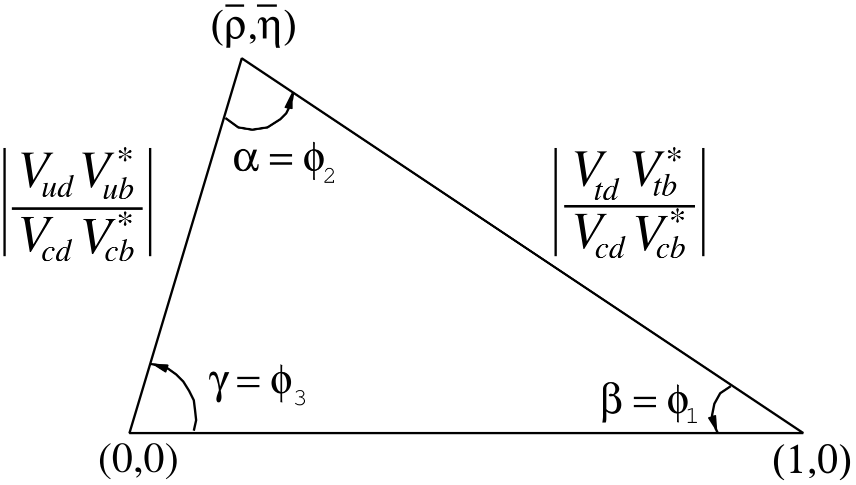
Problem 7

Violation of leptonic (muon and electron) numbers due to neutrino mixing. Estimate the branching ratio of the $\mu \longrightarrow e\gamma$ decay, which occurs in the Standard Model due to the analog of the penguin diagram from Fig. 8 without splitting of the photon.

7 Constraints on the unitarity triangle

7.1 Parameters of the CKM matrix

Four quantities are needed to specify the CKM matrix: s_{12}, s_{13}, s_{23} and δ, or λ, A, ρ, η. The areas shaded in Fig. 13 [23] show the domains of $\bar{\rho}$ and $\bar{\eta}$ allowed at 95% C.L. by different measurements ($\bar{\rho} \equiv \rho(1-\lambda^2/2)$, $\bar{\eta} \equiv \eta(1-\lambda^2/2)$).


$(\bar{\rho},\bar{\eta})$
$\alpha = \phi_2$
$\left|\frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*}\right|$
$\left|\frac{V_{td} V_{tb}^*}{V_{cd} V_{cb}^*}\right|$
$\gamma = \phi_3$
$\beta = \phi_1$
(0,0)
(1,0)

7.2 V_{cd}, V_{cb}, V_{ub}

The precise value of V_{us} follows from the extrapolation of the formfactor of $K \to \pi e \nu$ decay $f_+(q^2)$ to the point $q^2 = 0$, where q is the lepton pair momentum. Due to the Ademollo-Gatto theorem [24] the corrections to the CVC value $f_+(0) = 1$ are of the second order of flavor SU(3) violation, and these

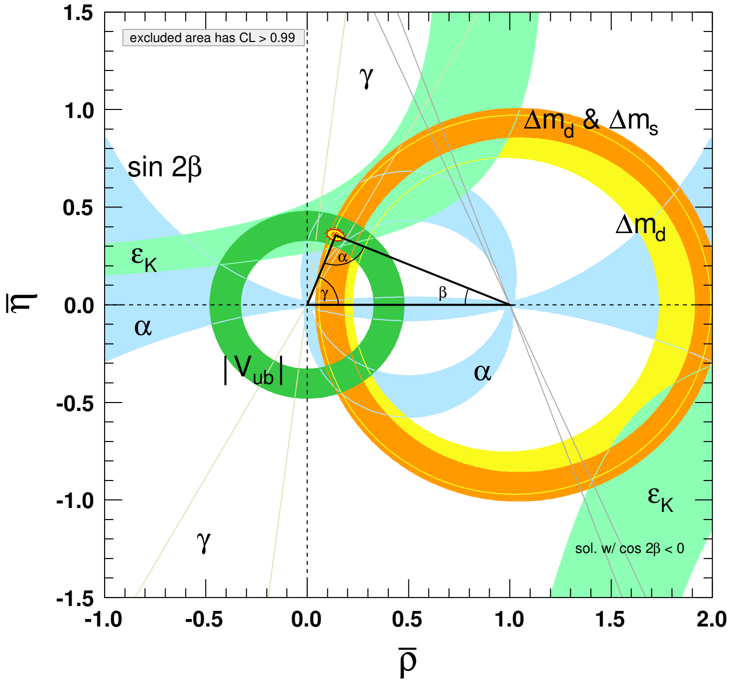


Fig. 13: Constraints on the apex of the unitarity triangle.

small terms were calculated. (For the case of isotopic SU(2) violation a similar theorem was proved in Ref. [25]). As a result of this (and other) analyses PDG gives the following value: $V_{us} \equiv \lambda = 0.2243(5)$.

The accuracy of λ is high: the other parameters of CKM matrix are known much worse. V_{cd} is measured in the processes with c-quark with an order of magnitude worse accuracy: $V_{cd} = 0.218(4)$.

The value of V_{cb} is determined from the inclusive and exclusive semileptonic decays of B-mesons to charm. At the level of quarks $b \to c l \nu$ transition is responsible for these decays: $V_{cb} = (42.2 \pm 0.8) \cdot 10^{-3}$.

The value of $|V_{ub}|$ is extracted from the semileptonic B-mesons decays without the charmed particles in the final state which originated from $b \to u l \nu$ transition: $V_{ub} = (3.94 \pm 0.36) \cdot 10^{-3}$.

The apex of the unitarity triangle should belong to a circle on $(\bar\rho, \bar\eta)$ plane with the center at the point $(0, 0)$. The area between such two circles (deep green color) corresponds to the domain allowed at 2σ.

7.3 $\varepsilon_K, \Delta m_{B^0}, \Delta m_{B^0_s}$

CPV in kaon mixing determines the hyperbola shown by light green color in Fig. 13, see Eq. (87).

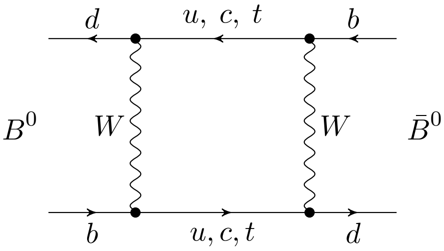


Fig. 14: B^0 – $\bar B^0$ mixing.

In the Standard Model the $B_d - \bar B_d$ transition occurs through the box diagram shown in Fig. 14.

Unlike the case of the $K^0 - \bar{K}^0$ transition the power of λ is the same for u, c and t quarks inside a loop, so the diagram with t-quarks dominates.

Calculating it in complete analogy with the K-meson case we get:

$$M_{12} = -\frac{G_F^2 B_{B_d} f_{B_d}^2}{12\pi^2} m_B m_t^2 \eta_B V_{tb}^2 V_{td}^{*^2} I(\xi) \ , \tag{118}$$

where $I(\xi)$ is the same function as that for K-mesons, and $\eta_B = 0.55 \pm 0.01$ (NLO).

Γ_{12} is determined by the absorptive part of the same diagram (so, 4 diagrams altogether: uu, uc, cu, cc quarks in the inner lines). The result of the calculation is:

$$\Gamma_{12} = \frac{G_F^2 B_{B_d} f_{B_d}^2 m_B^3}{8\pi}[V_{cb}V_{cd}^*(1 + O(\frac{m_c^2}{m_b^2})) + V_{ub}V_{ud}^*]^2 \ , \tag{119}$$

where the term $O(m_c^2/m_b^2)$ accounts for the nonzero c-quark mass.

Using the unitarity of the CKM matrix we get:

$$\Gamma_{12} = \frac{G_F^2 B_{B_d} f_{B_d}^2 m_B^3}{8\pi}[-V_{tb}V_{td}^* + O(\frac{m_c^2}{m_b^2})V_{cb}V_{cd}^*]^2 \ , \tag{120}$$

and the main term in Γ_{12} has the same phase as the main term in M_{12}. That is why CPV in the mixing of B-mesons is suppressed by an extra factor $(m_c/m_b)^2$ and is small. For the difference of masses of the two eigenstates from

$$M_+ - M_- - \frac{i}{2}(\Gamma_+ - \Gamma_-) = 2\sqrt{(M_{12} - \frac{i}{2}\Gamma_{12})(M_{12}^* - \frac{i}{2}\Gamma_{12}^*)} \tag{121}$$

we obtain:

$$\Delta m_{B^0} = -\frac{G_F^2 B_{B_d} f_B^2}{6\pi^2} m_B m_t^2 \eta_B \mid V_{tb}^2 V_{td}^{*^2} \mid I(\xi), \tag{122}$$

and Δm_{B^0} is negative as well as in the kaon system: a heavier state has a smaller width.

7.4 Δm_{B^0} and semileptonic $B^0(\bar{B}^0)$ decays

The B-meson semileptonic decays are induced by a semileptonic b-quark decay, $b \to c l^- \nu \ \ (u l^- \nu)$. In this way in the decays of $\bar{B}^0$ mesons l^- are produced, while in the decays of B^0 mesons l^+ are produced. However, B^0 and $\bar{B}^0$ are not the mass eigenstates and being produced at $t = 0$ they start to oscillate according to the following formulas:

$$B^0(t) = \frac{e^{-i\lambda_+ t} + e^{-i\lambda_- t}}{2} B^0 + \frac{q}{p}\frac{e^{-i\lambda_+ t} - e^{-i\lambda_- t}}{2}\bar{B}^0 \ , \tag{123}$$

$$\bar{B}^0(t) = \frac{e^{-i\lambda_+ t} + e^{-i\lambda_- t}}{2} \bar{B}^0 + \frac{p}{q}\frac{e^{-i\lambda_+ t} - e^{-i\lambda_- t}}{2} B^0 \ . \tag{124}$$

That is why in their semileptonic decays the "wrong sign leptons" are sometimes produced, l^- in the decays of the particles born as B^0 and l^+ in the decays of the particles born as $\bar{B}^0$. The number of these "wrong sign" events depends on the ratio of the oscillation frequency Δm and B-meson lifetime Γ (unlike the case of K-mesons for B-mesons $\Delta\Gamma \ll \Gamma$). For $\Delta m \gg \Gamma$ a large number of oscillations occurs, and the number of "the wrong sign leptons" equals that of a normal sign. If $\Delta m \ll \Gamma$, then B-mesons decay before they start to oscillate.

The pioneering detection of "the wrong sign events" by ARGUS collaboration in 1987 demonstrated that Δm is of the order of Γ, which in the framework of Standard Model could be understood only if the top quark is unusually heavy, $m_t \geq 100$ GeV [26]. Fast $B^0 - \bar{B}^0$ oscillations made possible the construction of asymmetric B-factories (suggested in [27]) where CPV in B^0 decays was observed. (Let us mention that UA1 collaboration saw the events which were interpreted as a possible manifestation of $B_s^0 - \bar{B}_s^0$ oscillations [28].)

Integrating the probabilities of B^0 decays in l^+ and l^- over t, we obtain for "the wrong sign lepton" probability:

$$W_{B^0 \to \bar{B}^0} \equiv \frac{N_{B^0 \to l^- X}}{N_{B^0 \to l^- X} + N_{B^0 \to l^+ X}} = \frac{\mid \frac{q}{p} \mid^2 (\frac{\Delta m}{\Gamma})^2}{2 + (\frac{\Delta m}{\Gamma})^2 + \mid \frac{q}{p} \mid^2 (\frac{\Delta m}{\Gamma})^2} \quad , \tag{125}$$

where we neglect $\Delta\Gamma$, the difference of B_+- and B_--mesons lifetimes. Precisely according to our discussion for $\Delta m/\Gamma \gg 1$ we have $W = 1/2$, while for $\Delta m/\Gamma \ll 1$ we have $W = 1/2(\Delta m/\Gamma)^2$ (with high accuracy $\mid p/q \mid = 1$).

For $\bar{B}^0$ decays we get the same formula with the interchange of q and p.

In ARGUS experiment B-mesons were produced in $\Upsilon(4S)$ decays: $\Upsilon(4S) \to B\bar{B}$. Υ resonances have $J^{PC} = 1^{--}$, that is why (pseudo)scalar B-mesons are produced in P-wave. It means that $B\bar{B}$ wave function is antisymmetric at the interchange of B and $\bar{B}$. This fact forbids the configurations in which due to $B - \bar{B}$ oscillations both mesons become B, or both become $\bar{B}$. However, after one of the B-meson decays the flavor of the remaining one is tagged, and it oscillates according to Eqs. (123) and (124).

If the first decay is semileptonic with l^+ emission indicating that a decaying particle was B^0, then the second particle was initially $\bar{B}^0$. Thus taking $\mid p/q \mid = 1$ we get for the relative number of the same sign dileptons born in semileptonic decays of B-mesons, produced in $\Upsilon(4S) \to B\bar{B}$ decays:

$$\frac{N_{l^+l^+} + N_{l^-l^-}}{N_{l^+l^-}} = \frac{W}{1-W} = \frac{x^2}{2+x^2} \; , \quad x \equiv \frac{\Delta m}{\Gamma} \quad . \tag{126}$$

Let us note that if B^0 and $\bar{B}^0$ are produced incoherently (say, in hadron collisions) a different formula should be used:

$$\frac{N_{l^+l^+} + N_{l^-l^-}}{N_{l^+l^-}} = \frac{2W - 2W^2}{1 - 2W + 2W^2} = \frac{x^2(2+x^2)}{2 + 2x^2 + x^4} \quad . \tag{127}$$

In the absence of oscillations ($x = 0$) both equations give zero; for high frequency oscillations ($x \gg 1$) both of them give one.

From the time integrated data of ARGUS and CLEO $W_d = 0.182 \pm 0.015$ follows. From the time-dependent analysis of B-decays at the high energy colliders (LEP II, Tevatron, SLC, LHC) and the time-dependent analysis at the asymmetric B-factories Belle and BaBar the following result was obtained :

$$x_d = 0.770(4) \quad . \tag{128}$$

By using the life time of B_d-mesons: $\Gamma_{B_d} = [1.52(1) \cdot 10^{-12} \text{ sec}]^{-1} \equiv [1.52(1)\text{ps}]^{-1}$ we get for the mass difference of B_d mesons:

$$\Delta m_d = 0.506(2)\text{ps}^{-1} \;\; \text{or, equivalently,} \, W_d = 0.1874 \pm 0.0018. \tag{129}$$

This Δm_d value can be used in Eq. (122) to extract the value of $|V_{td}|$. The main uncertainty is in a hadronic matrix element $f_{B_d}\sqrt{B_{B_d}} = 216 \pm 15$ MeV obtained from the lattice QCD calculations.

7.5 $\Delta m_{B_s^0}$

The theoretical uncertainty diminishes in the ratio

$$\frac{\Delta m_s}{\Delta m_d} = \frac{m_{B_s}}{m_{B_d}} \xi^2 \frac{|V_{ts}|^2}{|V_{td}|^2}, \tag{130}$$

where $\xi = (f_{B_s}\sqrt{B_{B_s}})/(f_{B_d}\sqrt{B_{B_d}}) = 1.24 \pm 0.05$.

Since the lifetimes of B_d - and B_s -mesons are almost equal, we get:

$$x_s \approx x_d \frac{|V_{ts}|^2}{|V_{td}|^2} \tag{131}$$

which means $x_s \gg 1$ and very fast oscillations. That is why W_{B_s} equals $1/2$ with very high accuracy and one cannot extract x_{B_s} from the time integrated measurements.

$B_s^0 - \bar{B}_s^0$ oscillations were first observed at Tevatron. The average of all published measurements

$$\Delta m_{B_s^0} = 17.757 \pm 0.020(\text{stat}) \pm 0.007(\text{syst}) \ \ (\text{ps}^{-1}) \tag{132}$$

is dominated by LHCb.

Thus we get

$$|V_{td}/V_{ts}| = 0.210 \pm 0.001(\text{exp}) \pm 0.008(\text{theor}), \tag{133}$$

which corresponds to yellow (only Δm_d) and brown (Δm_d and Δm_s) circles in Fig. 13.

What remains are the values of the angles of the unitarity triangle, which are determined by CP-violation measurements in B-meson decays. Soon we will go there.

7.6 $\Delta\Gamma/\Gamma$

For the difference of the width of B_{dL} and B_{dH} we obtain

$$\Delta\Gamma_{B_d} = 2\Gamma_{12} \approx \frac{G_F^2 B_{B_d} f_B^2 m_B^3}{4\pi} \mid V_{td} \mid^2 \ \ , \tag{134}$$

which is very small:

$$\frac{\Delta\Gamma_{B_d}}{\Gamma_{B_d}} < 1\% \quad , \tag{135}$$

as opposite to K-meson case, where K_S and K_L lifetimes differ strongly.

In the B_s-meson system a larger time difference was expected; substituting V_{ts} instead of V_{td} we obtain:

$$\frac{\Delta\Gamma_{B_s}}{\Gamma_{B_s}} \sim 10\% \ \ . \tag{136}$$

Here are the experimental results:

$$\Gamma_{B_{sL}^0} = (1.414(10)\text{ps})^{-1} \tag{137}$$

$$\Gamma_{B_{sH}^0} = (1.624(14)\text{ps})^{-1}, \tag{138}$$

where L is light, H - heavy.

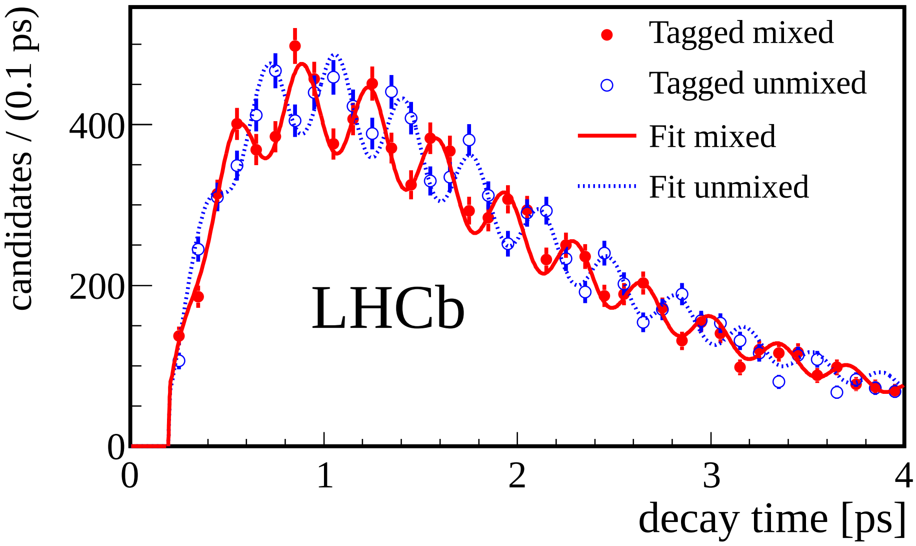


Fig. 15: $B_s - \bar{B}_s$ oscillations [29].

8 CPV in $B^0 - \bar{B}^0$ mixing

For a long time CPV in K-mesons was observed only in $K^0 - \bar{K}^0$ mixing. That is why it seems reasonable to start studying CPV in B-mesons from their mixing:

$$\left|\frac{q}{p}\right| = \left|\sqrt{1+\frac{i}{2}\left(\frac{\Gamma_{12}}{M_{12}} - \frac{\Gamma^*_{12}}{M^*_{12}}\right)}\right| = \left|1+\frac{i}{4}\left(\frac{\Gamma_{12}}{M_{12}} - \frac{\Gamma^*_{12}}{M^*_{12}}\right)\right| =$$
$$= 1 - \frac{1}{2}\mathrm{Im}\left(\frac{\Gamma_{12}}{M_{12}}\right) \approx 1 - \frac{m_c^2}{m_t^2}\mathrm{Im}\frac{V_{cb}V^*_{cd}}{V_{tb}V^*_{td}} \approx 1 - O(10^{-4}) \ . \tag{139}$$

We see that CPV in $B_d - \bar{B}_d$ mixing is very small because the t-quark is very heavy and CPV is even smaller in $B_s - \bar{B}_s$ mixing.

The experimental observation of $B_d - \bar{B}_d$ mixing comes from the detection of same sign leptons produced in the semileptonic decays of $B_d - \bar{B}_d$ pairs from $\Upsilon(4S)$ decay. Due to CPV in the mixing the number of l^-l^- events will differ from that of l^+l^+ and this difference is proportional to $|\frac{q}{p}| - 1 \sim 10^{-4}$:

$$A^B_{SL} = \frac{N(\bar{B}^0 \to l^+X) - N(B^0 \to l^-X)}{N(\bar{B}^0 \to l^+X) + N(B^0 \to l^-X)} = O(10^{-4}). \tag{140}$$

The experimental number is:

$$A^{B_d}_{SL} = 0.0021 \pm 0.0017 \ , \tag{141}$$

or

$$|q/p|_{B_d} = 1.0010 \pm 0.0008 \ . \tag{142}$$

This result shows no evidence of CPV and does not constrain the SM.

9 CPV in interference of mixing and decays, $B^0(\bar{B}^0) \to J/\Psi K$, and the angle β

9.1 General formulae

As soon as it became clear that CPV in $B - \bar{B}$ mixing is small theoreticians started to look for another way to find CPV in B decays. The evident alternative is the direct CPV. It is very small in K-mesons because:

a) the third generation almost decouples in K decays; b) due to $\Delta I = 1/2$ rule. Since in B-meson decays all three quark generations are involved and there are many different final states, large direct CPV does occur [30] - [33]. An evident drawback of this strategy: a branching ratio of B-meson decays into any particular exclusive hadronic mode is very small (just because there are many modes available), so a large number of B-meson decays are needed. The specially constructed asymmetric e^+e^--factories Belle (1999-2010) and BaBar (1999-2008) working at the invariant mass of $\Upsilon(4S)$ discovered CPV in $B^0(\bar{B}^0)$ decays in 2001 [16].

The time evolution of the states produced at $t = 0$ as B^0 or $\bar{B}^0$ is described by Eqs. (123) and (124). It is convenient to present these formulae in a little bit different form:

$$\mid B^0(t) >= e^{-i\frac{M_+ + M_-}{2}t - \frac{\Gamma t}{2}} \left[\cos(\frac{\Delta m t}{2}) \mid B^0 > + i\frac{q}{p}\sin(\frac{\Delta m t}{2}) \mid \bar{B}^0 >\right] \ , \qquad (143)$$

$$\mid \bar{B}^0(t) >= e^{-i\frac{M_+ + M_-}{2}t - \frac{\Gamma t}{2}} \left[+ i\frac{p}{q}\sin(\frac{\Delta m t}{2}) \mid B^0 > + \cos(\frac{\Delta m t}{2}) \mid \bar{B}^0 >\right] \ , \qquad (144)$$

where $\Delta m \equiv M_- - M_+ > 0$, and we take $\Gamma_+ = \Gamma_- = \Gamma$ neglecting their small difference (which should be accounted for in case of B_s).

Let us consider a decay in some final state f. Introducing the decay amplitudes according to the following definitions:

$$A_f = A(B^0 \to f) \ , \ \ \bar{A}_f = A(\bar{B}_0 \to f) \ \ , \qquad (145)$$

$$A_{\bar{f}} = A(B^0 \to \bar{f}) \ , \ \ \bar{A}_{\bar{f}} = A(\bar{B}_0 \to \bar{f}) \ \ , \qquad (146)$$

for the decay probabilities as functions of time we obtain:

$$P_{B^0 \to f}(t) = e^{-\Gamma t} \mid A_f \mid^2 \left[\cos^2(\frac{\Delta m t}{2}) + \left|\frac{q\bar{A}_f}{pA_f}\right|^2 \sin^2(\frac{\Delta m t}{2}) - \mathrm{Im}\left(\frac{q\bar{A}_f}{pA_f}\right)\sin(\Delta m t)\right] \ , \qquad (147)$$

$$P_{\bar{B}^0 \to \bar{f}}(t) = e^{-\Gamma t} \mid \bar{A}_{\bar{f}} \mid^2 \left[\cos^2(\frac{\Delta m t}{2}) + \left|\frac{pA_{\bar{f}}}{q\bar{A}_{\bar{f}}}\right|^2 \sin^2(\frac{\Delta m t}{2}) - \mathrm{Im}\left(\frac{pA_{\bar{f}}}{q\bar{A}_{\bar{f}}}\right)\sin(\Delta m t)\right] \ . \qquad (148)$$

The difference of these two probabilities signals different types of CPV: the difference in the first term in brackets appears due to direct CPV; the difference in the second term - due to CPV in mixing or due to direct CPV, and in the last term – due to CPV in the interference of $B^0 - \bar{B}^0$ mixing and decays.

Let f be a CP eigenstate: $\bar{f} = \eta_f f$, where $\eta_f = +(-)$ for CP even (odd) f. (Two examples of such decays: $B^0 \to J/\Psi K_{S(L)}$ and $B^0 \to \pi^+\pi^-$ are described by the quark diagrams shown in Fig. 16. The analogous diagrams describe $\bar{B}^0$ decays in the same final states.) The following equalities can be easily obtained:

$$A_{\bar{f}} = \eta_f A_f \ , \ \ \bar{A}_{\bar{f}} = \eta_f \bar{A}_f \ \ . \qquad (149)$$

In the absence of CPV the expressions in brackets are equal and the obtained formulas describe the exponential particle decay without oscillations. Taking CPV into account and neglecting a small deviation of $\mid p/q \mid$ from one, for CPV asymmetry of the decays into CP eigenstate we obtain:

$$a_{CP}(t) \equiv \frac{P_{\bar{B}^0 \to f} - P_{B^0 \to f}}{P_{\bar{B}^0 \to f} + P_{B_0 \to f}} = \frac{\mid \lambda \mid^2 - 1}{\mid \lambda \mid^2 + 1}\cos(\Delta m t) + \frac{2Im\lambda}{\mid \lambda \mid^2 + 1}\sin(\Delta m t) \equiv$$

$$\equiv -C_f \cos(\Delta m t) + S_f \sin(\Delta m t) \ \ , \qquad (150)$$

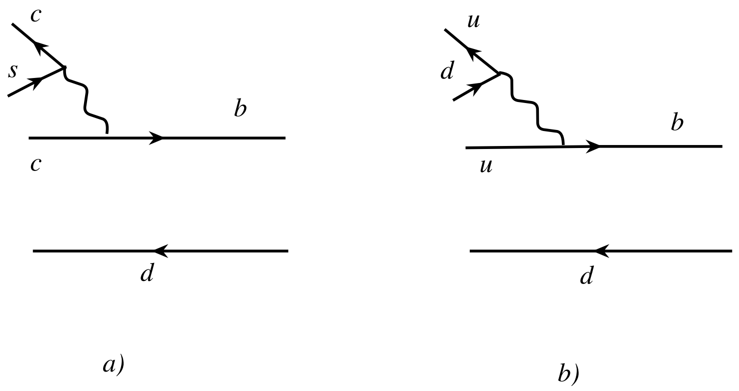


Fig. 16: Quark diagrams responsible for $B^0 \to J/\Psi K$ and $B^0 \to \pi\pi$-decays.

where $\lambda \equiv \frac{q\bar{A}_f}{pA_f}$ (not to be confused with the parameter of the CKM matrix).

The nonzero value of C_f corresponds to direct CPV; it occurs when more than one amplitude contribute to the decay. For extraction of CPV parameters (the angles of the unitarity triangle) in this case the knowledge of strong rescattering phases is necessary. The non-vanishing S_f describes CPV in the interference of mixing and decay. It is nonzero even when there is only one decay amplitude, and $|\lambda| = 1$. Such decays are of special interest since the extraction of CPV parameters becomes independent of poorly known strong phases of the final particles rescattering.

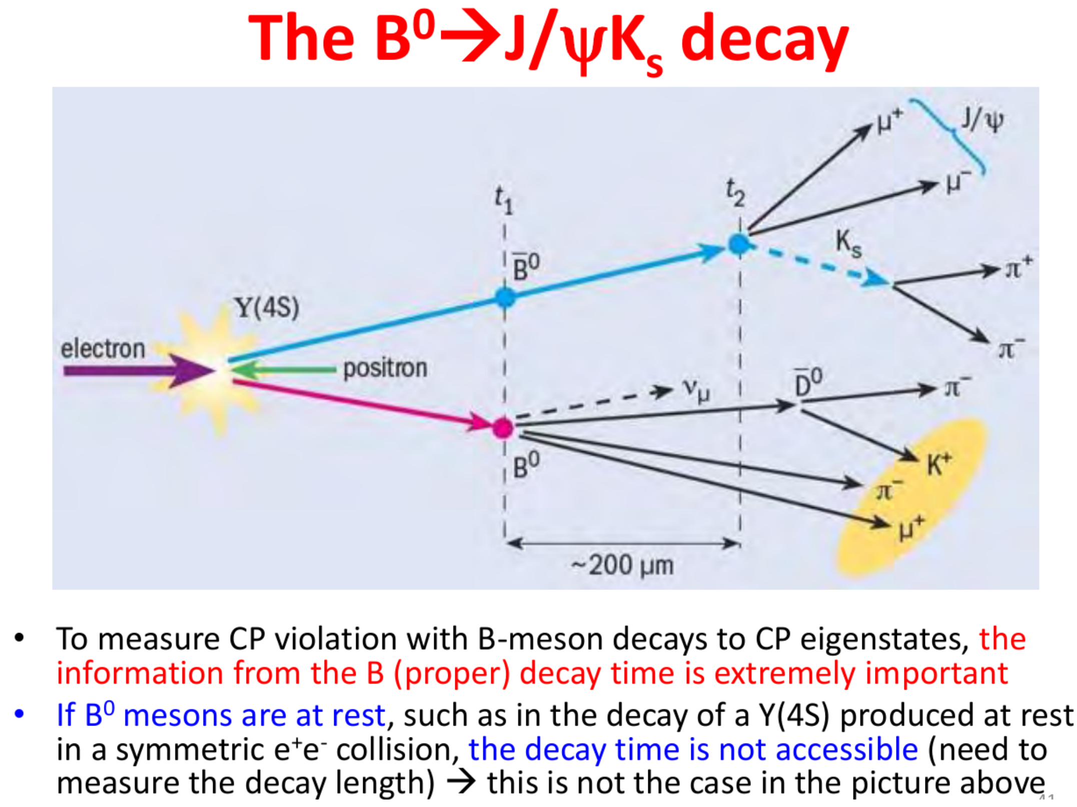
The decays of the $\Upsilon(4S)$ resonance produced in e^+e^- annihilation are a powerful source of $B^0\bar{B}^0$ pairs. A semileptonic decay of one of the B's tags "beauty" of the partner at the moment of decay (since (B^0B^0), $(\bar{B}^0\bar{B}^0)$ states are forbidden) thus making it possible to study CPV. However, the time-integrated asymmetry is zero for decays were C_f is zero. This happens since we do not know which of the two B-mesons decays earlier, and asymmetry is proportional to: $I = \int\limits_{-\infty}^{\infty} e^{-\Gamma|t|} \sin(\Delta m t) dt = 0$. The asymmetric B-factories provide the possibility to measure the time-dependence: $\Upsilon(4S)$ moves in a laboratory system, and since the energy release in $\Upsilon(4S) \to B\bar{B}$ decay is very small both B and $\bar{B}$ move with the same velocity as the original $\Upsilon(4S)$. This makes the resolution of B decay vertices possible unlike the case of $\Upsilon(4S)$ decay at rest, when non-relativistic B and $\bar{B}$ decay at almost the same point. The implementation of the time-dependent analysis for the search of CPV in B-mesons was suggested in [34] - [36].

9.2 $B_d^0(\bar{B}_d^0) \to J/\Psi K_{S(L)}$, sin 2β – straight lines

The tree diagram contributing to this decay is shown in Fig. 16 a). The product of the corresponding CKM matrix elements is: $V_{cb}^* V_{cs} \simeq A\lambda^2$. Also the penguin diagram $b \to sg$ with the subsequent $g \to c\bar{c}$ decay contributes to the decay amplitude. Its contribution is proportional to:

$$\begin{aligned} P \sim V_{us}V_{ub}^* f(m_u) + V_{cs}V_{cb}^* f(m_c) + V_{ts}V_{tb}^* f(m_t) = \\ = V_{us}V_{ub}^*(f(m_u) - f(m_t)) + V_{cs}V_{cb}^*(f(m_c) - f(m_t)) \ , \end{aligned} \tag{151}$$

where function f describes the contribution of quark loop and we have subtracted zero from the expression on the first line. The last term on the second line has the same weak phase as the tree amplitude, while the first term has a CKM factor $V_{us}V_{ub}^* \sim \lambda^4(\rho - i\eta)A$. Since the (one-loop) penguin amplitude should be in any case smaller than the tree one, we get that with 1% accuracy there is only one weak amplitude governing $B_d^0(\bar{B}_d^0) \to J/\Psi K_{S(L)}$ decays. This is the reason why this mode is called a "gold-plated mode" – the accuracy of the theoretical prediction of the CP-asymmetry is very high, and Br $(B_d \to J/\Psi K^0) \approx 10^{-3}$ is large enough to detect CPV.



The $B^0 \rightarrow J/\psi K_s$ decay

- To measure CP violation with B-meson decays to CP eigenstates, the information from the B (proper) decay time is extremely important
- If B^0 mesons are at rest, such as in the decay of a Y(4S) produced at rest in a symmetric e^+e^- collision, the decay time is not accessible (need to measure the decay length) → this is not the case in the picture above

Fig. 17: Tagging $\bar{B}^0$-meson by B^0-decay.

Substituting $|\lambda| = 1$ in the expression for $a_{CP}(t)$ we obtain:

$$a_{CP}(t) = \text{Im}\lambda \sin(\Delta m \Delta t) \quad , \tag{152}$$

where Δt is the time difference between the semileptonic decay of one of B-mesons produced in $\Upsilon(4S)$ decay and that of the second one to $J/\Psi K_{S(L)}$. Using the following equation

$$\bar{A}_f = \eta_f \bar{A}_{\bar{f}} \quad , \tag{153}$$

where η_f is CP parity of the final state, we obtain:

$$\lambda = \left(\frac{q}{p}\right)_{B_d} \frac{A_{\bar{B}^0 \to J/\Psi K_{S(L)}}}{A_{B^0 \to J/\Psi K_{S(L)}}} = \left(\frac{q}{p}\right)_{B_d} \eta_f \frac{A_{\bar{B}^0 \to \overline{J/\Psi K_{S(L)}}}}{A_{B^0 \to J/\Psi K_{S(L)}}} \quad . \tag{154}$$

The amplitude in the nominator contains $\bar{K}^0$ production. To project it on $\bar{K}_{S(L)}$ we should use:

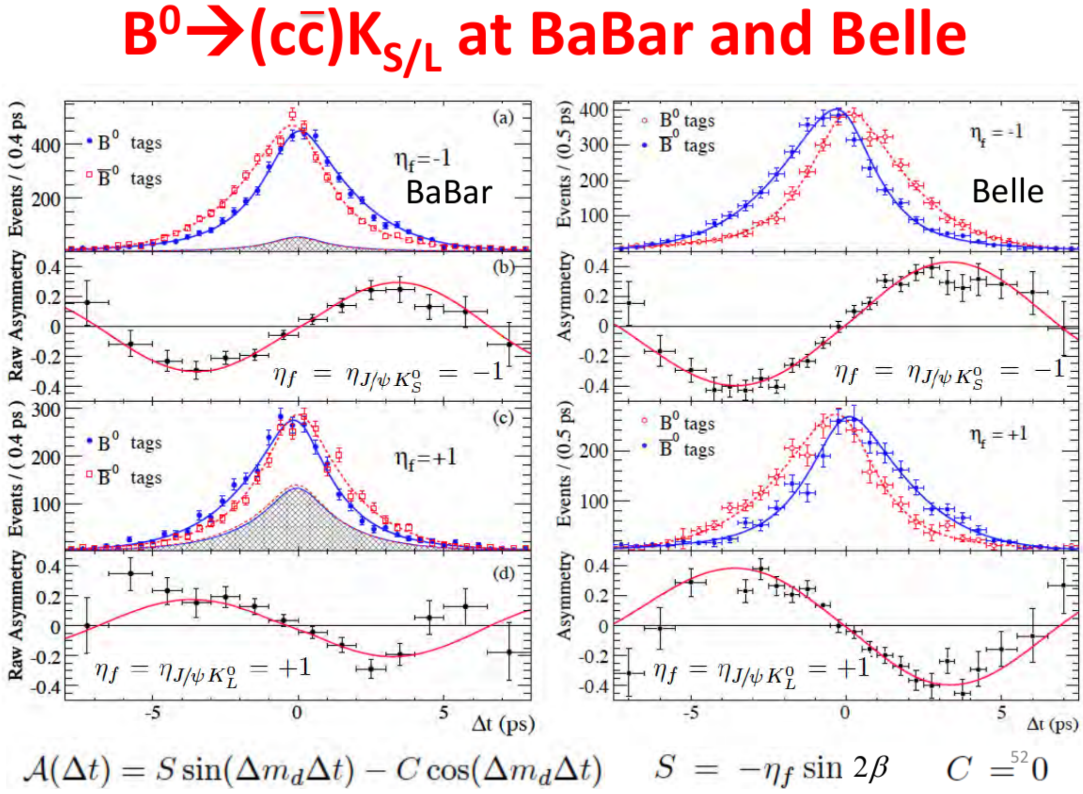
$$\overline{K^0} = \frac{K_S - K_L}{(q)_K} = \frac{\bar{K}_S + \bar{K}_L}{(q)_K} \quad , \tag{155}$$

getting $(q)_K$ in the denominator. The amplitude in the denominator contains K^0 production, and using:

$$K^0 = \frac{K_S + K_L}{(p)_K} \tag{156}$$

we obtain factor $(p)_K$ in the nominator. Collecting all the factors together and substituting CKM matrix elements for $\bar{A}_{\bar{f}}/A_f$ ratio we get:

$$\lambda = \eta_{S(L)} \left(\frac{q}{p}\right)_{B_d} \frac{V_{cb}V_{cs}^*}{V_{cb}^*V_{cs}} \left(\frac{p}{q}\right)_K \quad . \tag{157}$$



B⁰→(cc̄)K$_{S/L}$ at BaBar and Belle

$\mathcal{A}(\Delta t) = S\sin(\Delta m_d \Delta t) - C\cos(\Delta m_d \Delta t) \qquad S = -\eta_f \sin 2\beta \qquad C \overset{52}{=} 0$

Fig. 18: Measurements of CPV asymmetries.

Substituting the expressions for $(q/p)_{B_d}$ and $(p/q)_K$ we obtain:

$$\lambda(J/\Psi K_{S(L)}) = \eta_{S(L)} \frac{V_{td}V_{tb}^*}{V_{td}^*V_{tb}} \frac{V_{cb}V_{cs}^*}{V_{cb}^*V_{cs}} \frac{V_{cd}^*V_{cs}}{V_{cd}V_{cs}^*} \quad , \tag{158}$$

which is invariant under the phase rotation of any quark field. From the unitarity triangle figure we have

$$\arg(V_{tb}^*V_{td}) = \pi - \beta \quad , \tag{159}$$

and we finally obtain:

$$a_{CP}(t)\Bigg|_{J/\Psi K_{S(L)}} = -\eta_{S(L)} \sin(2\beta)\sin(\Delta m \Delta t) \quad , \tag{160}$$

which is a simple prediction of the Standard Model. Since in B decays J/Ψ and $K_{S(L)}$ are produced in P-wave, $\eta_{S(L)} = -1(+1)$ (CP of J/Ψ is $+1$, that of K_S is $+1$ as well, and $(-1)^l = -1$ comes from P-wave; CP of K_L is -1).

In this way the measurement of this asymmetry at B-factories provides the value of angle β of the unitarity triangle. The Belle, BaBar and LHCb average is:

$$\sin 2\beta = 0.691 \pm 0.017 \ , \tag{161}$$

which corresponds to

$$\beta = (21.9 \pm 0.7)^0. \tag{162}$$

As a final state not only $J/\Psi K_{S(L)}$ were selected, but neutral kaons with the other charmonium states as well.

Let us note that the decay amplitudes and $K^0 - \bar{K}^0$ mixing do not contain a complex phase, that is why the only source of it in $B^0 \to \text{charmonium}\ K_{S(L)}$ decays is $B^0 - \bar{B}^0$ mixing:

$$\left(\frac{q}{p}\right)_{B_d} = \sqrt{\frac{M_{12}^*}{M_{12}}} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*} \ , \tag{163}$$

thus the phase comes from V_{td}, that is why the final expression contains angle 2β – the phase of V_{td}/V_{td}^*.

Fig. 17 and Fig. 18 (see [37]) illustrate the above discussion.

10 Probability of the $\Upsilon(4S) \to B_d^0 \bar{B}_d^0 \to J/\Psi K_S\ J/\Psi K_S$ decay

The following parameters are used to describe the time evolution of B-mesons: $m \equiv (m_H + m_L)/2$, $\Delta m \equiv m_H - m_L$, $\Gamma_H = \Gamma_L = \Gamma$.

Since $J^{PC}(\Upsilon) = 1^{--}$, B-mesons are produced in P-wave, so their wave function is C-odd: $\Psi(t_1, t_2) = B^0(t_1)\bar{B}^0(t_2) - B^0(t_2)\bar{B}^0(t_1)$.

For the decay amplitude we get:

$$\begin{aligned} \langle J/\Psi K_S\ J/\Psi K_S | \Psi(t_1,t_2)\rangle &= e^{-imt_1 - \frac{\Gamma t_1}{2}} \left[A \cos\frac{\Delta m t_1}{2} + i\frac{q}{p}\sin\left(\frac{\Delta m t_1}{2}\right)\bar{A}\right] \times \\ &\times e^{-imt_2 - \frac{\Gamma t_2}{2}} \left[\cos\left(\frac{\Delta m t_2}{2}\right)\bar{A} + i\frac{p}{q}\sin\left(\frac{\Delta m t_2}{2}\right) A\right] - (t_1 \leftrightarrow t_2) = \\ &= e^{-im(t_1+t_2) - \Gamma\frac{t_1+t_2}{2}} \left[(i\frac{p}{q}A^2 - i\frac{q}{p}\bar{A}^2)\cos\left(\frac{\Delta m t_1}{2}\right)\sin\left(\frac{\Delta m t_2}{2}\right) + \right. \\ &\left. + (i\frac{q}{p}\bar{A}^2 - i\frac{p}{q}A^2)\sin\left(\frac{\Delta m t_1}{2}\right)\cos\left(\frac{\Delta m t_2}{2}\right)\right] = -e^{-2imt - \Gamma t}(i\frac{p}{q}A^2)[1-\lambda^2]\sin\left(\frac{\Delta m \Delta t}{2}\right) \ , \end{aligned} \tag{164}$$

where $t \equiv \frac{t_1+t_2}{2}$, $\Delta t \equiv t_1 - t_2$, $\frac{q}{p} = e^{-2i\beta}$.

The decay probability equals

$$P(J/\Psi K_S, J/\Psi K_S) = e^{-2\Gamma t}|A|^4[1 - e^{4i\beta}][1 - e^{-4i\beta}]\sin^2\left(\frac{\Delta m \Delta t}{2}\right) \sim e^{-2\Gamma t}\sin^2(2\beta)\sin^2\frac{(\Delta m \Delta t)}{2} \ . \tag{165}$$

Changing integration variables in the expression for the decay probability according to

$$\int\limits_0^\infty dt_1 \int\limits_0^\infty dt_2 = \int\limits_{-\infty}^\infty d(\Delta t) \int\limits_{|\Delta t|/2}^\infty dt \tag{166}$$

and performing integration over t we get:

$$N(\Delta t) \sim \sin^2 2\beta[1 - \cos(\Delta m \Delta t)]e^{-\Gamma|\Delta t|} \ , \tag{167}$$

which is zero when $\Delta t = 0$ due to Bose statistics, when $\Delta m = 0$ – no oscillations, and for $\beta = 0$ – no CPV (CP $\Upsilon = +$, CP $(J/\Psi K_S\ J/\Psi K_S) = -$).

For the total number of $\Upsilon(4S) \to J/\Psi K_S\ J/\Psi K_S$ decays integrating over Δt we obtain:

$$N(J/\Psi K_S\ J/\Psi K_S) \sim \sin^2 2\beta \left(\frac{\Delta m^2}{\Delta m^2 + \Gamma^2} \right) \tag{168}$$

After one of B decays to $J/\Psi K_S$ the second one starts to oscillate and may decay to $J/\Psi K_S$ as well. The initial state is CP even, the final state is CP odd, so no decays without CPV would occur.

Taking different initial and final states one may solve many problems the same way as we have just shown.

C-even initial state:

$$\Psi(t_1, t_2) = B^0(t_1)\bar{B}^0(t_2) + B^0(t_2)\bar{B}^0(t_1) \ . \tag{169}$$

The "classical“ initial state (produced in hadron collisions):

$$\Psi(t_1, t_2) = B^0(t_1)\bar{B}^0(t_2) \ . \tag{170}$$

11 CPV in the $b \to sg \to ss\bar{s}$ transition: penguin domination

The decays $B_d \to \phi K^0, K^+K^-K^0, \eta' K^0$ proceed through the diagrams shown in Fig. 19.

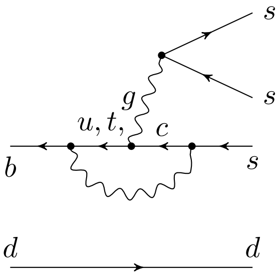


Fig. 19: Penguin diagram describing $b \to ss\bar{s}$-transition.

The diagram with an intermediate u-quark is proportional to λ^4, while those with intermediate c- and t-quarks are proportional to λ^2. In this way the main part of the decay amplitude is free of the CKM phase, just like in case of $B_d \to J/\Psi K$ decays. A nonzero phase which leads to time-dependent CP asymmetry comes from the $B_d - \bar{B}_d$ transition:

$$a_{CP}(t) = -\eta_f \sin(2\beta)\sin(\Delta m \Delta t) \ , \tag{171}$$

analogously to $B_d \to J/\Psi K$ decays.

The main interest in these decays is to look for phases of NP which may be hidden in loops. According to Fig. 20 [38] SM nicely describes the experimental data within their present day accuracy.

12 $B_s(\bar{B}_s) \to J/\Psi\phi$, ϕ_s

This decay is an analog of $B^0(\bar{B}^0) \to J/\Psi K$ decay: the tree amplitude dominates and CP asymmetry could appear from the $B_s \leftrightarrow \bar{B}_s$ transition. V_{ts} unlike V_{td} is almost real, so the asymmetry should be

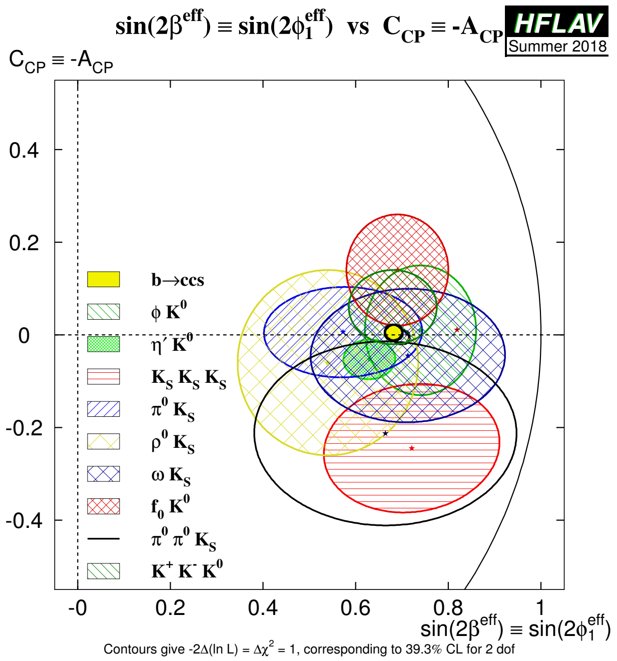


Fig. 20: CP-asymmetries from B_d-decays with production of three strange quarks.

very small in the SM – a good place to look for New Physics. The angular analysis of $J/\Psi \to \mu^+\mu^-$ and $\phi \to KK$ decays is necessary to select the final states with definite CP parity.

Taking the difference of the width of two eigenstates into account ($\Delta\Gamma = \Gamma_L - \Gamma_H$) we get:

$$P_{B_s \to f}(t) = \frac{1}{2}e^{-\Gamma t}|A_f|^2(1+|\lambda_f|^2)[\cosh(\Delta\Gamma t/2) - D_f \sinh(\Delta\Gamma t/2) + C_f \cos(\Delta m t) - S_f \sin(\Delta m t)] \ , \tag{172}$$

$$P_{\bar{B}_s \to f}(t) = \frac{1}{2}e^{-\Gamma t}|\frac{p}{q}A_f|^2(1+|\lambda_f|^2)[\cosh(\Delta\Gamma t/2) - D_f \sinh(\Delta\Gamma t/2) - C_f \cos(\Delta m t) + S_f \sin(\Delta m t)] \ , \tag{173}$$

$$D_f = \frac{2Re\lambda_f}{1+|\lambda_f|^2}, \quad C_f = \frac{1-|\lambda_f|^2}{1+|\lambda_f|^2}, \quad S_f = \frac{2Im\lambda_f}{1+|\lambda_f|^2} \ . \tag{174}$$

$$A_{CP}(t)(|p/q|=1) = \frac{-C_f \cos(\Delta m t) + S_f \sin(\Delta m t)}{\cosh(\Delta\Gamma t/2) - D_f \sinh(\Delta\Gamma t/2)} \ . \tag{175}$$

The Standard Model prediction is $\phi_s^{SM} = -\arg\frac{V_{ts}V_{tb}^*}{V_{ts}^*V_{tb}} = -2\lambda^2\eta$ =-0.036 rad, while $\phi_s^{exp} = -0.040 \pm 0.025$ rad. No New Physics in this decay as well.

13 The angles α and γ

13.1 $\alpha : B \longrightarrow \pi\pi, \rho\rho, \pi\rho$

Since α is the angle between $V_{tb}^* V_{td}$ and $V_{ub}^* V_{ud}$, the time dependent CP asymmetries in $b \longrightarrow u\bar{u}d$ decay dominated modes directly measure $\sin(2\alpha)$.

$b \longrightarrow d$ penguin amplitudes have different CKM phases compared to the tree amplitude and are of the same order in λ. Thus the penguin contribution can be sizeable, making the determination of α complicated.

Fortunately $Br(B \to \rho^0\rho^0) \ll Br(B \to \rho^+\rho^-), Br(B^+ \to \rho^+\rho^0)$, which proves that the contribution of the penguins in $B \longrightarrow \rho\rho$ decays is small.

Moreover, the longitudinal polarization fractions in $B \to \rho^+\rho^-, B^+ \to \rho^+\rho^0$ decays appeared to be close to unity, which means that the final states are CP even and the following relations should be valid:

$$S_{\rho^+\rho^-} = \sin(2\alpha), \quad C_{\rho^+\rho^-} = 0 \ . \tag{176}$$

The experimental numbers are:

$$S_{\rho^+\rho^-} = -0.05 \pm 0.17, \quad C_{\rho^+\rho^-} = -0.06 \pm 0.13 \ . \tag{177}$$

So, C is compatible with zero, while from S we get

$$\alpha = (91 \pm 5)^0 \ . \tag{178}$$

Finally from the combination of the $B \longrightarrow \pi\pi, \rho\rho, \pi\rho$ modes the following result is obtained: $\alpha = (85 \pm 4)^0$.

Problem 8

In the decays considered in this section the quarks of the first and the third generations participate, so only 2 generations are involved. As it has been stated and demonstrated, at least 3 generations are needed for CPV. So, how does it happen that in $B \longrightarrow \rho\rho$ decays CP is violated?

13.2 γ

The next task is to measure the angle γ, or the phase of V_{ub}. In B_d decays the angle β enters the game through $B_d - \bar{B}_d$ mixing. To avoid it in order to single out angle γ we should consider B_s decays, or the decays of charged B-mesons [39]. The interference of $B^- \longrightarrow D^0 K^- (b \longrightarrow c\bar{u}s)$ and $B^- \longrightarrow \bar{D}^0 K^- (b \longrightarrow u\bar{c}s)$ transitions in the final states accessible in both D^0 and $\bar{D}^0$ decays (such as $K_S^0 \pi^+ \pi^-$) provides the best accuracy in γ determination [40]. Combining all the existing methods, the following result was obtained:

$$\gamma = (74 \pm 5)^0 \ . \tag{179}$$

Here the LHCb measurement is significantly more precise than the old Belle and BaBar results and it undergoes continuous improvement.

14 CKM fit

The UTfit and CKMfitter collaborations are making fits of available data by four Wolfenstein parameters. Here are the UTfit results:

$$\begin{aligned} \lambda &= 0.225(1) \ , \\ A &= 0.83(1) \ , \\ \eta &= 0.36(1) \ , \\ \rho &= 0.15(1) \ . \end{aligned} \tag{180}$$

For the angles of the unitarity triangle the result of the fit is:

$$\alpha = (90 \pm 2)^0, \quad \beta = (24 \pm 1)^0, \quad \gamma = (66 \pm 2)^0 \ . \tag{181}$$

So $\alpha + \beta + \gamma = 180^0$ – no traces of New Physics yet.

The quality of fit is high and CKMfitter results are approximately the same.

15 Perspectives: $K \longrightarrow \pi\nu\nu$, Belle II, LHC

Two running experiments are measuring the probabilities of $K^+ \to \pi^+\nu\bar{\nu}$ (NA62 at SPS, CERN) and $K_L \to \pi^0\nu\bar{\nu}$ (KOTO at J-PARC, Japan) decays. These decays are very rare. In the framework of the SM the branching ratios of these decays are predicted with high accuracy: $\mathrm{Br}(K^+ \to \pi^+\nu\bar{\nu}) = (8.4 \pm 1)10^{-11}$, $\mathrm{Br}(K_L \to \pi^0\nu\bar{\nu}) = (3.4 \pm 0.6)10^{-11}$. The smallness of branching ratios in the SM makes these decays a proper place to look for indirect manifestations of New Physics.

The Belle II experiment at KEK laboratory started taking data in 2019. With much higher luminosity than that collected by Belle and BaBar it will also contribute to the search for New Physics. The planned Belle II sensitivities for the measurement of the angles of the unitarity triangle are 1%.

Knowledge of the unitarity triangle parameters with better accuracy is expected from the future LHC data. Assuming a reasonable improvement of non-perturbative quantities from lattice QCD we can hope that it will be sufficient to crack the triangle.

Useful introductions to flavor physics and CP violation can be found in Refs. [41–44].

Acknowledgements

I am grateful to the organizers of the PINP 2019 Winter School and the European School of High Energy Physics 2019 for the invitations to lecture at the Schools, to A.E. Bondar, V.F. Obraztsov and P.N. Pakhlov for useful comments and to S.I. Godunov and E.A. Ilyina for the help in preparation slides and lectures. This work was supported by RSF grant 19-12-00123.

References

[1] S.L. Glashow, *Nucl. Phys.* **22** (1961) 579, doi:10.1016/0029-5582(61)90469-2;
S. Weinberg, *Phys. Rev. Lett.* **19** (1967) 1264, doi:10.1103/PhysRevLett.19.1264;
A. Salam, Weak and electromagnetic interactions, Proc. 8th Nobel Symposium, Ed. N. Svartholm (Almqvist & Wiksell, Stockholm, 1968), pp. 367–377, reprinted in *Selected Papers of Abdus Salam*, Eds. A. Ali *et al.*, (World Scientific, Singapore, 1994), pp. 244–254, doi:10.1142/9789812795915_0034.

[2] N. Cabibbo, *Phys. Rev. Lett.* **10** (1963) 531, doi:10.1103/PhysRevLett.10.531.

[3] M. Gell-Mann, *Phys. Rev.* **125** (1962) 1067, doi:10.1103/PhysRev.125.1067.

[4] S.L. Glashow, J. Iliopoulos, L. Maiani, *Phys. Rev.* **D2** (1970) , doi:10.1103/PhysRevD.2.1285.

[5] M. Kobayashi, T. Maskawa, *Progr. Theor. Phys.* **49** (1973) 652, doi:10.1143/PTP.49.652.

[6] L. Wolfenstein, *Phys. Rev. Lett.* **51** (1983) 1945, doi:10.1103/PhysRevLett.51.1945.

[7] C. Jarlskog, *Phys. Rev. Lett.* **55** (1985) 1039, doi:10.1103/PhysRevLett.55.1039.

[8] T.D. Lee, C.N. Yang, *Phys. Rev.* **104** (1956) 254, doi:10.1103/PhysRev.104.254.

[9] B.L. Ioffe, L.B. Okun, A.P. Rudik, *Zh. Eksp. Teor. Fiz.* **32** (1957) 396, English transl. publ. in *Sov. Phys. JETP* **5** (1957) 328, http://jetp.ras.ru/cgi-bin/e/index/e/5/2/p328?a=list.

[10] M. Gell-Mann, A. Pais, *Phys. Rev.* **97** (1955) 1387, doi:10.1103/PhysRev.97.1387.

[11] T.D. Lee. C.N Yang, R. Oehme, *Phys. Rev.* **106** (1957) 340, doi:10.1103/PhysRev.106.340.

[12] L.D. Landau, *Zh. Eksp. Teor. Fiz.* **32** (1957) 405, English transl. publ. in *Sov. Phys. JETP* **5** (1957) 336, http://jetp.ras.ru/cgi-bin/e/index/r/32/2/p405?a=list;
L.D. Landau, *Nucl. Phys.* **3** (1957) 127, doi:10.1016/0029-5582(57)90061-5.

[13] L.B. Okun, *Slaboe vzaimodeistvie elementarnykh chastits* (M.: Fizmatgiz, 1963) (in Russian).

[14] J.H. Christenson, J.W. Cronin, V.L. Fitch, R. Turlay, *Phys. Rev.* **13** (1964) 138, doi:10.1103/PhysRevLett.13.138.

[15] V. Fanti *et. al.* (NA 48 Collaboration), *Phys. Lett.* **B465** (1999) 335, doi:10.1016/S0370-2693(99)01030-8;

A. Alavi-Harati *et al.* (KTeV Collaboration), *Phys. Rev. Lett.* **83** (1999) 22, doi:10.1103/PhysRevLett.83.22.

[16] B. Aubert *et. al.* (BaBar Collaboration), *Phys. Rev. Lett.* **87** (2001) 091801, doi:10.1103/PhysRevLett.87.091801;
K. Abe *et al.* (Belle Collaboration), *Phys. Rev. Lett.* **87** (2001) 091802, doi:10.1103/PhysRevLett.87.091802.

[17] R.Aaij *et. al.* (LHCb Collaboration), *Phys. Rev. Lett.* **122** (2019) , doi:10.1103/PhysRevLett.122.211803.

[18] A.D. Sakharov, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32, English transl. publ. in *JETP Lett.* **5** (1967) 24, transl. reprinted in *Sov.Phys.Usp.* **34** (1991) 392, doi:10.1070/PU1991v034n05ABEH002497.

[19] M.I. Vysotsky, *Yad. Fiz.* **31** (1980) 1535, English transl. publ. in *Sov. J. Nucl. Phys.* **31** (1980) 797.

[20] R.Aaij *et al.* (LHCb Collaboration), *Phys. Rev. Lett.* **110** (2013) 221601, doi:10.1103/PhysRevLett.110.221601.

[21] M. Tanabashi *et al.* (Particle Data Group), *Phys. Rev.* **D98** (2018) 030001, doi:10.1103/PhysRevD.98.030001.

[22] B. Pontecorvo, *Zh. Eksp. Teor. Fiz.* **34** (1957) 247, English transl. publ. in *Sov. Phys. JETP* **7** (1958) 172, http://jetp.ras.ru/cgi-bin/e/index/r/34/1/p247?a=list.

[23] A. Ceccucci, Z. Ligeti, Y. Sakai, "CKM quark-mixing matrix", Review of particle physics, Eds. M. Tanabashi *et al.* [Particle Data Group], *Phys. Rev.* **D98** (2018) 030001; pp. 229–237, doi:10.1103/PhysRevD.98.030001.

[24] M. Ademollo, R. Gatto, *Phys. Rev. Lett.* **13** (1964) 264, doi:10.1103/PhysRevLett.13.264.

[25] M.V. Terent'ev, *Zh. Eksp. Teor. Fiz.* **44** (1963) 1320, English transl. publ. in *Sov. Phys. JETP* **17** (1963) 890, http://jetp.ras.ru/cgi-bin/e/index/e/17/4/p890?a=list.

[26] H. Albrecht *et al.*, *Phys. Lett.* **B192** (1987) 245, doi:10.1016/0370-2693(87)91177-4.

[27] P. Oddone, "Detector considerations", Proc. UCLA Workshop Linear Collider $B\bar{B}$ Factory Conceptual Design, Los Angeles, California, 26–30 Jan. 1987, Ed. D.H. Stork (World Scientific, Singapore, 1987), pp. 423–446, https://inspirehep.net/literature/256027.

[28] C. Albajar *et al.*, *Phys. Lett.* **B186** (1987) 247, doi:10.1016/0370-2693(87)90288-7, Erratum: *Phys. Lett.* **B197** 565 (1987), doi:10.1016/0370-2693(87)91057-4.

[29] O. Schneider, "$B^0 - \bar{B}^0$ mixing", Review of particle physics, Eds. M. Tanabashi *et al.* [Particle Data Group], *Phys. Rev.* **D98** (2018) 030001, pp. 725–730, doi:10.1103/PhysRevD.98.030001.

[30] A.A. Anselm, Ya.I. Azimov, *Phys. Lett.* **B85** (1979) 72, doi:10.1016/0370-2693(79)90779-2.

[31] M. Bander, D. Silverman, A. Soni, *Phys. Rev. Lett.* **43** (1979) 242, doi:10.1103/PhysRevLett.43.242.

[32] A. Carter, A. Sanda, *Phys. Rev. Lett.* (1980) 952, doi:10.1103/PhysRevLett.45.952.

[33] I.I. Bigi, A.I. Sanda, *Nucl. Phys.* **B193** (1981) 85, doi:10.1016/0550-3213(81)90519-8.

[34] I. Dunietz, J. Rosner, *Phys. Rev.* **D34** (1986) 1404, doi:10.1103/PhysRevD.34.1404.

[35] Ya.I. Azimov, N.G. Uraltzev, V.A. Khoze, *Yad. Fiz.* **45** (1987) 1412, English transl. publ. in *Sov. J. Nucl. Phys.* **45** (1987) 878.

[36] Ya.I. Azimov, N.G. Uraltzev, V.A. Khoze, *Proc. XXI Winter School Leningrad Institute of Nuclear Physics* (LIYaF, 1986) p. 178 (*in Russian*).

[37] Eds. A.J. Bevan *et al.*, *Eur. Phys. J.* **C74** (2014) 3026, doi:10.1140/epjc/s10052-014-3026-9, arXiv: 1406.6311 [hep-ex].

[38] T. Gershon and Y. Nir, "CP violation in the quark sector", Review of particle physics, Eds. M. Tanabashi *et al.* [Particle Data Group], *Phys. Rev.* **D98** (2018) 030001, pp. 238–250 doi10.1103/PhysRevD.98.030001.

[39] M. Gronau, D. Wyler, *Phys. Lett.* **B265** (1991) 172, doi:10.1016/0370-2693(91)90034-N; M. Gronau, D. London, *Phys. Lett.* **B253** (1991) 483, doi:10.1016/0370-2693(91)91756-L; D. Atwood, I. Dunietz, A. Soni, *Phys. Rev. Lett.* **78** (1997) 3257, doi:10.1103/PhysRevLett.78.3257, arXiv:hep-ph/9612433.

[40] A. Bondar, Proc. BINP Special Meeting on Dalitz Analysis, 24–26 Sep. 2002 (unpublished); A. Giri *et al.*, *Phys. Rev.* **D68** (2003) 054018, doi:10.1103/PhysRevD.68.054018, arXiv:hep-ph/0303187.

[41] J. Zupan, "Introduction to flavour physics", Proc. 2018 European School of High-Energy Physics, Maratea, Italy, 20 Jun. –3 Jul. 2018, Eds. M. Mulders and C. Duhr (CERN, Geneva, 2019), pp. 181–212, doi:10.23730/CYRSP-2019-006.181, arXiv:1903.05062 [hep-ph].

[42] M. Blanke, "Introduction to flavour physics and CP violation", Proc. 2016 European School of High-Energy Physics, Skeikampen, Norway, 15–28 Jun. 2016, Eds. M. Mulders and G. Zanderighi (CERN, Geneva, 2017), pp. 71-100, doi:10.23730/CYRSP-2017-005.71, arXiv:1704.03753 [hep-ph].

[43] B. Grinstein, "Lectures on flavor physics and CP violation", Proc. 8th CERN–Latin-American School of High-Energy Physics, Ibarra, Ecuador, Mar. 5–17 2015, Eds. M. Mulders and G. Zanderighi (CERN, Geneva, 2016), pp. 43–84, doi:10.5170/CERN-2016-005.43, arXiv:1701.06916 [hep-ph].

[44] S. Gori, "Three lectures of flavor and CP violation within and beyond the Standard Model", Proc. 2015 European School of High-Energy Physics, Bansko, Bulgaria 2–15 Sep. 2015, Eds. M. Mulders and G. Zanderighi (CERN, Geneva, 2017), pp. 65–90, doi:10.23730/CYRSP-2017-004.65, arXiv:1610.02629 [hep-ph].

Proceedings of the 2019 European School of High-Energy Physics, St Petersburg, Russia, 14–27 September 2019, edited by C. Duhr and M. Mulders, CERN-2021-005 (CERN, Geneva, 2021)

Neutrino physics

M.C. Gonzalez-Garcia
Departament de Física Quàntica i Astrofísica and Institut de Ciencies del Cosmos, Universitat de Barcelona, Barcelona, Spain,
Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain,
C.N. Yang Institute for Theoretical Physics, Stony Brook University, Stony Brook, USA.

Abstract

The purpose of these lectures is to quantitatively summarize the present status of the phenomenology of massive neutrinos. In the first lecture I will present the low energy formalism for adding neutrino masses to the Standard Model and the induced leptonic mixing, and I will describe the status of the existing probes of the absolute neutrino mass scale. The second lecture is devoted to describing the phenomenology associated with neutrino flavour oscillations in vacuum and in matter and the corresponding experimental results observing these phenomena. In the third lecture I will present the minimal 3ν mixing picture emerging from the global description of the data. I will briefly comment on the status of extensions of this picture with additional light states and the possibility of non-standard neutrino interactions. I will also discuss some theoretical implications of these results, such as the existence of new physics, the estimate of the scale of this new physics, leptogenesis and collider signatures.

Keywords

Neutrinos, flavour oscillations, neutrino masses, sterile neutrinos, leptogenesis, lectures.

1 LECTURE I: Neutrino properties

1.1 Introduction

In 1930 Wolfgang Pauli postulated the existence of a new particle in order to reconcile the observed continuous spectrum of nuclear beta decay with energy conservation. The postulated particle had no electric charge and, in fact, Pauli himself pointed out that in order to do the job it had to weight less than one percent of the proton mass, thus establishing the first limit on the *neutrino* mass. It was Fermi, who, in 1934 [1], gave its name to the neutrino and first proposed the four-fermion theory of beta decay. The neutrino was first observed by Cowan, Reines and collaborators [2] in 1956 in a reactor experiment. Soon after, in 1958 its helicity was determined by Goldhaber and collaborators [3] to be always -1 (*left-handed*) and as such were introduced in the Standard Model (SM).

Neutrinos are copiously produced in natural sources: in the burning of the stars, in the interaction of cosmic rays, in the Earth radioactivity... even as relics of the Big Bang. In the 1960's, neutrinos produced in the sun and in the atmosphere were first observed. In 1987, neutrinos from a supernova in the Large Magellanic Cloud were also detected. In 2013 the ICECUBE experiment detected high energy neutrinos from extragalactic sources. Neutrinos are also produced in *man-made* facilities, starting with the nuclear reactors which were the first source to be detected, and continuing with dedicated beams produced with particle accelerators. All these observations play an important role in understanding the properties of the neutrinos. In particular they allowed to establish that neutrinos carry *lepton flavour* characterizing them by the charged lepton with which they are produced in a SM weak current interaction.

The properties of the neutrino and in particular the question of its mass have intrigued physicists' minds ever since it was proposed. In the laboratory, neutrino masses have been kinematically searched

http://doi.org/10.23730/CYRSP-2021-005.85

for without any positive result. Experiments achieved higher and higher precision, reaching upper limits for the electron-neutrino mass of 10^{-9} the proton mass, rather than the 10^{-2} originally obtained by Pauli. This raised the question of whether neutrinos are truly massless like photons.

It is clear that the answer to this question is limited by our capability of detecting the effect of a non-zero neutrino mass. This is a very difficult task in direct kinematic measurements. In 1957, however, Bruno Pontecorvo [4, 5] realized that the existence of neutrino masses may not only reveal itself in kinematic effects but it implies also the possibility of neutrino oscillations. Flavor oscillations of neutrinos were searched for using either neutrino beams from reactors or accelerators, or natural neutrinos generated at astrophysical sources (the Sun giving the largest flux) or in the atmosphere. The longer the distance that the neutrinos travel from their production point to the detector, the smaller masses that can be signaled by their oscillation. Indeed, the solar neutrinos allow us to search for masses that are as small as 10^{-5} eV, that is 10^{-14} of the proton mass!

Experiments studying natural neutrino fluxes were the first to provide us with strong evidence of neutrino masses and lepton flavour mixing. Experiments that measure the flux of atmospheric neutrinos found results that suggested the disappearance of muon-neutrinos when propagating over distances of order hundreds (or more) kilometers. Experiments that measured the flux of solar neutrinos found results that eventually demonstrated the disappearance of electron-neutrinos while propagating within the Sun. The disappearance of both atmospheric ν_μ's and solar ν_e's was most easily explained in terms of neutrino flavour transitions associated to neutrino masses and mixing. These results were tested and eventually confirmed with increasing precision in experiments using laboratory beams from nuclear reactors and accelerators. With the exception of a set of unconfirmed "hints" of possible eV scale mass states, all the oscillation signatures can be explained with the three flavor neutrinos (ν_e, ν_μ, ν_τ) expressed as quantum superposition of three massive states ν_i ($i = 1, 2, 3$) with different masses m_i.

In these lectures I first discuss some generic properties of the neutrinos related to the question of their mass and describe the low energy formalism for adding neutrino masses to the SM and the induced leptonic mixing. In the second lecture I describe the phenomenology associated with neutrino flavour oscillations in vacuum and transitions in matter and present the experimental evidence of neutrino oscillations. In the third lecture I will first present the derived values of neutrino masses and mixing when the bulk of data is consistently analyzed in the framework of mixing between the three active neutrinos. I will briefly comment on the status of extensions of this picture with additional light states and the possibility of non-standard neutrino interactions. I will also discuss some theoretical implications and some avenues open by these results: the existence of new physics, the estimate of the scale of this new physics, leptogenesis, collider signatures, etc…

In preparing these lectures, I have benefited from the many excellent books, such as Refs. [6–10], and several review articles. In the writing of these notes, I have used material from my review articles [11–13].

1.2 Standard Model of massless neutrinos

The Standard Model (SM) is based on the gauge group

$$G_{\text{SM}} = SU(3)_{\text{C}} \times SU(2)_{\text{L}} \times U(1)_{\text{Y}}, \tag{1}$$

with three fermion generations, where a single generation consists of five different representations of the gauge group,

$$Q_L(3,2,\frac{1}{6})\ ,\ U_R(3,1,\frac{2}{3})\ ,\ D_R(3,1,-\frac{1}{3})\ ,\ L_L(1,2,-\frac{1}{2})\ ,\ E_R(1,1,-1). \tag{2}$$

where the numbers in parenthesis represent the corresponding charges under the group (1).

The model contains a single Higgs boson doublet, $\phi(1,2,1/2)$, whose vacuum expectation value breaks the gauge symmetry,

$$\langle\phi\rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \quad \Longrightarrow \quad G_{\rm SM} \to SU(3)_{\rm C} \times U(1)_{\rm EM}. \tag{3}$$

Neutrinos are fermions that have neither strong nor electromagnetic interactions, *i.e.* they are singlets of $SU(3)_{\rm C} \times U(1)_{\rm EM}$. Active neutrinos have weak interactions, that is, they are not singlets of $SU(2)_{\rm L}$. They reside in the lepton doublets L_L. Sterile neutrinos are define as having no SM gauge interactions, this is, they are singlets of the SM gauge group..

The SM has three active neutrinos accompanying the charged lepton mass eigenstates, e, μ and τ:

$$L_{L\ell} = \begin{pmatrix} \nu_{L\ell} \\ \ell^-_L \end{pmatrix}, \quad \ell = e, \mu, \tau. \tag{4}$$

Thus the charged current interaction terms for leptons read

$$-\mathcal{L}_{\rm CC} = \frac{g}{\sqrt{2}} \sum_\ell \overline{\nu_{L\ell}} \gamma^\mu \ell^-_L W^+_\mu + \text{h.c.}. \tag{5}$$

In addition, the SM neutrinos have neutral current (NC) interactions,

$$-\mathcal{L}_{\rm NC} = \frac{g}{2\cos\theta_W} \sum_\ell \overline{\nu_{L\ell}} \gamma^\mu \nu_{L\ell} Z^0_\mu. \tag{6}$$

Equations (5) and (6) give all the neutrino interactions within the SM. In particular, Eq. (6) determines the decay width of the Z^0 boson into neutrinos which is proportional to the number of light left-handed neutrinos. At present the measurement of the invisible Z width yields $N_\nu = 2.984 \pm 0.008$ [14] making the existence of three, and only three, light (that is, $m_\nu \leq m_Z/2$) active neutrinos an experimental fact.

An important feature of the SM, which is relevant to the question of the neutrino mass is the fact that the SM with the gauge symmetry of Eq. (1) and the particle content of Eq. (2) presents an accidental global symmetry:

$$G^{\rm global}_{\rm SM} = U(1)_B \times U(1)_e \times U(1)_\mu \times U(1)_\tau. \tag{7}$$

$U(1)_B$ is the baryon number symmetry, and $U(1)_{e,\mu,\tau}$ are the three lepton flavor symmetries, with total lepton number given by $L = L_e + L_\mu + L_\tau$. It is an accidental symmetry because we do not impose it. It is a consequence of the gauge symmetry and the representations of the physical states.

In the SM fermion masses arise from the Yukawa interactions which couple a right-handed fermion with its left-handed doublet and the Higgs field (i, j are generation index),

$$-\mathcal{L}_{\rm Yukawa} = Y^d_{ij} \overline{Q_{Li}} \phi D_{Rj} + Y^u_{ij} \overline{Q_{Li}} \tilde{\phi} U_{Rj} + Y^\ell_{ij} \overline{L_{Li}} \phi E_{Rj} + \text{h.c.}, \tag{8}$$

(where $\tilde{\phi} = i\tau_2\phi^\star$) and after spontaneous symmetry breaking generates a mass for fermions f

$$m^f_{ij} = Y^f_{ij} \frac{v}{\sqrt{2}}\ . \tag{9}$$

However, since no right-handed neutrinos exist in the model, the Yukawa interactions of Eq. (8) leave the neutrinos massless.

One may wonder if neutrino masses could arise from loop corrections or even by nonperturbative effects, however this cannot happen because any neutrino mass term that can be constructed with the SM fields would violate the total lepton symmetry, which, as mentioned above, is a global symmetry of the model so this is not allowed. I will return to this point in the last lecture.

It follows that the SM predicts that neutrinos are precisely massless. In order to add a mass to the neutrino the SM has to be extended.

1.3 Introducing massive neutrinos

As discussed above with the fermionic content and gauge symmetry of the SM one cannot construct a renomalizable mass term for the neutrinos. So in order to introduce a neutrino mass one must either extend the particle contents of the model or abandon gauge invariance and/or renormalizability. I will go back to this point in the last lecture.

Here I will assume that we want to keep the gauge symmetry and the renormalizability condition and we are going to explore the possibilities that we have to introduce a neutrino mass term if one adds to the SM an arbitrary number m of sterile neutrinos $\nu_{si}(1,1,0)$.

As we are going to see, related to the way we introduce the neutrino mass, it comes the fact that for the neutrino because it is the only neutral fermion, one can ask the question of whether a neutrino is a different particle than the antineutrino or they are both the same state.

If the neutrino is a different particle than the antineutrino we say that the neutrino is a *Dirac*-type particle, similar to any of the other charged fermions in the theory. Neutrino and antineutrino are then described by two different fields which involve two sets of creation-annihilation operators. If the neutrino and antineutrino are the same particle we say that the neutrino is a *Majorana*-type particle. This implies that there is only one field which describes both states and involves only one set of creation-annihilation operators. Mathematically this implies that it must be verified that:

$$\nu(x) = \nu^c(x) \tag{10}$$

Here ν^c indicates a charge conjugated field, $\nu^c \equiv C\overline{\nu}^T$ and C is the charge conjugation matrix. Notice that this condition implies that there is only one field which describes both neutrino and antineutrino states. Thus a Majorana neutrino can be described by a two-component spinor unlike the charged fermions, which are Dirac particles, and are represented by four-component spinors.

With the particle contents of the SM and the addition of an arbitrary m number of sterile neutrinos one can construct two types of mass terms that arise from *renormalizable* terms:

$$-\mathcal{L}_{M_\nu} = M_{Dij}\bar{\nu}_{si}\nu_{Lj} + \frac{1}{2}M_{Nij}\bar{\nu}_{si}\nu^c_{sj} + \text{h.c.}. \tag{11}$$

M_D is a complex $m \times 3$ matrix and M_N is a symmetric matrix of dimension $m \times m$.

The first term is a Dirac mass term. It is generated after spontaneous electroweak symmetry breaking from Yukawa interactions

$$Y^\nu_{ij}\bar{\nu}_{si}\tilde{\phi}^\dagger L_{Lj} \Rightarrow M_{Dij} = Y^\nu_{ij}\frac{v}{\sqrt{2}} \tag{12}$$

similarly to the charged fermion masses. It conserves total lepton number but it breaks the lepton flavor number symmetries.

The second term in Eq. (11) is a Majorana mass term. It is different from the Dirac mass terms in many important aspects. It is a singlet of the SM gauge group. Therefore, it can appear as a bare mass term. Furthermore, since it involves two neutrino fields, it breaks lepton number by two units. More generally, such a term is allowed only if the neutrinos carry no additive conserved charge.

In general Eq. (11) can be rewritten as:

$$-\mathcal{L}_{M_\nu} = \frac{1}{2}\overline{\vec{\nu}^c}M_\nu\vec{\nu} + \text{h.c.}\,, \tag{13}$$

where

$$M_\nu = \begin{pmatrix} 0 & M_D^T \\ M_D & M_N \end{pmatrix}, \tag{14}$$

and $\vec{\nu} = (\vec{\nu}_L, \vec{\nu_s^c})^T$ is a $(3+m)$-dimensional vector. The matrix M_ν is complex and symmetric. It can be diagonalized by a unitary matrix of dimension $(3+m)$, V^ν, so that

$$(V^\nu)^T M_\nu V^\nu = \mathrm{diag}(m_1, m_2, \ldots, m_{3+m})\,. \tag{15}$$

In terms of the resulting $3+m$ mass eigenstates

$$\vec{\nu}_{\text{mass}} = (V^\nu)^\dagger \vec{\nu}\,, \tag{16}$$

Eq. (13) can be rewritten as:

$$-\mathcal{L}_{M_\nu} = \frac{1}{2}\sum_{k=1}^{3+m} m_k \left(\bar{\nu}^c_{\text{mass},k}\nu_{\text{mass},k} + \bar{\nu}_{\text{mass},k}\nu^c_{\text{mass},k}\right) = \frac{1}{2}\sum_{k=1}^{3+m} m_k \bar{\nu}_{Mk}\nu_{Mk}\,, \tag{17}$$

where

$$\nu_{Mk} = \nu_{\text{mass},k} + \nu^c_{\text{mass},k} = ({V^\nu}^\dagger \vec{\nu})_k + ({V^\nu}^\dagger \vec{\nu})^c_k \tag{18}$$

which clearly obey the Majorana condition Eq. (10).

From Eq. (18) we find that the weak-doublet components of the neutrino fields are:

$$\nu_{Li} = P_L \sum_{j=1}^{3+m} V^\nu_{ij}\nu_{Mj} \quad i = 1,2,3\,, \tag{19}$$

where P_L is the left-handed projector.

There are three interesting cases, differing in the hierarchy of scales between M_N and M_D:

(1) The scale of the mass eigenvalues of M_N is much higher than the scale of electroweak symmetry breaking $\langle\phi\rangle$. In this case the scale of the mass eigenvalues of M_N is much higher than the scale of electroweak symmetry breaking $\langle\phi\rangle$. The diagonalization of M_ν leads to three light, ν_l, and m heavy, N, neutrinos:

$$-\mathcal{L}_{M_\nu} = \frac{1}{2}\bar{\nu}_l M^l \nu_l + \frac{1}{2}\bar{N} M^h N \tag{20}$$

with

$$M^l \simeq -V_l^T M_D^T M_N^{-1} M_D V_l, \qquad M^h \simeq V_h^T M_N V_h \tag{21}$$

and

$$V^\nu \simeq \begin{bmatrix} \left(1 - \frac{1}{2}M_D^\dagger M_N^*{}^{-1} M_N^{-1} M_D\right) V_l & M_D^\dagger M_N^*{}^{-1} V_h \\ -M_N^{-1} M_D V_l & \left(1 - \frac{1}{2}M_N{}^{-1} M_D M_D^\dagger M_N^*{}^{-1}\right) V_h \end{bmatrix} \tag{22}$$

where V_l and V_h are 3×3 and $m\times m$ unitary matrices respectively. So the heavier are the heavy states, the lighter are the light ones. This is the *see-saw mechanism* [15–19]. Also, as seen from Eq. (22), the heavy states are mostly right-handed while the light ones are mostly left-handed. Both the light and the heavy neutrinos are Majorana particles. Two well-known examples of extensions of the SM leading to a see-saw mechanism for neutrino masses are SO(10) Grand Unified Theories [16, 17] and left-right symmetry [19]. In this case the SM is a good effective low energy theory.

(2) The scale of some eigenvalues of M_N is not higher than the electroweak scale. Now the SM is not even a good low energy effective theory: there are more than three light neutrinos, and they are mixtures of doublet and singlet fields. Again both light fields and the heavy ones are all of the Majorana-type.

(3) $M_N = 0$. This is equivalent to imposing lepton number symmetry on this model. Again, the SM is not even a good low energy theory: both the fermionic content and the assumed symmetries are different. Now only the first term in Eq. (11) is present, which is a Dirac mass term. It is generated by

the Higgs mechanism in the same way that charged fermions masses are generated. If indeed it is the only neutrino mass term present and $m = 3$, the six massive Majorana neutrinos combine to form three massive neutrino Dirac states, equivalently to the charged fermions. Technically in this particular case the 6×6 diagonalizing matrix in Eq. (15) is block diagonal and it can be written in terms of two 3×3 unitary matrices, here denoted by V^ν and V^ν_R, such that

$$V^{\nu\dagger}_R M_D V^\nu = \mathrm{diag}(m_1, m_2, m_3)\,. \tag{23}$$

So the neutrino mass term can be written as:

$$-\mathcal{L}_{M_\nu} = \sum_{k=1}^{3} m_k \bar{\nu}_{Dk} \nu_{Dk} \tag{24}$$

where

$$\nu_{Dk} = (V^{\nu\dagger} \vec{\nu}_L)_k + (V^{\nu\dagger}_R \vec{\nu}_s)_k\,. \tag{25}$$

So in this we identify the three sterile neutrinos with the right handed component of a four-component spinor neutrino field while the weak-doublet components of the neutrino fields are

$$\nu_{Li} = P_L \sum_{j=1}^{3} V^\nu_{ij} \nu_{Dj}\,, \qquad i = 1, 2, 3\,. \tag{26}$$

As we will see the analysis of neutrino oscillations is the same whether the light neutrinos are of the Majorana- or Dirac-type. Only in the discussion of neutrinoless double beta decay the question of Majorana versus Dirac neutrinos is crucial.

1.4 Lepton mixing

The possibility of arbitrary mixing between two massive neutrino states was first introduced in Ref. [20]. In the general case general, we denote the neutrino mass eigenstates by $(\nu_1, \nu_2, \nu_3, \ldots, \nu_n)$ where $n = 3 + m$, and the charged lepton mass eigenstates by (e, μ, τ). The corresponding interaction eigenstates are denoted by (e^I, μ^I, τ^I) and $\vec{\nu} = (\nu_{Le}, \nu_{L\mu}, \nu_{L\tau}, \nu_{s1}, \ldots, \nu_{sm})$. In the mass basis, leptonic charged current interactions are given by

$$-\mathcal{L}_{\mathrm{CC}} = \frac{g}{\sqrt{2}} (\overline{e_L}\ \overline{\mu_L}\ \overline{\tau_L}) \gamma^\mu U \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \cdot \\ \cdot \\ \nu_n \end{pmatrix} W^+_\mu - \text{h.c.}. \tag{27}$$

Here U is a $3 \times n$ matrix which verifies

$$U U^\dagger = I_{3\times 3} \tag{28}$$

but in general $U^\dagger U \neq I_{n\times n}$.

Given the charged lepton mass matrix M_ℓ and the neutrino mass matrix M_ν in some interaction basis,

$$-\mathcal{L}_M = (\overline{e^I_L}\ \overline{\mu^I_L}\ \overline{\tau^I_L})\ M_\ell \begin{pmatrix} e^I_R \\ \mu^I_R \\ \tau^I_R \end{pmatrix} + \frac{1}{2} \overline{\vec{\nu}^c} M_\nu \vec{\nu} + \text{h.c.}\,, \tag{29}$$

we can find the diagonalizing matrices V^ℓ and V^ν:

$$V^{\ell\dagger} M_\ell M_\ell^\dagger V^\ell = \mathrm{diag}(m_e^2, m_\mu^2, m_\tau^2), \quad V^{\nu\dagger} M_\nu^\dagger M_\nu V^\nu = \mathrm{diag}(m_1^2, m_2^2, m_3^2, \ldots, m_n^2). \tag{30}$$

Here V^ℓ is a unitary 3×3 matrix while V^ν the $n\times n$ unitary matrix in Eq. (15). The $3\times n$ mixing matrix U can be found from these diagonalizing matrices:

$$U_{ij} = P_{\ell,ii}\, V^{\ell\,\dagger}_{ik}\, V^\nu_{kj}\, (P_{\nu,jj}). \tag{31}$$

P_ℓ is a diagonal 3×3 phase matrix, that is introduce to reduce by three the number of phases in U. P_ν is a diagonal matrix with additional arbitrary phases (chosen to reduce the number of phases in U) only for Dirac states. For Majorana neutrinos, this matrix is simply a unit matrix. The reason for that is that if one rotates a Majorana neutrino by a phase, this phase will appear in its mass term which will no longer be real. Thus, the number of phases that can be absorbed by redefining the mass eigenstates depends on whether the neutrinos are Dirac or Majorana particles. In particular, if there are only three Majorana neutrinos, U is a 3×3 matrix analogous to the Cabibbo-Kobayashi-Maskawa (CKM) matrix for the quarks [21,22] but due to the Majorana nature of the neutrinos it depends on six independent parameters: three mixing angles and three phases. This is to be compared to the case of three Dirac neutrinos [1] where the number of physical phases is one, similarly to the CKM matrix. Note, however, that the two extra Majorana phases are very hard to measure since they are only physical if neutrino mass is non-zero and therefore the amplitude of any process involving them is suppressed a factor m_ν/E to some power where E is the energy involved in the process which is typically much larger than the neutrino mass. The most sensitive experimental probe of Majorana phases is the rate of neutrinoless $\beta\beta$ decay. If no new interactions for the charged leptons are present we can identify their interaction eigenstates with the corresponding mass eigenstates after phase redefinition. In this case the charged current lepton mixing matrix U is simply given by a $3\times n$ sub-matrix of the unitary matrix V^ν. It worth noticing that while for the case of 3 light Dirac neutrinos the procedure leads to a fully unitary U matrix for the light states, generically for three light Majorana neutrinos this is not the case when the full spectrum contains heavy neutrino states which have been integrated out as can be seen, from Eq. (22). However, as seen in Eq. (22), the unitarity violation is of the order $\mathcal{O}(M_D/M_N)$ and it is expected to be very small (at it is also severely constrained experimentally). Consequently in the analysis of oscillation data presented in next lectures the U matrix is assumed to be unitary.

1.5 Laboratory probes of ν mass scale and its nature

Kinematic constraints from weak decays

It was Fermi who first proposed a kinematic search for the neutrino mass from the hard part of the beta spectra in ^{3}H beta decay $^3\text{H}\to{}^3\text{He}+e^- + \overline{\nu}_e$. This is a superallowed transition, which means that the nuclear matrix elements do not generate any energy dependence, so that the electron spectrum is given by the phase space alone

$$\frac{dN}{dT} = CpE(Q-T)\,\sqrt{(Q-T)^2 - m_\nu^2}\,F(E)\ . \tag{32}$$

where $E = T + m_e$, Q is the maximum energy and $F(E)$ is the Fermi function which incorporates final state Coulomb interactions.

Plotted in terms of the Kurie function $K(T) \equiv \sqrt{\frac{dN}{dT}\frac{1}{pEF(E)}}$ a non-vanishing neutrino mass m_ν provokes a distortion from the straight-line T-dependence at the end point: for $m_\nu = 0 \to T_{\max} = Q$ whereas for $m_\nu \neq 0 \to T_{\max} = Q - m_\nu$ as illustrated in Fig. 1. ^{3}H beta decay has a a very small energy release $Q = 18.6\ KeV$ which makes it particularly sensitive to this kinematic effect. In the presence of mixing these limits have to be modified and in general they involve more than one flavor parameter. For neutrinos with small mass differences the distortion of the beta spectrum can be described by the single

[1] In this case, as discussed above the 6×6 neutrino diagonalizing matrix is block diagonal and the V^ν in Eq. (31) is the 3×3 block introduced in Eq. (23).

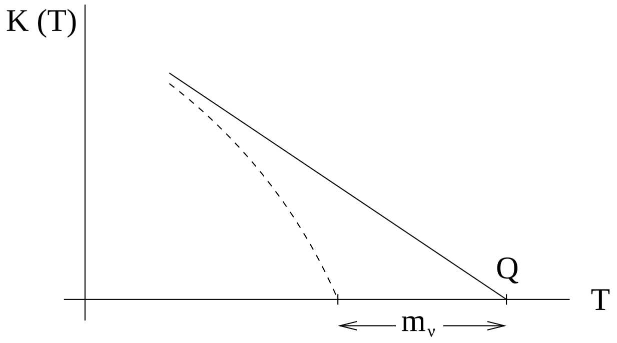


Fig. 1: Kinematic determination of m_ν

parameter substituting m_ν by

$$\left(m^{\rm eff}_{\nu_e}\right)^2 = \sum_i m_i^2 |U_{ei}|^2 \tag{33}$$

The most recent result on the kinematic search for neutrino mass in tritium decay is from KATRIN [23], an experiment that so far has found no indication of $m_{\nu_e} \neq 0$ and sets an upper limit

$$m^{\rm eff}_{\nu_e} < 1.1 \text{ eV}\,, \tag{34}$$

at 90% CL improving over the previous bound from the Mainz [24, 25] and Troitsk [26] experiments which constrained $m^{\rm eff}_{\nu_e} < 2.2$ eV at 95% CL. KATRIN continues running with an estimated sensitivity limit of $m^{\rm eff}_{\nu_e} \sim 0.2$ eV.

For the other flavours the present limits are [14]

$$m^{\rm eff}_{\nu_\mu} = \sqrt{\sum_i m_i^2 |U_{\mu i}|^2} < 190 \text{ keV} \quad (90\% \text{ CL}) \qquad \text{from} \qquad \pi^- \to \mu^- + \overline{\nu}_\mu \tag{35}$$

$$m^{\rm eff}_{\nu_\tau} = \sqrt{\sum_i m_i^2 |U_{\tau i}|^2} < 18.2 \text{ MeV} \quad (95\% \text{ CL}) \qquad \text{from} \qquad \tau^- \to n\pi + \nu_\tau \tag{36}$$

Thus, in the presence of non-vanishing mixing the most stringent constraint on the absolute mass of any of the neutrinos is set by the limit from tritium beta decay in Eq. (34).

Dirac vs Majorana: neutrinoless double-beta decay

The most sensitive probe to whether neutrinos are Dirac or Majorana states is the neutrinoless double beta decay ($0\nu\beta\beta$):

$$(A, Z) \to (A, Z + 2) + e^- + e^-\,. \tag{37}$$

In the presence of neutrino masses and mixing the process in Eq.(37) can be generated at lower order in perturbation theory by the term represented by the diagram in Fig. 2 The amplitude of this process is proportional to the product of the two leptonic currents

$$M_{\alpha\beta} \propto [\bar e \gamma_\alpha (1-\gamma_5)\nu_e]\ [\bar e \gamma_\beta (1-\gamma_5)\nu_e] \propto \sum_i (U_{ei})^2\, [\bar e \gamma_\alpha (1-\gamma_5)\nu_i]\ [\bar e \gamma_\beta (1-\gamma_5)\nu_i]\ . \tag{38}$$

The neutrino propagator in Fig. 2 can only arise from the contraction $\langle 0\,|\,\nu_i(x)\nu_i(y)^T\,|\,0\rangle$. But if the neutrino is a Dirac particle ν_i field annihilates a neutrino states and creates an antineutrino state which are different, so the contraction $\langle 0\,|\,\nu_i(x)\nu_i(y)^T\,|\,0\rangle = 0$ and $M_{\alpha\beta} = 0$. On the other hand, if ν_i is a Majorana particle, neutrino and antineutrino are described by the same field and $\langle 0\,|\,\nu_i(x)\nu_i(y)^T\,|\,0\rangle \neq 0$.

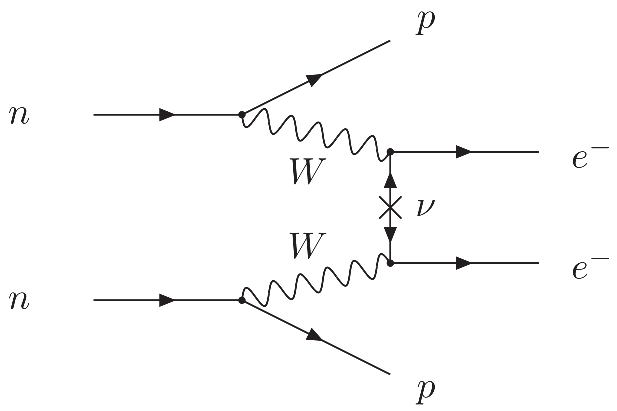


Fig. 2: Feynman diagram for neutrinoless double-beta decay.

The conclusion is that in order to induce the $0\nu\beta\beta$ decay, neutrinos must be Majorana particles. This is consistent with the fact that the process (37) violates total lepton number by two units. Conversely, if $0\nu\beta\beta$ decay is observed, massive neutrinos cannot be exact Dirac states [27].

After some algebra one finds that the rate of the process is proportional to the *effective Majorana mass of* ν_e,

$$m_{ee} = \left| \sum_i m_i U_{ei}^2 \right| \tag{39}$$

which, in addition to the masses and mixing parameters that affect the tritium beta decay spectrum, depends also on the leptonic CP violating phases.

The observable determined by the experiments is the half-life of the decay. Under the assumption that the Majorana neutrino mass is the only source of lepton number violation at low energies, the decay half-life is given by:

$$(T_{1/2}^{0\nu})^{-1} = G^{0\nu} \left| M^{0\nu} \right|^2 \left(\frac{m_{ee}}{m_e} \right)^2 , \tag{40}$$

where $G^{0\nu}$ is the phase space integral taking into account the final atomic state, and $|M^{0\nu}\ |$ is the nuclear matrix element of the transition.

At present the strongest bound on $0\nu\beta\beta$ decay lifetime comes from the search in KamLAND-Zen experiment [28] which uses 13 Tons of Xe-loaded liquid scintillator to search for the decay $0\nu\beta\beta$ of ^{136}Xe and has set a bound on the half-life of $T_{1/2}^{0\nu} > 1.07 \times 10^{26}$ yr at 90% CL. From Eq. (40) we see that nuclear structure details enter relation between the decay rate (or lifetime) and the effective Majorana mass. As a consequence uncertainties in the nuclear structure calculations result in a spread of m_{ee} values for a given $T_{1/2}^{0\nu}$ by a factor of 2–3 [29]. Using a variety of nuclear matrix element calculations, the corresponding upper bound on the effective Majorana mass is

$$m_{ee} < 61 - 165 \text{ meV} . \tag{41}$$

This bound is stronger than the one from tritium beta decay but it is model dependent because it requires that neutrinos are Majorana particles and that their mass is the only source of lepton number violation generating neutrinoless double beta decay.

Cosmological bounds

Neutrinos, like any other particles, contribute to the total energy density of the Universe. Furthermore light neutrinos are relativist through most of the evolution of the Universe. As a consequence they can play a relevant role in large scale structure formation and leave clear signatures in several cosmological observables.

Within what we presently know of their masses, neutrinos are relativistic through most of the evolution of the Universe and being very weakly interacting they decoupled early in cosmic history. Depending on their exact masses they can impact the cosmic microwave background spectra, in particular by altering the value of the redshift for matter-radiation equality. More importantly, their free streaming suppresses the growth of structures on scales smaller than the horizon at the time when they become non-relativistic and therefore affects the matter power spectrum which is probed from surveys of the Large Scale Structure distribution. Because of these effects it is possible to infer constraints, although indirect, on the neutrino masses by comparing the most recent cosmological data with the current theoretical predictions.

The relevant quantity in these studies is the total neutrino energy density in our Universe, $\Omega_\nu h^2$ (where h is the Hubble constant normalized to $H_0 = 100$ km s^{-1} Mpc^{-1}). At present $\Omega_\nu h^2$ is related to the total mass in the form of neutrinos

$$\Omega_\nu h^2 = \sum_i m_i/(94\text{eV}) \; . \tag{42}$$

Therefore cosmological data gives information on the sum of the neutrino masses and has very little to say on their mixing.

Because of these effects, the recent precise astrophysical and cosmological observations can provide indirect upper limits on absolute neutrino masses competitive with those from laboratory experiments. At present the most robust bounds come from the analysis of Planck results which within the Λ-Cold-Dark-Matter model imply $\sum_i m_i \leq 0.17 - 0.74$ eV where the range includes variations of the data sets included in the analysis. One must always keep in mind that these bounds apply *within a given cosmological model* and consequently variations of the model can relax the bounds.

1.6 Summary

In the SM neutrinos are purely left-handed and strictly massless. Neutrino masses can be introduced in the model at the expense of adding new right-handed – hence sterile – states, and/or breaking total lepton number. Depending on the way the mass term is introduced, the massive neutrinos are Dirac particles, as any other fermions of the SM for which neutrinos and antineutrinos are different states, or Majorana particles, being their own antiparticles. In this second case one may gain an understanding of why neutrino masses are smaller than other fermion masses. Massive neutrinos open up the possibility of flavour mixing and CP violation in the lepton sector similar to the quark sector. So far direct searches for neutrino masses have result only into limits, the strongest model independent bound is $\sim$ eV from tritium β decay.

2 LECTURE II: Flavour oscillations

2.1 Mass-induced flavour oscillations in vacuum

If neutrinos have masses and lepton flavours are mixed in the weak CC interactions, lepton flavour is not conserved in neutrino propagation [4,5]. This phenomenon is usually referred to as *neutrino oscillations*. In brief, a weak eigenstates, ν_α, which by default is the state produced in the weak CC interaction of a charged lepton ℓ_α, is the linear combination determined by the mixing matrix U

$$|\nu_\alpha\rangle = \sum_{i=1}^{n} U^*_{\alpha i} |\nu_i\rangle \,, \tag{43}$$

where ν_i are the mass eigenstates and here n is the number of light neutrino species (implicit in our definition of the state $|\nu\rangle$ is its energy-momentum and space-time dependence). After traveling a distance L ($L \simeq ct$ for relativistic neutrinos), that state evolves as:

$$|\nu_\alpha(t)\rangle = \sum_{i=1}^{n} U^*_{\alpha i} |\nu_i(t)\rangle \,. \tag{44}$$

This neutrino can then undergo a charged-current (CC) interaction producing a charge lepton ℓ_β, $\nu_\alpha(t)N' \to \ell_\beta N$, with a probability

$$P_{\alpha\beta} = |\langle \nu_\beta|\nu_\alpha(t)\rangle|^2 = |\sum_{i=1}^{n}\sum_{j=1}^{n} U^*_{\alpha i} U_{\beta j} \langle \nu_j|\nu_i(t)\rangle|^2 \,. \tag{45}$$

Assuming that $|\nu\rangle$ is a plane wave, $|\nu_i(t)\rangle = e^{-i\,E_i t}|\nu_i(0)\rangle$, [2] with $E_i = \sqrt{p_i^2 + m_i^2}$ and m_i being, respectively, the energy and the mass of the neutrino mass eigenstate ν_i. In all practical cases neutrinos are very relativistic ,so $p_i \simeq p_j \equiv p \simeq E$. We can then write

$$E_i = \sqrt{p_i^2 + m_i^2} \simeq p + \frac{m_i^2}{2E} \,, \tag{46}$$

and use the orthogonality of the mass eigenstates, $\langle \nu_j|\nu_i\rangle = \delta_{ij}$, to arrive to the following form for $P_{\alpha\beta}$:

$$P_{\alpha\beta} = \delta_{\alpha\beta} - 4\sum_{i<j}^{n} \text{Re}[U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j}] \sin^2 X_{ij} + 2\sum_{i<j}^{n} \text{Im}[U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j}] \sin 2X_{ij} \,, \tag{47}$$

where

$$X_{ij} = \frac{(m_i^2 - m_j^2)L}{4E} = 1.267 \, \frac{\Delta m^2_{ij}}{\text{eV}^2} \, \frac{L/E}{\text{m/MeV}} \,. \tag{48}$$

If we had made the same derivation for antineutrino states we would have ended with a similar expression but with the exchange $U \to U^*$. Consequently we conclude that the first term in the right-hand-side of Eq. (47) is CP conserving since it is the same for neutrinos and antineutrinos, while the last one is CP violating because it has opposite sign for neutrinos and antineutrinos.

Equation (47) oscillates in distance with oscillation lengths

$$L^{\text{osc}}_{0,ij} = \frac{4\pi E}{|\Delta m^2_{ij}|} \,, \tag{49}$$

[2] For a pedagogical discussion of the quantum mechanical description of flavour oscillations in the wave package approach see for example Ref. [8]. A recent review of the quantum mechanical aspects and subtleties on neutrino oscillations can be found in in Ref. [30].

and with amplitudes proportional to products of elements in the mixing matrix. Thus, neutrinos must have different masses ($\Delta m^2_{ij} \neq 0$) and they must have not vanishing mixing ($U_{\alpha_i} U_{\beta i} \neq 0$) in order to undergo flavour oscillations. Also, from Eq. (47) we see that the Majorana phases cancel out in the oscillation probability. This is expected because flavour oscillation is a total lepton number conserving process.

Ideally, a neutrino oscillation experiment would like to measure an oscillation probability over a distance L between the source and the detector, for neutrinos of a definite energy E. In practice, neutrino beams, both from natural or artificial sources, are never monoenergetic, but have an energy spectrum $\Phi(E)$. In addition each detector has a finite energy resolution. Under these circumstances what is measured is an average probability

$$\begin{aligned}\langle P_{\alpha\beta}\rangle &= \frac{\int dE \frac{d\Phi}{dE}\sigma(E) P_{\alpha\beta}(E)\epsilon(E)}{\int dE \frac{d\Phi}{dE}\sigma_{CC}(E)\epsilon(E)} \\ &= \delta_{\alpha\beta} - 4\sum_{i<j}^{n} \mathrm{Re}[U_{\alpha i}U^*_{\beta i}U^*_{\alpha j}U_{\beta j}]\langle \sin^2 X_{ij}\rangle + 2\sum_{i<j}^{n} \mathrm{Im}[U_{\alpha i}U^*_{\beta i}U^*_{\alpha j}U_{\beta j}]\langle \sin 2X_{ij}\rangle\,.\end{aligned} \tag{50}$$

σ is the cross section for the process in which the neutrino flavour is detected, and $\epsilon(E)$ is the detection efficiency. The minimal range of the energy integral is determined by the energy resolution of the experiment.

It is clear from the above expression that if $(E/L) \gg |\Delta m^2_{ij}|$ ($L \ll L^{\rm osc}_{0,ij}$) so $\sin^2 X_{ij} \ll 1$, the oscillation phase does not give any appreciable effect. Conversely if $L \gg L^{\rm osc}_{0,ij}$, many oscillation cycles occur between production and detection so the oscillating term is averaged to $\langle \sin^2 X_{ij}\rangle = 1/2$.

We summarize in Table 1. the typical values of L/E for different types of neutrino sources and experiments and the corresponding ranges of Δm^2 to which they can be most sensitive.

Table 1: Characteristic values of L and E for experiments performed using various neutrino sources and the corresponding ranges of $|\Delta m^2|$ to which they can be most sensitive to flavour oscillations in vacuum. SBL stands for short baseline and LBL for long baseline.

Experiment		L (m)	E (MeV)	$\|\Delta m^2\|$ (eV^2)
Solar		10^{10}	1	10^{-10}
Atmospheric		$10^4 - 10^7$	10^2–10^5	$10^{-1} - 10^{-4}$
Reactor	SBL	$10^2 - 10^3$	1	$10^{-2} - 10^{-3}$
	LBL	$10^4 - 10^5$		$10^{-4} - 10^{-5}$
Accelerator	SBL	10^2	10^3–10^4	> 0.1
	LBL	$10^5 - 10^6$	$10^3 - 10^4$	$10^{-2} - 10^{-3}$

Historically, the results of neutrino oscillation experiments were interpreted assuming two-neutrino states so there is only one oscillating phase, the mixing matrix depends on a single mixing angle θ and no CP violation effect in oscillations is possible. At present, as we will discuss in the third lecture we need at least the mixing among three-neutrino states to fully describe the bulk of experimental results. However, in many cases, the observed results can be understood in terms of oscillations dominantly driven by one Δm^2. In this limit $P_{\alpha\beta}$ of Eq. (47) takes the form [5]

$$P_{\alpha\beta} = \delta_{\alpha\beta} - (2\delta_{\alpha\beta} - 1)\sin^2 2\theta \sin^2 X\,. \tag{51}$$

In this effective $2-\nu$ limit, changing the sign of the mass difference, $\Delta m^2 \to -\Delta m^2$, and changing the octant of the mixing angle, $\theta \to \frac{\pi}{2} - \theta$, is just redefining the mass eigenstates, $\nu_1 \leftrightarrow \nu_2$: $P_{\alpha\beta}$ must be invariant under such transformation. So the physical parameter space can be covered with either $\Delta m^2 \geq 0$ with $0 \leq \theta \leq \frac{\pi}{2}$, or, alternatively, $0 \leq \theta \leq \frac{\pi}{4}$ with either sign for Δm^2.

However, from Eq. (51) we see that $P_{\alpha\beta}$ is actually invariant under the change of sign of the mass splitting *and* the change of octant of the mixing angle separately. This implies that there is a two-fold discrete ambiguity since the two different sets of physical parameters, $(\Delta m^2, \theta)$ and $(\Delta m^2, \frac{\pi}{2} - \theta)$, give the same transition probability in vacuum. In other words, one could not tell from a measurement of, say, $P_{e\mu}$ in vacuum whether the larger component of ν_e resides in the heavier or in the lighter neutrino mass eigenstate. This symmetry is broken when one considers mixing of three or more neutrinos in the flavour evolution and/or when the neutrinos traverse regions of dense matter as we describe in the following.

2.2 Propagation of massive neutrinos in matter

When neutrinos propagate in dense matter, the interactions with the medium affect their properties. These effects are either coherent or incoherent. For purely incoherent inelastic ν-p scattering, the characteristic cross section is very small:

$$\sigma \sim \frac{G_F^2 s}{\pi} \sim 10^{-43}\mathrm{cm}^2 \left(\frac{E}{1\ \mathrm{MeV}}\right)^2 . \tag{52}$$

The smallness of this cross section is demonstrated by the fact that if a beam of 10^{10} neutrinos with $E \sim 1$ MeV was aimed at the Earth, only one would be deflected by the Earth matter. It may seem then that for neutrinos matter is irrelevant. However, one must take into account that Eq. (52) does not contain the contribution from forward elastic coherent interactions. In coherent interactions, the medium remains unchanged and it is possible to have interference of scattered and unscattered neutrino waves which enhances the effect. Coherence further allows one to decouple the evolution equation of the neutrinos from the equations of the medium. In this approximation, the effect of the medium is described by an effective potential which depends on the density and composition of the matter [31].

For example the effective potential for the evolution of ν_e in a medium with electrons, protons and neutrons due to its CC interactions is given by (a detailed derivation of this result can be found, for instance, in Refs. [8, 11, 12])

$$V_C = \sqrt{2} G_F N_e . \tag{53}$$

where N_e is the electron number density. For $\overline{\nu_e}$ the sign of V_C is reversed. This potential can also be expressed in terms of the matter density ρ:

$$V_C = \sqrt{2} G_F N_e \simeq 7.6\, Y_e \frac{\rho}{10^{14}\mathrm{g/cm}^3}\ \mathrm{eV} , \tag{54}$$

where $Y_e = \frac{N_e}{N_p + N_n}$ is the relative number density. Three examples that are relevant to observations are the following:

- At the Earth core $\rho \sim 10$ g/cm^3 and $V_C \sim 10^{-13}$ eV;
- At the solar core $\rho \sim 100$ g/cm^3 and $V_C \sim 10^{-12}$ eV

In the same way we can obtain the effective potentials for any flavour neutrino or antineutrino due to interactions with different particles in the medium. For ν_μ and ν_τ, $V_C = 0$ for most media while for any active neutrino the effective potential due to NC interactions in neutral medium is $V_N = -1/\sqrt{2} G_F N_n$ where N_n is the number density of neutrons.

There are several derivations in the literature of the evolution equation of a neutrino system in matter (see, for instance, Refs. [32–34]). In here we start by considering a state which is an admixture of two neutrino species $|\nu_\alpha\rangle$ and $|\nu_\beta\rangle$ or, equivalently, of $|\nu_1\rangle$ and $|\nu_2\rangle$:

$$|\Phi(x)\rangle = \Phi_\alpha(x)|\nu_\alpha\rangle + \Phi_\beta(x)|\nu_\beta\rangle = \Phi_1(x)|\nu_1\rangle + \Phi_2(x)|\nu_2\rangle \tag{55}$$

We decompose the neutrino wave function: $\Phi_i(x) = \nu_i(x)\phi_i(x)$ where $\phi_i(x)$ is the spinor part.

The evolution of Φ in a medium is described by a system of coupled Dirac equations, but after several approximations the spinorial part can be drop out and we end up with an equation which can be written in matrix form as [31]:

$$-\,i\frac{\partial}{\partial x}\begin{pmatrix}\nu_\alpha\\ \nu_\beta\end{pmatrix}=\left(-\frac{M_w^2}{2E}\right)\begin{pmatrix}\nu_\alpha\\ \nu_\beta\end{pmatrix}\,, \tag{56}$$

where we have defined an effective mass matrix in matter:

$$M_w^2=\begin{pmatrix}\frac{m_1^2+m_2^2}{2}+2EV_\alpha-\frac{\Delta m^2}{2}\cos 2\theta & \frac{\Delta m^2}{2}\sin 2\theta\\ \frac{\Delta m^2}{2}\sin 2\theta & \frac{m_1^2+m_2^2}{2}+2EV_\beta+\frac{\Delta m^2}{2}\cos 2\theta\end{pmatrix}\,. \tag{57}$$

Here $\Delta m^2=m_2^2-m_1^2$.

We define the instantaneous mass eigenstates in matter, ν_i^m, as the eigenstates of M_w for a fixed value of x (or t). They are related to the interaction eigenstates through a unitary rotation,

$$\begin{pmatrix}\nu_\alpha\\ \nu_\beta\end{pmatrix}=U(\theta_m)\begin{pmatrix}\nu_1^m\\ \nu_2^m\end{pmatrix}=\begin{pmatrix}\cos\theta_m & \sin\theta_m\\ -\sin\theta_m & \cos\theta_m\end{pmatrix}\begin{pmatrix}\nu_1^m\\ \nu_2^m\end{pmatrix}\,. \tag{58}$$

The eigenvalues of M_w, that is, the effective masses in matter are given by [31,35]:

$$\mu_{1,2}^2(x)=\frac{m_1^2+m_2^2}{2}+E(V_\alpha+V_\beta)\mp\frac{1}{2}\sqrt{(\Delta m^2\cos 2\theta-A)^2+(\Delta m^2\sin 2\theta)^2}\,, \tag{59}$$

while the mixing angle in matter is given by

$$\tan 2\theta_m=\frac{\Delta m^2\sin 2\theta}{\Delta m^2\cos 2\theta-A}. \tag{60}$$

The quantity A is defined by

$$A\equiv 2E(V_\alpha-V_\beta). \tag{61}$$

In Fig. 3 we plot the effective masses and the mixing angle in matter as functions of the potential A, for $A>0$ and $\Delta m^2\cos 2\theta>0$. Notice that even massless neutrinos acquire non-vanishing effective masses in matter. Also the sign of A depends on the composition of the medium and on the flavour composition of the neutrino state considered. From the expressions above we see that for a given sign of A the mixing angle in matter is larger(smaller) than in vacuum if this last one is in the first (second) octant. Therefore the symmetry about 45 degrees which existing in vacuum oscillations between two neutrino states is broken by the matter potential in propagation in a medium.

The expressions above show that very important effects are present when A, is close to $\Delta m^2\cos 2\theta$. In particular, as seen in Eq. (60), the tangent of the mixing angle changes sign if, along its path, the neutrino passes by some matter density region satisfying, for its energy, the *resonance condition*

$$A_R=\Delta m^2\cos 2\theta\,. \tag{62}$$

This implies that if the neutrino is created in a region where the relevant potential satisfies $A_0>A_R$ (A_0 here is the value of the relevant potential at the production point), then the effective mixing angle in matter at the production point is such that $\mathrm{sgn}(\cos 2\theta_{m,0})=-\mathrm{sgn}(\cos 2\theta)$. So the flavour component of the mass eigenstates is inverted as compared to their composition in vacuum. In particular, if at production point we have $A_0=2A_R$, then $\theta_{m,0}=\frac{\pi}{2}-\theta$. Asymptotically, for $A_0\gg A_R$, $\theta_{m,0}\to\frac{\pi}{2}$. In other words, if in vacuum the lightest (heaviest) mass eigenstate has a larger projection on the flavour α (β), inside a matter with density and composition such that $A>A_R$, the opposite holds. So if the neutrino system is traveling across a monotonically varying matter potential, the dominant flavour component of a given mass eigenstate changes when crossing the region with $A=A_R$. This phenomenon is known as *level crossing*.

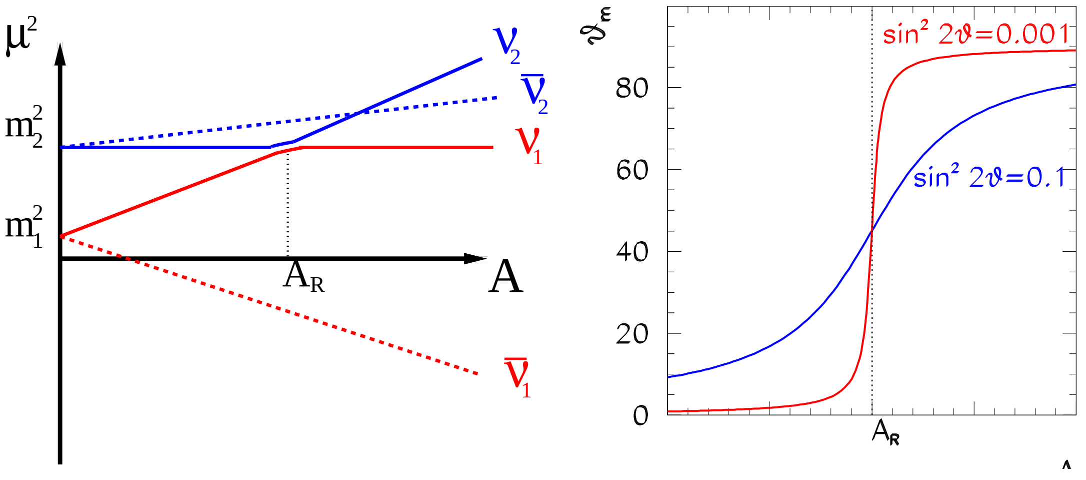


Fig. 3: Effective masses (left) and mixing(right) acquired in the medium by a system of two massive neutrinos as a function of the potential A [see Eq. (59)].

From the expression above we see that the oscillation length in matter,

$$L^{\rm osc} = \frac{L_0^{\rm osc}\Delta m^2}{\sqrt{(\Delta m^2\cos 2\theta - A)^2 + (\Delta m^2 \sin 2\theta)^2}}, \tag{63}$$

where the oscillation length in vacuum, $L_0^{\rm osc}$, was defined in Eq. (49), presents a resonant behaviour. At the resonance point the oscillation length is

$$L_R^{\rm osc} = \frac{L_0^{\rm osc}}{\sin 2\theta}. \tag{64}$$

The width (in distance) of the resonance, δr_R, corresponding to $\delta A_R = 2\Delta m^2 \sin^2 2\theta$ is

$$\delta r_R = \frac{\delta A_R}{|\frac{dA}{dr}|_R} \tag{65}$$

For constant A, *i.e.*, for constant matter density, the evolution of the neutrino system is described just in terms of the masses and mixing in matter. But for varying A, this is in general not the case.

In the general case, taking the time derivative of Eq. (58), we find:

$$\frac{\partial}{\partial t}\begin{pmatrix}\nu_\alpha\\ \nu_\beta\end{pmatrix} = \dot U(\theta_m)\begin{pmatrix}\nu_1^m\\ \nu_2^m\end{pmatrix} + U(\theta_m)\begin{pmatrix}\dot\nu_1^m\\ \dot\nu_2^m\end{pmatrix}. \tag{66}$$

Using the evolution equation in the flavor basis, Eq. (56), we get

$$i\begin{pmatrix}\dot\nu_1^m\\ \dot\nu_2^m\end{pmatrix} = \frac{1}{2E}U^\dagger(\theta_m)M_w^2U(\theta_m)\begin{pmatrix}\nu_1^m\\ \nu_2^m\end{pmatrix} - i\,U^\dagger\dot U(\theta_m)\begin{pmatrix}\nu_1^m\\ \nu_2^m\end{pmatrix}. \tag{67}$$

For constant matter density, θ_m is constant and the second term vanishes. In general, using the definition of the effective masses $\mu_i(t)$ in Eq. (59), and subtracting a diagonal piece $(\mu_1^2+\mu_2^2)/2E \times I$, we can rewrite the evolution equation as:

$$i\begin{pmatrix}\dot\nu_1^m\\ \dot\nu_2^m\end{pmatrix} = \frac{1}{4E}\begin{pmatrix}-\Delta(t) & -4iE\dot\theta_m(t)\\ 4iE\dot\theta_m(t) & \Delta(t)\end{pmatrix}\begin{pmatrix}\nu_1^m\\ \nu_2^m\end{pmatrix} \tag{68}$$

where we defined $\Delta(t) \equiv \mu_2^2(t) - \mu_1^2(t)$.

The evolution equations, Eq. (68), constitute a system of coupled equations: the instantaneous mass eigenstates, ν_i^m, mix in the evolution and are not energy eigenstates. The importance of this effect is controlled by the relative size of the off-diagonal piece $4\,E\,\dot{\theta}_m(t)$ with respect to the diagonal one $\Delta(t)$. When $\Delta(t) \gg 4\,E\,\dot{\theta}_m(t)$, the instantaneous mass eigenstates, ν_i^m, behave approximately as energy eigenstates and they do not mix in the evolution. This is the *adiabatic* transition approximation. From the definition of θ_m in Eq. (60) we find that the adiabaticity condition can be expressed in terms of the adiabaticity parameter Q as

$$\frac{Q}{2} \equiv \frac{\Delta(t)}{4E\dot{\theta}_m(t)} = \frac{\Delta(t)^3}{2EA\Delta m^2 \sin 2\theta}\left|\frac{A}{\dot{A}}\right| \gg 1\;. \tag{69}$$

Since for small mixing angles the maximum of $\dot{\theta}_m$ occurs at the resonance point (as seen in Fig. 3), the strongest adiabaticity condition is obtained when Eq. (69) is evaluated at the resonance

$$Q = \frac{2\,\pi\,\delta r_R}{L_R^{osc}}\;, \tag{70}$$

where we used the definitions of A_R and δr_R in Eqs. (62) and (65). Written in this form, we see that the adiabaticity condition, $Q \gg 1$, implies that many oscillations take place in the resonant region. Conversely, when $Q \leq 1$ the transition is non-adiabatic.

From the expressions above we see that, for example, the amplitude of a ν_α produced in matter at t_0 and exiting the matter at $t > t_0$ as ν_β can be written as follows:

$$\begin{aligned}
&\mathcal{A}(\nu_\alpha \to \nu_\beta; t) = \sum_{i,j} \mathcal{A}(\nu_\alpha(t_0) \to \nu_i(t_0))\;\mathcal{A}(\nu_i(t_0) \to \nu_j(t))\;\mathcal{A}(\nu_j(t) \to \nu_\beta(t)) \\
&\mathcal{A}(\nu_\alpha(t_0) \to \nu_i(t_0)) = \langle \nu_i(t_0)|\nu_\alpha(t_0)\rangle = U_{\alpha i}^*(\theta_{m,0}) \\
&\mathcal{A}(\nu_j(t) \to \nu_\beta(t)) = \langle \nu_\beta(t)|\nu_j(t)\rangle = U_{\beta j}(\theta)
\end{aligned} \tag{71}$$

where $U_{\alpha i}^*(\theta_{m,0})$ is the (αi) element of the mixing matrix in matter at the production point and $U_{\beta j}(\theta)$ is the (βj) element of the mixing matrix in vacuum.

In the adiabatic approximation the mass eigenstates do not mix so

$$\mathcal{A}(\nu_i(t_0) \to \nu_j(t)) = \delta_{ij}\,\langle \nu_i(t)|\nu_i(t_0)\rangle = \delta_{ij}\,\exp\left\{i\int_{t_0}^{t} E_i(t')dt'\right\}\;. \tag{72}$$

Note that E_i is a function of time because the effective mass μ_i is a function of time,

$$E_i(t') \simeq p + \frac{\mu_i^2(t')}{2p}\;. \tag{73}$$

Thus the transition probability for the adiabatic case is given by

$$P(\nu_\alpha \to \nu_\beta; t) = \left|\sum_i U_{\beta i}(\theta) U_{\alpha i}^\star(\theta_{m,0}) \exp\left(-\frac{i}{2E}\int_{t_0}^{t} \mu_i^2(t')dt'\right)\right|^2\;. \tag{74}$$

For the case of two-neutrino mixing Eq. (74) for $\alpha = \beta$ takes the form

$$P(\nu_\alpha \to \nu_\alpha; t) = \cos^2\theta_{m,0}\cos^2\theta + \sin^2\theta_{m,0}\sin^2\theta + \frac{1}{2}\sin 2\theta_{m,0}\sin 2\theta\cos\left(\frac{\delta(t)}{2E}\right)\;, \tag{75}$$

where

$$\delta(t) = \int_{t_0}^{t} \Delta(t')dt' = \int_{t_0}^{t} \sqrt{(\Delta m^2\cos 2\theta - A(t'))^2 + (\Delta m^2 \sin 2\theta)^2 dt'}\;,$$

which, in general, has to be evaluated numerically. There are some analytic approximations for specific forms of $A(t')$: exponential, linear ...(see, for instance, Ref. [36]). For $\delta(t) \gg E$ the last term in Eq. (75) is averaged and the survival probability takes the form

$$P(\nu_\alpha \to \nu_\alpha; t) = \frac{1}{2}\left[1 + \cos 2\theta_{m,0} \cos 2\theta\right] \tag{76}$$

The Mihheev-Smirnov-Wolfenstein effect for solar neutrinos

The matter effects discussed in the previous section are of special relevance for solar neutrinos. As the Sun produces ν_e's in its core, here we shall consider the propagation of a $\nu_e - \nu_X$ neutrino system (X is some superposition of μ and τ, which is arbitrary because ν_μ and ν_τ have only and equal neutral current interactions) in the matter density of the Sun.

The density of solar matter is a monotonically decreasing function of the distance R from the center of the Sun, and it can be approximated by an exponential for $R < 0.9R_\odot$

$$n_e(R) = n_e(0) \exp\left(-R/r_0\right) \ , \tag{77}$$

with $r_0 = R_\odot/10.54 = 6.6 \times 10^7$ m $= 3.3 \times 10^{14}$ eV^{-1}.

As mentioned above, the nuclear reactions in the Sun produce electron neutrinos. After crossing the Sun, the composition of the neutrino state exiting the Sun will depend on the relative size of $\Delta m^2 \cos 2\theta$ versus $A_0 = 2\, E\, G_F\, n_{e,0}$ (here 0 refers to the neutrino production point which is near but no exactly at the center of the Sun, $R = 0$).

If the relevant matter potential at production is well below the resonant value, $A_R = \Delta m^2 \cos 2\theta \gg A_0$, matter effects are negligible. With the characteristic matter density and energy of the solar neutrinos, this condition is fulfilled for values of Δm^2 such that $\Delta m^2/E \gg L_{\text{Sun-Earth}}$. So the propagation occurs as in vacuum with the oscillating phase averaged to 1/2 and the survival probability at the exposed surface of the Earth is

$$P_{ee}(\Delta m^2 \cos 2\theta \gg A_0) = 1 - \frac{1}{2}\sin^2 2\theta > \frac{1}{2}\,. \tag{78}$$

If the relevant matter potential at production is only slightly below the resonant value, $A_R = \Delta m^2 \cos 2\theta \gtrsim A_0$, the neutrino does not cross a region with resonant density, but matter effects are sizable enough to modify the mixing. The oscillating phase is averaged in the propagation between the Sun and the Earth. This regime is well described by an adiabatic propagation, Eq. (76)

$$P_{ee}(\Delta m^2 \cos 2\theta \geq A_0) = \frac{1}{2}\left[1 + \cos 2\theta_{m,0} \cos 2\theta\right] \ . \tag{79}$$

This expression reflects that an electron neutrino produced at A_0 is an admixture of ν_1 with fraction $P_{e1,0} = \cos^2\theta_{m,0}$ and ν_2 with fraction $P_{e2,0} = \sin^2\theta_{m,0}$. On exiting the Sun, ν_1 consists of ν_e with fraction $P_{1e} = \cos^2\theta$, and ν_2 consists of ν_e with fraction $P_{2e} = \sin^2\theta$ so $P_{ee} = P_{e1,0}P_{1e} + P_{e2,0}P_{2e} = \cos^2\theta_{m,0}\cos^2\theta + \sin^2\theta_{m,0}\sin^2\theta$ [37–39], exactly as given in Eq. (79). Since $A_0 < A_R$ the resonance is not crossed so $\cos 2\theta_{m,0}$ has the same sign as $\cos 2\theta$ and still $P_{ee} \geq 1/2$.

Finally, in the case that $A_R = \Delta m^2 \cos 2\theta < A_0$, the neutrino can cross the resonance on its way out. In the convention of $\Delta m^2 > 0$ this occurs if $\cos 2\theta > 0$ ($\theta < \pi/4$). which means that in vacuum ν_e is a combination of ν_1 and ν_2 with larger ν_1 component, while at the production point ν_e is a combination of ν_1^m and ν_2^m with larger ν_2^m component. In particular, if the density at the production point is much higher than the resonant density, $\Delta m^2 \cos 2\theta \ll A_0$,

$$\theta_{m,0} = \frac{\pi}{2} \quad \Rightarrow \quad \cos 2\theta_{m,0} = -1\,, \tag{80}$$

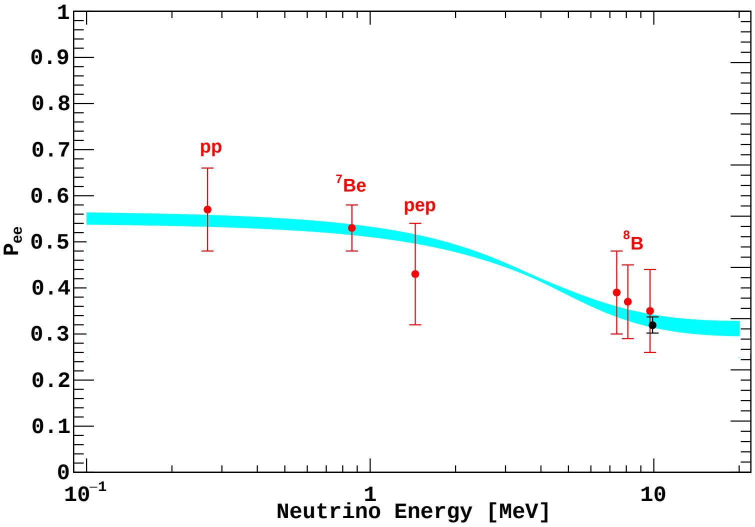


Fig. 4: Electron neutrino survival probability as function of neutrino energy. The points represent, from left to right, the Borexino pp, ^{7}Be, pep, and ^{8}B data (red points) and the SNO+SK ^{8}B data (black point). The three Borexino [40] ^{8}B data points correspond, from left to right, to the low-energy (LE) range, LE+HE range, and the high-energy (HE) range. The electron neutrino survival probabilities from experimental points are determined using a high metallicity SSM from Ref. [41]. The error bars represent the $\pm 1\sigma$ experimental + theoretical uncertainties. The curve corresponds to the $\pm 1\sigma$ prediction of the MSW-LMA solution using the parameter values given in Ref. [42]. This figure is taken from Ref. [13] and it was provided by A. Ianni.

and the produced ν_e is purely ν_2^m.

In this regime, the evolution of the neutrino ensemble can be adiabatic or non-adiabatic depending on the particular values of Δm^2 and the mixing angle. We now know that the neutrino masses and mixing happen to be such that the transition is adiabatic in all ranges of solar neutrino energies. Thus the survival probability at the exposed surface of the Earth is given by Eq. (79) but now with mixing angle, Eq. (80), so

$$P_{ee}(\Delta m^2 \cos 2\theta < A_0) = \frac{1}{2}\left[1 + \cos 2\theta_{m,0} \cos 2\theta\right] = \sin^2\theta \; . \tag{81}$$

So in this case P_{ee} can be much smaller than $1/2$ because $\cos 2\theta_{m,0}$ and $\cos 2\theta$ have opposite signs. This is referred to as the Mihheev-Smirnov-Wolfenstein (MSW) effect [31, 35] which plays a fundamental role in the interpretation of the solar neutrino data.

The resulting energy dependence of the survival probability of solar neutrinos is shown in Fig. 4 (together with a compilation of data from solar experiments). The plotted curve corresponds to $\Delta m^2 \sim 7.5 \times 10^{-5}$ eV^2 and $\sin^2\theta \sim 0.3$ (the so-called large mixing angle, LMA, solution). The figure illustrates the regimes described above. For these values of the oscillation parameters, neutrinos with $E \ll 1$ MeV are in the regime with $\Delta m^2 \cos 2\theta \gg A_0$ so the curve represents the value of vacuum averaged survival probability, Eq. (78), and therefore $P_{ee} > 0.5$. For $E > 10$ MeV, on the contrary, $\Delta m^2 \cos 2\theta \ll A_0$ and the survival probability is given by Eq. (81), so $P_{ee} = \sin^2\theta \sim 0.3$. In between, the survival probability is given by Eq. (79) with θ_0 changing rapidly from its vacuum value to the asymptotic matter value, Eq. (80), 90°.

2.3 Experimental evidence of neutrino oscillations

Neutrino flavour transitions have been searched for and observed in a variety of experiments using different neutrino sources and detection techniques. Generically the signatures can be classified in *disap-*

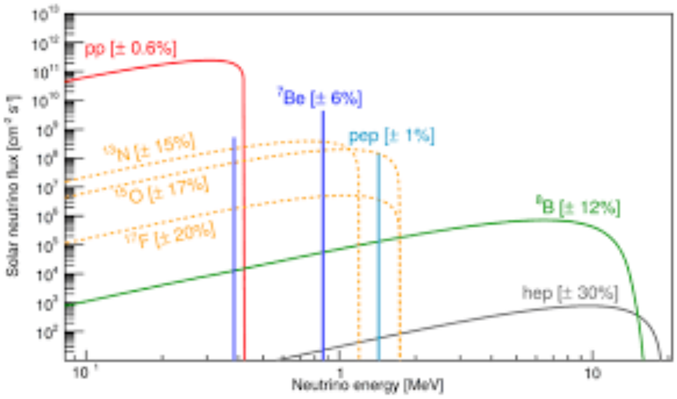


Fig. 5: Neutrino fluxes predicted by the SSM [41] as a function of the neutrino energy.

pearance signals, in which the number of observed neutrino events with the flavour of the original beam is below expectation, and *appearance* signals, in which neutrino events with different flavour than the expected in the beam are observed. Furthermore, to fully establish that the mechanism of flavour transition is that of mass-induced flavour oscillations and to best determine the corresponding mass difference and mixing angles, the experiments study the dependence of the event rates with the distance from the source or with the neutrino energy as well reconstructed as possible.

Solar neutrinos

Solar neutrinos are electron neutrinos produced in the thermonuclear reactions which generate the solar energy. These reactions occur via two main chains, the pp chain and the CNO cycle. There are five reactions which produce ν_e in the pp chain and three in the CNO cycle. Both chains result in the overall fusion of protons into ^{4}He:

$$4p \to {}^4\text{He} + 2e^+ + 2\nu_e + \gamma, \tag{82}$$

where the energy released in the reaction, $Q = 4m_p - m_{^4\text{He}} - 2m_e \simeq 26$ MeV, is mostly radiated through the photons and only a small fraction is carried by the neutrinos, $\langle E_{2\nu_e} \rangle = 0.59$ MeV.

In order to precisely determine the rates of the different reactions in the two chains which would give us the final neutrino fluxes and their energy spectrum, a detailed knowledge of the Sun and its evolution is needed. Solar Models (SSM) describe the properties of the Sun and its evolution after entering the main sequence. The models are based on a set of observational parameters and on several basic assumptions: spherical symmetry, hydrostatic and thermal equilibrium, equation of state of an ideal gas, and present surface abundances of elements similar to the primordial composition. I show in Fig. 5 the energy spectrum of the neutrino fluxes from the different reactions together with their present uncertainties as predicted by the SSM in Ref. [41] which is the last version of the Solar Model calculations initiated by Bahcall *et. al* [43]. It is customary to refer to the neutrino fluxes by the corresponding source reaction, so, for instance, the neutrinos produced from ^{8}B decay are called ^{8}B neutrinos.

Solar neutrinos were observed for the first time in 1968 in the Chlorine experiment located in the Homestake mine [44]. Since then they have been detected in a variety of experiments. They can generically be classified as:

- *Radiochemical* detectors, which detect solar $\nu_e's$ by capture in some inverse β decay reaction which leaves as signal the daughter nucleus which are recounted every certain period of time.
  - Chlorine in which ν_e's are captured via ^{37}Cl (ν, e^-) ^{37}Ar. The energy threshold for this reaction is 0.814 MeV, so the relevant fluxes are the ^{7}Be and ^{8}B neutrinos. For the SSM fluxes, 78% of the expected number of events are due to ^{8}B neutrinos while 13% arise from

  ^{7}Be neutrinos. The average ν_e event rate measured during the more than 20 years of operation was $\sim 30\%$ of that expected in the SSM [45].

  - Gallium experiments: SAGE [46] and GALLEX/GNO [47, 48]. In these experiments the solar neutrinos are captured via ^{71}Ga$(\nu, e^-)^{71}$Ge. The special properties of this target include a low threshold (0.233 MeV) and a strong transition to the ground level of ^{71}Ge, which gives a large cross section for the lower energy pp neutrinos. According to the SSM, approximately 54% of the events are due to pp neutrinos, while 26% and 11% arise from ^{7}Be and ^{8}B neutrinos, respectively. The average ν_e event rate measured in both experiments is $\sim 55\%$ of that expected in the SSM.

- *Real time* detectors in which the interaction of the solar neutrino is recorded in real time.
  - Water Cherenkov detectors: Kamiokande [49,49] and SuperKamiokande (SK) [50,51]. They are able to detect in real time the electrons which are emitted from the water by the elastic scattering (ES) of the solar neutrinos, $\nu_a + e^- \to \nu_a + e^-$. The detection threshold is above ~ 5 MeV. This means that these experiments are able to measure only the ^{8}B neutrinos (and the very small hep neutrino flux). They observe a rate of about $\sim 40\%$ of the SSM prediction. Notice that, while the detection process in radiochemical experiments is purely a CC (W-exchange) interaction, the detection ES process goes through both CC NC (Z-exchange) interactions. Consequently, the ES detection process is sensitive to all active neutrino flavors, although ν_e's (which are the only ones to scatter via W-exchange) give a contribution that is about 6 times larger than that of ν_μ's or ν_τ's.
  - SNO: The Sudbury Neutrino Observatory (SNO) is a Cherenkov detector using heavy water D$_2$O as target. Solar neutrinos can interact in the D$_2$O of via three different reactions. Electron neutrinos may interact via the CC reaction $\nu_e + d \to p + p + e^-$, and can be detected above an energy threshold of a few MeV. All active neutrinos ($\nu_a = \nu_e, \nu_\mu, \nu_\tau$) interact via the NC reaction $\nu_a + d \to n + p + \nu_a$ with an energy threshold of 2.225 MeV. The non-sterile neutrinos can also interact via ES, $\nu_a + e^- \to \nu_a + e^-$, but with smaller cross section. The comparison of the observed event rates in the different reactions allow to address the flavour dependence of the solar neutrinos arriving at the Earth. The reactions in the Sun only produce ν_e', however SNO observed rates which could only be understood if other flavours were present, confirming the flavour transition of solar ν_e'.

These real time experiments have provided us also with information on the time, direction and energy for each event. Signatures of neutrino oscillations might include distortion of the recoil electron energy spectrum, difference between the night-time solar neutrino flux and the day-time flux, or a seasonal variation in the neutrino flux. Observation of these effects were searched for and generically no significant energy or time dependence of the event rates beyond the expected ones in the SSM was observed.

With all the data collected in these experiments it was established that solar neutrinos undergo flavour transitions and they have to be due to the MSW effect in the Sun matter in the adiabatic regime, the so-called Large Mixing Angle (LMA) solution. In Fig. 6 I show the region of masses and mixing which better describe the bulk of solar neutrino data when interpreted in terms of mixing between 2ν states. As seen from the figure these results determine a non-zero $\Delta m^2 \sim \mathcal{O}(10^{-5})$ eV^2 and a mixing angle $\sim 32°$.

  - Borexino employs a liquid scintillator that produces sufficient light to observe low energy neutrino events via elastic scattering by electrons. The reaction is sensitive to all neutrino flavors by the neutral current interaction, but the cross section for ν_e is larger due to the combination of charged and neutral currents. It has a much lower threshold and better energy resolution than Cherenkov detectors which allows for detail determination of the observed spectrum rates and disentangling the different components once the oscillation parameters are known [40]. A compilation of their results is shown in Fig. 4.

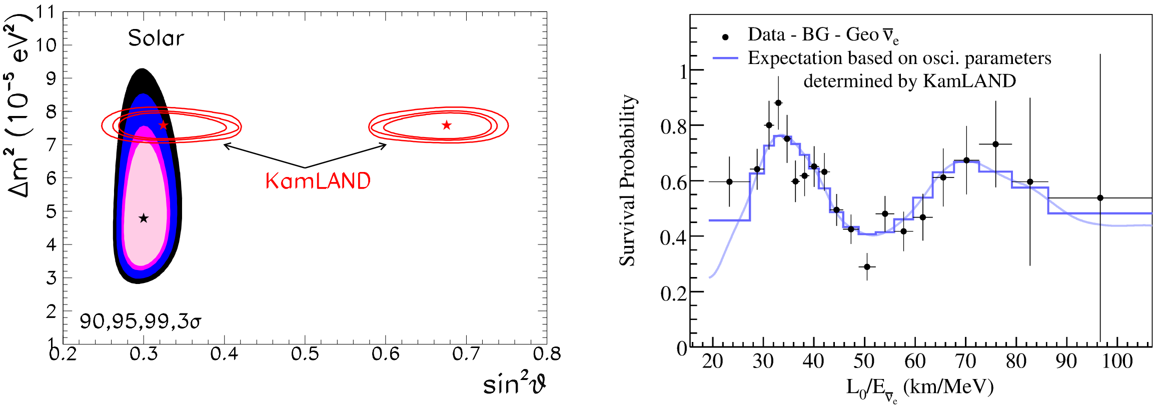


Fig. 6: **Left**: Allowed region of Δm^2 and $\sin^2\theta$ which better describe the bulk observation of solar data (full regions) and KamLAND spectral data (void regions) at different Confidence Levels (CL) as indicated in the figure when interpreted in terms of flavour oscillations driven by the mixing between 2ν states. **Right**: Ratio of the observed spectrum to the expectation for no-oscillation versus L_0/E for the KamLAND data. L_0 = 180 km is the flux-weighted average reactor baseline. The blue line corresponds to the expectation from oscillations of ν_e, taken from Ref. [52].

Reactor neutrinos at long baseline: KamLAND

Neutrino oscillations are also searched for using neutrino beams from nuclear reactors. Nuclear reactors produce $\bar{\nu}_e$ beams with $E_\nu \sim$ MeV. Due to the low energy, e^+'s are the only charged leptons which can be produced in the $\bar{\nu}_e$ CC interaction. If the $\bar{\nu}_e$ oscillated to another flavor, its CC interaction could not be observed. Therefore oscillation experiments performed at reactors are disappearance experiments. They have the advantage that small values of Δm^2 can be accessed due to the low beam energy. In particular values of Δm^2 as small as $\mathcal{O}(10^{-5})$ eV^2 can be accessed in a reactor experiment using a $\mathcal{O}(100)$ km baseline. Pursuing this idea, the KamLAND experiment, a 1000 ton liquid scintillation detector operated in the Kamioka mine in Japan which is located at an average distance of 150–210 km from several Japanese nuclear power stations. The measurement of the energy spectrum of the $\bar{\nu}_e$'s detected in KamLAND [52] is shown in the left panel of Fig. 6 and confirms $\bar{\nu}_e$ oscillations with parameters compatible with those observed in MSW flavour conversion of solar ν_e's. In the left panel of the same figure I show the parameters region obtained from the fit of KamLAND data in comparison with that from the analysis of solar neutrino data. The figure illustrates the compatibility of the observations. It also illustrates the degeneracy of solutions associated to θ and $\frac{\pi}{2} - \theta$ in 2ν oscillations in vacuum which is broken in the case of flavor transitions in matter as discussed in the previous sections.

Atmospheric neutrinos

Cosmic rays interacting with the nitrogen and oxygen in the Earth's atmosphere at an average height of 15 kilometers produce mostly pions and some kaons that decay into electron and muon neutrinos and anti-neutrinos.

Since ν_e is produced mainly from the decay chain $\pi \to \mu\nu_\mu$ followed by $\mu \to e\nu_\mu\nu_e$, one naively expects a $2:1$ ratio of ν_μ to ν_e. For higher energy events the expected ratio is larger because some of the muons arrive to Earth before they had time to decay. In practice, however, the theoretical calculation of the ratio of muon-like interactions to electron-like interactions in each experiment is more complicated. A set of increasingly more sophisticated calculations of the atmospheric fluxes

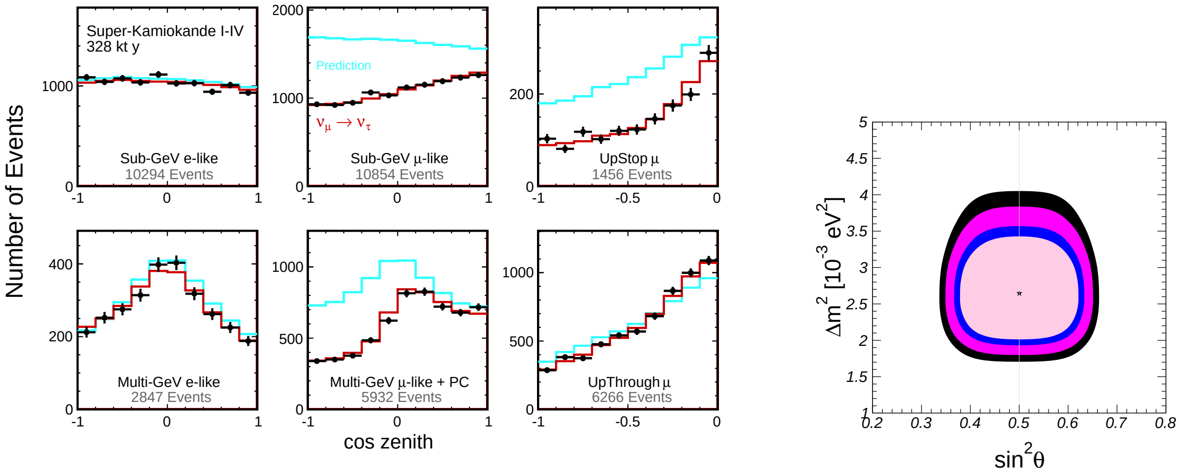


Fig. 7: Left: The zenith angle distribution of different event samples from SK experiment [13]. The points show the data, blue histograms show the non-oscillated expectations and the lines show the best-fit expectations for oscillations.**Right:** The allowed regions (same CL as Fig. 6) of Δm^2 and $\sin^2\theta$ by the global analysis of SK atmospheric data in the framework of $\nu_\mu \to \nu_\tau$ vacuum oscillations .

have been performed [53–56] over the years showing that the predicted absolute fluxes of neutrinos produced by cosmic-ray interactions in the atmosphere can vary at the 20% level among the different simulations while their zenith angular dependence, the ratio of neutrinos of different flavor, and the neutrino/antineutrino ratio are much more precisely determined.

Atmospheric neutrinos were first detected in the 1960's by the underground experiments in South Africa [57] and the Kolar Gold Field experiment in India [58]. A set of modern experiments were proposed and built starting the 1970's. The original purpose was to search for nucleon decay, for which atmospheric neutrinos constitute a background. But eventually the study of atmospheric neutrino events turned out to be a focus of study following a set of anomalies observed. This culminated with the first evidence of ν_μ oscillation presented by SK. in 1998 [59].

In Fig. 7 [13] I show the data accumulated in SK in its four phases of operation in different event categories and plotted as function of the zenith angle which defines the direction of the observed charged lepton produced in the interaction and which for energies above GeV is very well aligned with the neutrino direction. Upgoing stopping muons arise from neutrinos $E_\nu \sim 10$ GeV, and Upthrough-going muons are originated by neutrinos with energies of the order of hundreds of GeV. Comparing the observed and the expected distributions, we can make the following statements:

- ν_e distributions are well described by the expectations while ν_μ presents a deficit. Thus the atmospheric neutrino deficit is mainly due to disappearance of ν_μ and not the appearance of ν_e.
- The suppression of contained μ-like events is stronger for larger $\cos\theta$, which implies that the deficit grows with the distance traveled by the neutrino from its production point to the detector which ranges from $L \sim 10$ km for $\cos(\text{zenith}) = 1$ to $L \sim 10^4$ km for $\cos(\text{zenith}) = -1$. This effect is more obvious for multi-GeV events because at higher energy the direction of the charged lepton is more aligned with the direction of the neutrino.
- There is very little deficit on the number of through-going muons which implies that at larger energy the neutrino is less likely to disappear.

The simplest and most direct interpretation of the atmospheric neutrino anomaly is that of muon neutrino oscillations $\nu_\mu \to \nu_\tau$ with parameters as shown in the right of Fig. 7 As seen from the figure these results

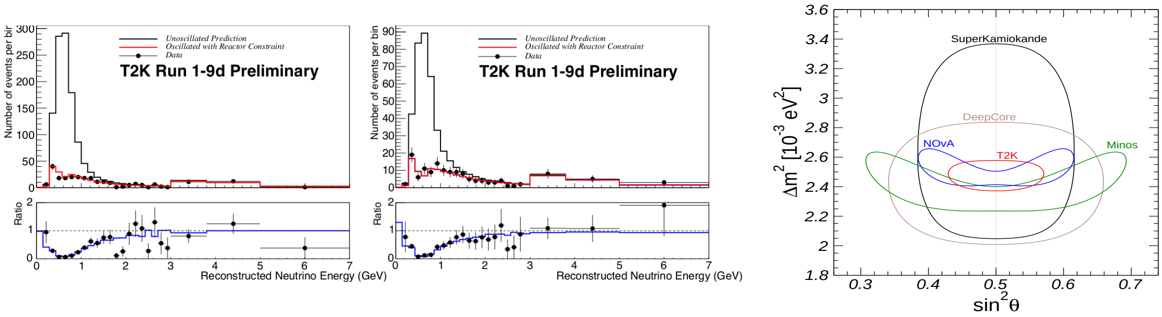


Fig. 8: Spectrum of ν_μ (left) and $\bar{\nu}_\mu$ (center) events observed in T2K. Data points with statistical error bars are shown together with the prediction without (black line) and including (red line) neutrino oscillation. Figure from Ref. [13]. The right panel shows the allowed regions at 95% CL from the analysis of the data in terms of ν_μ disappearance due to oscillations in the 2ν approximation. For comparison the corresponding regions obtained from the analysis of atmospheric neutrino experiments SK and ICECUBE are also shown.

.

determine a non-zero $\Delta m^2 \sim \mathcal{O}(10^{-3})$ eV^2 and a mixing angle $\sim 45°$.

The neutrino telescopes primarily built for the high energy neutrino astronomy such as ANTARES and IceCube can also measure neutrino oscillations with atmospheric neutrinos. IceCube DeepCore [60] provided a precision comparable to the measurements by Super-Kamiokande.

Accelerator neutrinos at long baselines

Conventional neutrino beams from accelerators are mostly produced by π decays (and some K decays), with the pions produced by the scattering of the accelerated protons on a fixed target:

$$\begin{aligned} p + \text{target} \to \quad & \pi^\pm + X \\ & \pi^\pm \to \quad \mu^\pm + \nu_\mu(\bar{\nu}_\mu) \\ & \qquad\qquad \mu^\pm \to e^\pm + \nu_e(\bar{\nu}_e) + \bar{\nu}_\mu(\nu_\mu) \end{aligned} \tag{83}$$

Thus the beam can contain both μ- and e-neutrinos and antineutrinos. The final composition and energy spectrum of the neutrino beam is determined by selecting the sign of the decaying π and by stopping the produced μ in the beam line. There is an additional contribution to the electron neutrino and antineutrino flux from kaon decay.

Indeed the accelerator neutrino beams are very similar in nature to the atmospheric neutrinos and they can be used to test the observed oscillation signal with a controlled beam. Given the characteristic Δm^2 involved in the interpretation of the atmospheric neutrino signal, the intense neutrino beam from the accelerator must be aimed at a detector located underground at a distance of several hundred kilometers.

The first LBL accelerator experiment was the K2K experiment [61] which run with a baseline of about 235 km from KEK to SK. The MINOS experiment used a beam from Fermilab and a detector in Soudan mine 735 km away [62]. The results from both K2K and MINOS both in the observed deficit of events and in their energy dependence confirmed that accelerator ν_μ oscillate over distances of several hundred kilometers as expected from oscillations with the parameters compatible with those inferred from the atmospheric neutrino data.

In the last decade a second generation of LBL experiments came to operation with the aim at precise determination of the ν_μ disappearance, looking for ν_e appearance and testing the possibility of CP violation. T2K uses the high-intensity beam from the new constructed proton synchrotron J-PARC and the Super-Kamiokande detector at 295 km. The NOvA experiment uses the NuMI beamline with an off-axis configuration. The far detector is located in Minnesota, at 810 km from the source.

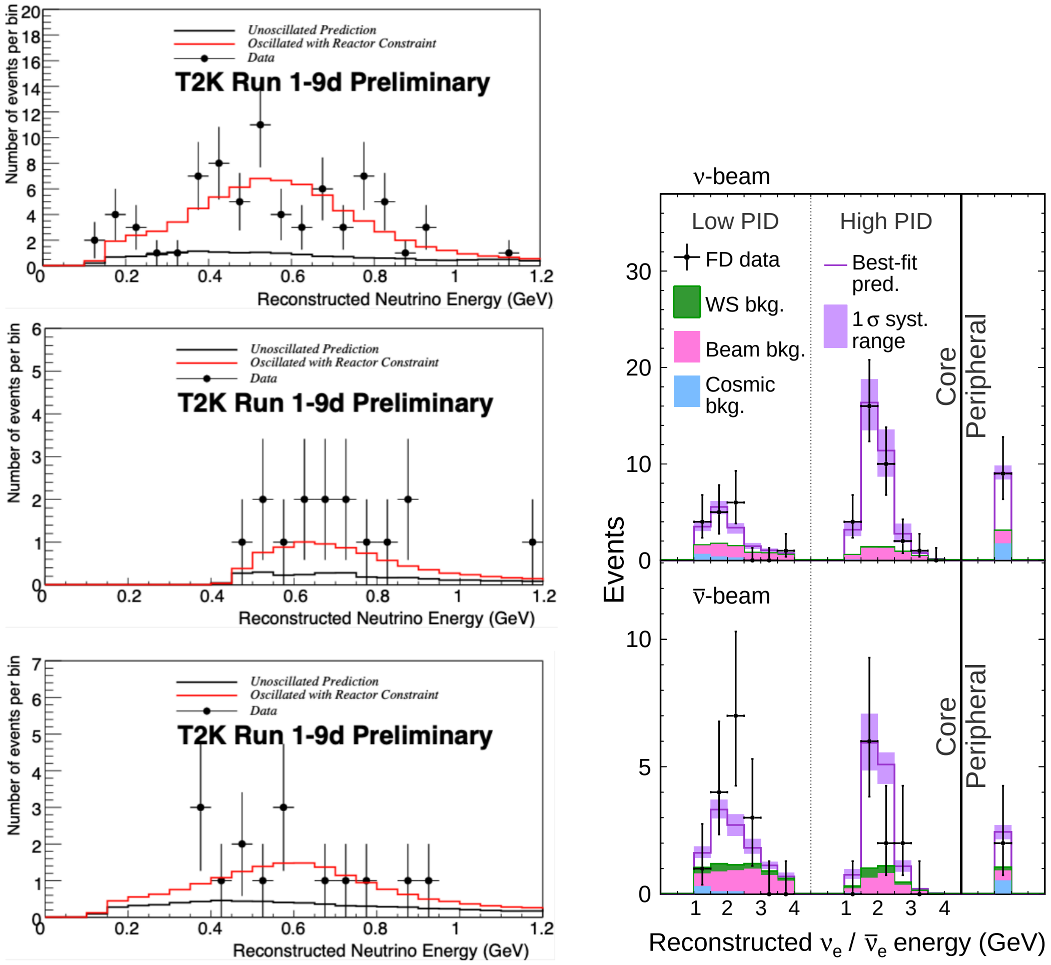


Fig. 9: Spectrum of ν_e and $\bar\nu_e$ events observed in T2K [13] (left panels). and NOvA [63].

Both experiments have taken data with ν and with $\bar\nu$ beam. Their measured spectrum of μ events allow for precise determination of the same oscillation parameters measured with atmospheric neutrinos. We show in Fig. 8 the observed spectrum of ν_μ and $\bar\nu_\mu$ events in T2K together with the allowed regions at 95% CL from the analysis of the data from the different LBL experiments in terms of $\bar\nu_\mu$ disappearance due to oscillations in the 2ν approximation compared to those from atmospheric neutrino experiments SK and ICECUBE.

Both experiments have also observed $\nu_\mu \to \nu_e$ and $\bar\nu_\mu \to \bar\nu_e$ transitions. In Fig. 9 I show the spectrum of ν_e and $\bar\nu_e$ events in both experiments. If due to oscillations, these results could be explained with a $\Delta m^2 \sim \mathcal{O}(10^{-3})$ eV^2 is compatible with that inferred from the analysis of $\nu_\mu \to \nu_\tau$ in atmospheric and LBL neutrinos but with a much smaller mixing angle. Also comparison of the observations in neutrino and antineutrino mode allow for test of CP symmetry. The present situation is that T2K claims a CP violation effect. NOvA indication of leptonic CP violation is less conclusive.

Reactor neutrinos at $\mathcal{O}$(km) baseline

Over several decades neutrino oscillations were also searched with $\bar\nu_e$ fluxes produced by reactors but at baselines of order of kilometer or shorter. Originally they all reported negative results when compared with the expected reactor fluxes obtained with the best calculations of the time. The strongest bounds were established by CHOOZ [64] and Palo Verde [65]. which searched for neutrino oscillations in the $\Delta m^2 \sim 10^{-2}$–10^{-3} eV^2 range and set a limit on the corresponding mixing angle $\sin^2\theta \lesssim 0.025$ at 90%

CL.

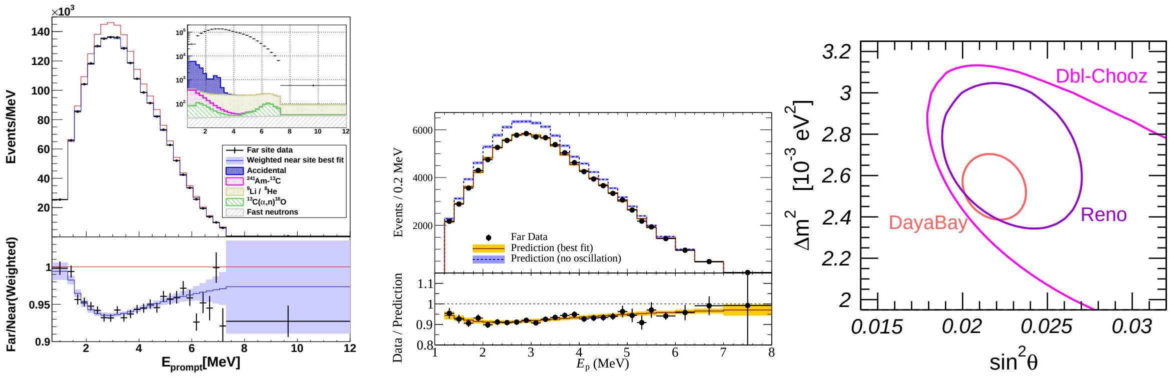


Fig. 10: Energy spectra for prompt events at the far detectors for Daya Bay [66] (left) and RENO [67] (center). The right panel show the allowed regions at 95% CL from the analysis of the data in terms of $\bar{\nu}_e$ disappearance due to oscillations in the 2ν approximation.

.

This changed over the last decade with three experiments, Double Chooz [68] in France, Daya Bay [66], in China, and RENO [67] in Korea, which to achieve better precision made use of at least two detectors – one near the reactor and other at kilometer distance – allowing to minimize systematics and flux calculation uncertainties. All three report a deficit of events in the far detectors compared with expectation from the observation in the near detector in the absence of oscillations. Furthermore they all measure a distortion of the observed spectrum in the far detectors consistent with oscillations. We show in Fig. 10 the spectrum of events observed in the far detectors in Daya Bay (left) and RENO (center). In the right panel I show the allowed regions at 95% CL from the analysis of this data in terms of $\bar{\nu}_e$ disappearance due to oscillations in the 2ν approximation. As see the $\Delta m^2 \sim \mathcal{O}(10^{-3})$ eV^2 is compatible with that inferred from the analysis of $\nu_\mu \to \nu_\tau$ in atmospheric and LBL neutrinos. But the mixing angle $\sim 9°$ is different, and also, unlike in atmospheric and LBL ν_μ disappearance, ν_e's are involved.

2.4 Summary

Neutrino masses and mixing imply flavour oscillation in vacuum and flavour transitions in matter with a well determined dependence on the distance from the source and the energy of the neutrino. Presently these phenomena have been observed in a variety of experiments. In brief:

- Atmospheric ν_μ and $\bar{\nu}_\mu$ disappear most likely converting to ν_τ and $\bar{\nu}_\tau$. The results show an energy and distance dependence perfectly described by mass-induced oscillations.
- Accelerator ν_μ and $\bar{\nu}_\mu$ disappear over distances of $\sim$ 200 to 800 km. The energy spectrum of the results show a clear oscillatory behaviour also in accordance with mass-induced oscillations with wavelength in agreement with the effect observed in atmospheric neutrinos.
- Accelerator ν_μ and $\bar{\nu}_\mu$ appear as ν_e and $\bar{\nu}_e$ at distances $\sim$ 200 to 800 km.
- Solar ν_e convert to ν_μ and/or ν_τ. The observed energy dependence of the effect is well described by massive neutrino conversion in the Sun matter according to the MSW effect.
- Reactor $\bar{\nu}_e$ disappear over distances of $\sim$ 200 km and $\sim$ 1.5 km with different probabilities. The observed energy spectra show two different mass-induced oscillation wavelengths: at short distances in agreement with the one observed in accelerator ν_μ disappearance, and at long distance compatible with the required parameters for MSW conversion in the Sun.

3 LECTURE III: Implications

3.1 The new minimal Standard Model

From the experimental situation described in the second lecture we conclude that the description of all the data requires an effective model consisting of the SM minimally extended to include neutrino masses with mixing between the three flavour neutrinos of the SM in three distinct mass eigenstates. As mentioned in the first lecture this can be effectively achieved in two different ways:

- Introduce ν_R and impose L conservation so after spontaneous electroweak symmetry breaking

$$\mathcal{L}_D = \mathcal{L}_{SM} - M_\nu \bar{\nu}_L \nu_R + h.c. \tag{84}$$

In this case mass eigenstate neutrinos are Dirac fermions, ie $\nu^C \neq \nu$.

- Construct a mass term only with the SM left-handed neutrinos by allowing L violation

$$\mathcal{L}_M = \mathcal{L}_{SM} - \frac{1}{2} M_\nu \bar{\nu}_L \nu_L^c + h.c. \tag{85}$$

In this case the mass eigenstates are Majorana fermions.

In either case U is a 3×3 matrix but which for Majorana (Dirac) neutrinos depends on six (four) independent parameters: three mixing angles and three (one) phases

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \cdot \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{\rm CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{\rm CP}} & 0 & c_{13} \end{pmatrix} \cdot \begin{pmatrix} c_{21} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} e^{i\eta_1} & 0 & 0 \\ 0 & e^{i\eta_2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{86}$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$. In addition to the Dirac-type phase $\delta_{\rm CP}$, analogous to that of the quark sector, there are two physical phases η_i associated to the Majorana character of neutrinos.

There are several possible conventions for the ranges of the angles and ordering of the states. The community finally agreed to a parametrization of the leptonic mixing matrix as in Eq. (86). The angles θ_{ij} can be taken without loss of generality to lie in the first quadrant, $\theta_{ij} \in [0, \pi/2]$, and the phase $\delta_{\rm CP} \in [0, 2\pi]$. Values of $\delta_{\rm CP}$ different from 0 and π imply CP violation in neutrino oscillations in vacuum [69–71]. The Majorana phases η_1 and η_2 play no role in neutrino oscillations [70, 72].

In this convention there are two non-equivalent orderings for the spectrum of neutrino masses:

- Spectrum with Normal Ordering (NO) with $m_1 < m_2 < m_3 \Rightarrow \Delta m^2_{31,32} > 0$.
- Spectrum Inverted ordering (IO) with $m_3 < m_1 < m_2 \Rightarrow \Delta m^2_{31,32} < 0$.

Furthermore the data show a hierarchy between the mass splittings, $\Delta m^2_{21} \ll |\Delta m^2_{31}| \simeq |\Delta m^2_{32}|$. So in total, the 3-$\nu$ oscillation analysis of the existing data involves six parameters: 2 mass differences (one of which can be positive or negative), 3 mixing angles, and the CP phase. I summarize in Table 2 the different experiments which dominantly contribute to the present determination of the different parameters in the chosen convention. The table illustrates that the determination of the leptonic parameters requires global analysis of the data from the different experiments. Over the years these analyses have been in the hands of a few phenomenological groups (see for example Refs. [73–76]). In Fig. 11 I show the determination of the six parameters from the updated analysis in Ref. [73]. Defining the 1σ relative precision of the parameter by $2(x^{\rm up} - x^{\rm low})/3(x^{\rm up} + x^{\rm low})$, where $x^{\rm up}$ ($x^{\rm low}$) is the upper (lower) bound on a parameter x at the 3σ level, one reads the following 1σ relative precision (marginalizing over ordering) for the better determined parameters:

$$4\% \, (\sin^2\theta_{12}) \, , \quad 2.3\% \, (\sin^2\theta_{13}) \, , \quad 16\% \, (\Delta m^2_{21}) \, . \quad 1.3\% \, (|\Delta m^2_{3\ell}|) \tag{87}$$

The issues which still require clarification are: the mass ordering discrimination, the determination of θ_{23} and the leptonic CP phase $\delta_{\rm CP}$:

Table 2: Experiments contributing to the present determination of the oscillation parameters.

Experiment	Dominant	Important
Solar Experiments	θ_{12}	Δm^2_{21} , θ_{13}
Reactor LBL (KamLAND)	Δm^2_{21}	θ_{12} , θ_{13}
Reactor MBL (Daya-Bay, Reno, D-Chooz)	θ_{13}, $\lvert\Delta m^2_{31,32}\rvert$	
Atmospheric Experiments (SK, IC-DC)		θ_{23},$\lvert\Delta m^2_{31,32}\rvert$, θ_{13},$\delta_{\rm CP}$
Accel LBL ν_μ,$\bar\nu_\mu$, Disapp (K2K, MINOS, T2K, NOνA)	$\lvert\Delta m^2_{31,32}\rvert$, θ_{23}	
Accel LBL ν_e,$\bar\nu_e$ App (MINOS, T2K, NOνA)	$\delta_{\rm CP}$	θ_{13} , θ_{23}

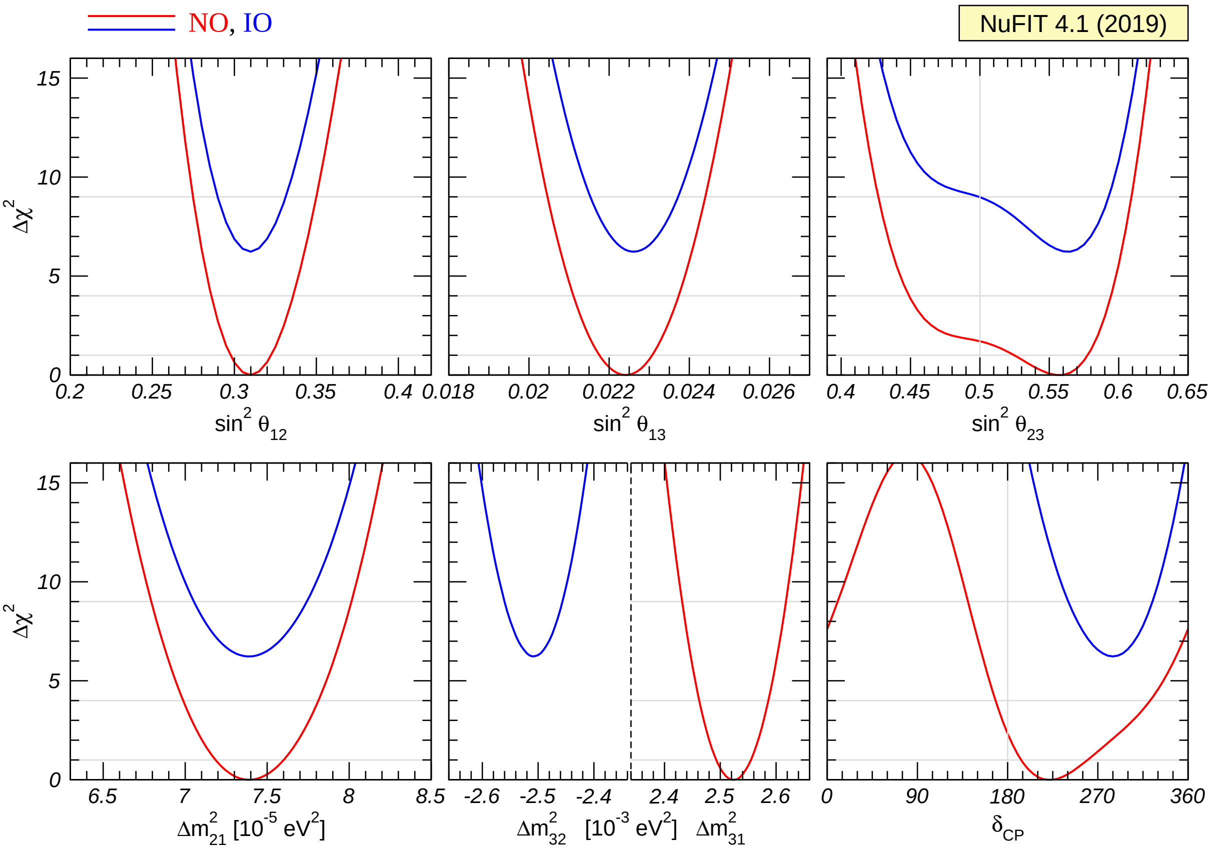


Fig. 11: Global 3ν oscillation analysis. The red (blue) curves are for Normal (Inverted) Ordering. Results for different assumptions concerning the analysis of data from reactor experiments are shown as explained in the text.

- The best fit is for the normal mass ordering. Inverted ordering is disfavoured with a $\Delta\chi^2$ which ranges from slightly above 2σ – driven by the interplay of long-baseline accelerator and short-baseline reactor data – to 3σ when adding the atmospheric χ^2 (not shown in the figure) from Ref. [77].
- The analysis find some preference for the second octant of θ_{23} but with statistical significance still well below 3σ.
- The best fit for the complex phase in NO is at $\delta_{\rm CP} \sim 120°$ but CP conservation (for $\delta_{\rm CP} \sim 180°$) is still allowed at a confidence level (CL) of 1-2σ. We notice that, at present, the significance of CP violation in the global analysis is reduced with respect to that reported by T2K [78] because NOvA data does not show a significant indication of CP violation.

These results yield the present determination of the modulus of the leptonic mixing matrix

$$|U|_{3\sigma} = \begin{pmatrix} 0.797 \to 0.842 & 0.518 \to 0.585 & 0.143 \to 0.156 \\ 0.233 \to 0.495 & 0.448 \to 0.679 & 0.639 \to 0.783 \\ 0.287 \to 0.532 & 0.486 \to 0.706 & 0.604 \to 0.754 \end{pmatrix} , \tag{88}$$

which is still much less precisely known than the corresponding quark CKM mixing matrix [14]

$$|V|_{\text{CKM}} = \begin{pmatrix} 0.97427 \pm 0.00015 & 0.22534 \pm 0.0065 & (3.51 \pm 0.15) \times 10^{-3} \\ 0.2252 \pm 0.00065 & 0.97344 \pm 0.00016 & (41.2^{+1.1}_{-5}) \times 10^{-3} \\ (8.67^{+0.29}_{-0.31}) \times 10^{-3} & (40.4^{+1.1}_{-0.5}) \times 10^{-3} & 0.999146^{+0.000021}_{-0.000046} \end{pmatrix} . \tag{89}$$

It is also clear by comparing them that they are very different in structure. Quark CKM matrix is rather *hierarchical* with mixing angles relatively small and smaller for the heavier generation. On the contrary two leptonic mixings are large and even the smaller one, $\theta_{13} \sim 9^\circ$, is not very small.

In the framework of 3ν mixing leptonic CP violation can be quantified in terms of a unique leptonic Jarlskog invariant [79], defined by:

$$\begin{aligned} J_{\text{CP}} &\equiv \text{Im}\left[U_{\alpha i} U^*_{\alpha j} U^*_{\beta i} U_{\beta j}\right] \\ &\equiv J^{\max}_{\text{CP}} \sin\delta_{\text{CP}} = \cos\theta_{12}\sin\theta_{12}\cos\theta_{23}\sin\theta_{23}\cos^2\theta_{13}\sin\theta_{13}\sin\delta_{\text{CP}} \; . \end{aligned} \tag{90}$$

For example from the analysis in Refs. [73, 74]

$$J^{\max}_{\text{CP}} = 0.03359 \pm 0.0006\,(\pm 0.0019)\,, \tag{91}$$

at 1σ (3σ) for both orderings, and the preference of the present data for non-zero δ_{CP} implies a non-zero best fit value $J^{\text{best}}_{\text{CP}} = -0.019$. This can be directly compared with the value of the corresponding invariant in the quark sector $J^{\text{quarks}}_{\text{CP}} = (3.18 \pm 0.15) \times 10^{-5}$ [14].

The status of the determination of leptonic CP violation can also be graphically displayed by projecting the results of the global analysis in terms of leptonic unitarity triangles [80–82]. Since in the analysis U is unitary by construction, any given pair of rows or columns can be used to define a triangle in the complex plane. There a total of six possible triangles corresponding to the unitary conditions

$$\sum_{i=1,2,3} U_{\alpha i} U^*_{\beta i} = 0 \text{ with } \alpha \neq \beta\,, \qquad \sum_{\alpha=e,\mu,\tau} U_{\alpha i} U^*_{\alpha j} = 0 \text{ with } i \neq j\,. \tag{92}$$

As illustration we show in Fig. 12 the recasting of the allowed regions of the analysis in Refs. [73, 74] in terms of one leptonic unitarity triangle. We show the triangle corresponding to the unitarity conditions on the first and third columns (after the shown rescaling) which is the equivalent to the one usually shown for the quark sector. In this figure the absence of CP violation would imply a flat triangle, *i.e.*, $\text{Im}(z) = 0$. So the CL at which leptonic CP violation is being observed would be given by the CL at which the region crosses the horizontal axis. For comparison we show in the right panel the present determination of the corresponding unitary triangle in the quark sector as given in Ref. [14]. Notice that the tiny yellow region in the apex of the triangle in the quark sector is the equivalent to the whole blue region in the leptonic sector.

Projections on neutrino mass scale observables

As discussed in the first lecture, information on the neutrino masses, rather than mass differences, can be extracted from kinematic studies of reactions in which a neutrino or an anti-neutrino is involved. In

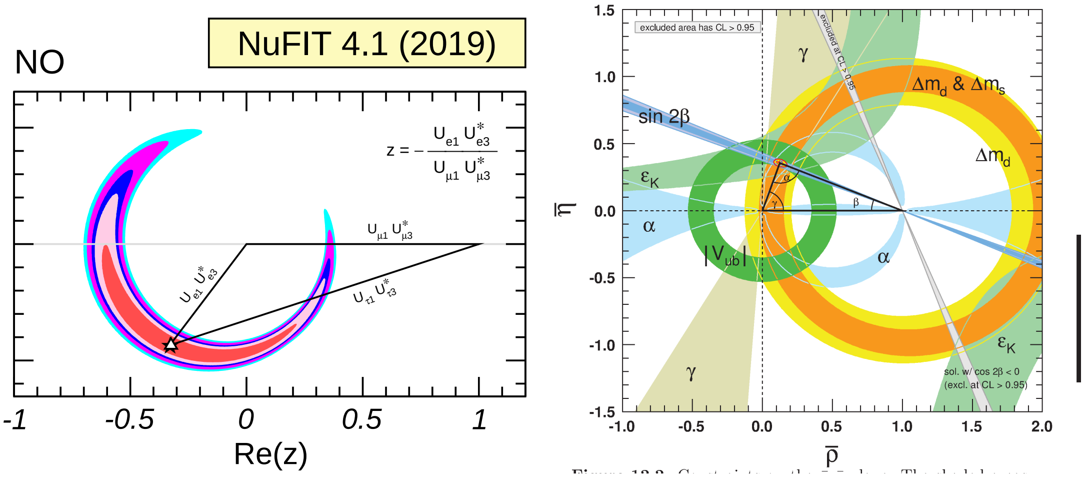


Fig. 12: Left: Leptonic unitarity triangle for the first and third columns of the mixing matrix. After scaling and rotating the triangle so that two of its vertices always coincide with $(0,0)$ and $(1,0)$ the figure shows the 1σ, 90%, 2σ, 99%, 3σ CL (2 dof) allowed regions of the third vertex for the NO from the analysis in Refs. [73, 74]. **Right:** The corresponding determination of the unitary triangle in the quark sector.

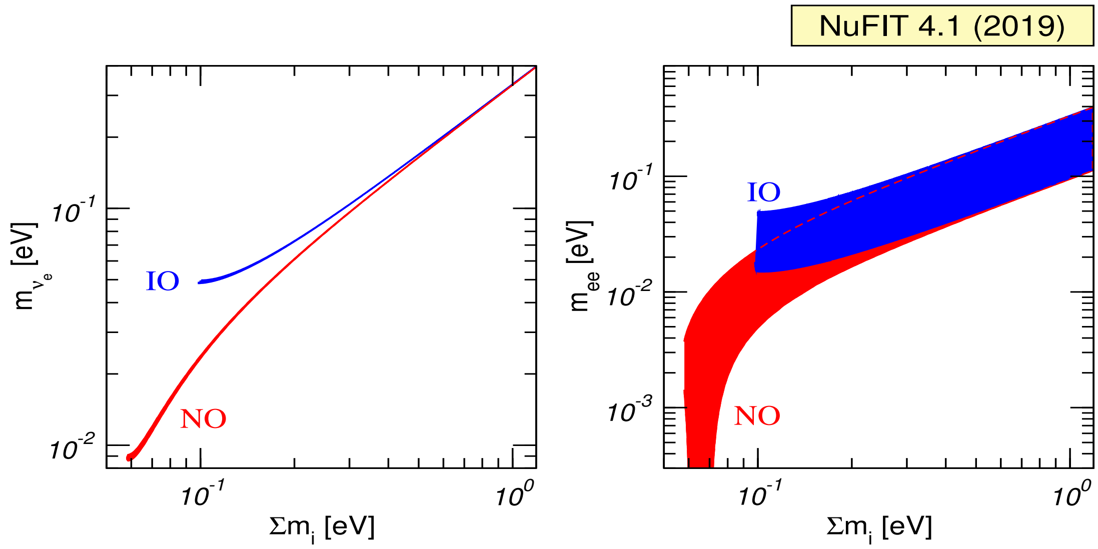


Fig. 13: 95% allowed regions (for 2 dof) in the planes $(m_{\nu_e}^{\rm eff}, \sum m_\nu)$ and $(m_{ee}, \sum m_\nu)$ obtained from projecting the results of the global analysis of oscillation data. The regions are defined with respect to the minimum for each ordering.

the presence of mixing the most relevant constraint comes from the study of the end point of the electron spectrum in Tritium beta decay and for 3ν mixing the $m_{\nu_e}^{\rm eff}$ introduced in Eq. (33) reads:

$$
\begin{aligned}
m_{\nu_e}^{\rm eff} &= \frac{\sum_i m_i^2 |U_{ei}|^2}{\sum_i |U_{ei}|^2} = \sum_i m_i^2 |U_{ei}|^2 = c_{13}^2 c_{12}^2 m_1^2 + c_{13}^2 s_{12}^2 m_2^2 + s_{13}^2 m_3^2 \\
&= \begin{cases} \text{NO:} \quad m_0^2 + \Delta m_{21}^2 c_{13}^2 s_{12}^2 + \Delta m_{3\ell}^2 s_{13}^2 \,, \\ \text{IO:} \quad m_0^2 - \Delta m_{21}^2 c_{13}^2 c_{12}^2 - \Delta m_{3\ell}^2 c_{13}^2 \end{cases}
\end{aligned}
\tag{93}
$$

where the second equality holds if unitarity is assumed and $m_0 = m_1\,(m_3)$ in NO (IO) denotes the lightest neutrino mass.

In what respects the effective Majorana mass of the ν_e which determines the rate of the rate of $0\nu\beta\beta$ decay in the 3ν scenario reads:

$$\begin{aligned} m_{ee} &= \left|\sum_i m_i U_{ei}^2\right| = \left|m_1 c_{13}^2 c_{12}^2 e^{i2\alpha_1} + m_2 c_{13}^2 s_{12}^2 e^{i2\alpha_2} + m_3 s_{13}^2 e^{-i2\delta_{\rm CP}}\right| \\ &= \begin{cases} \text{NO:} & m_0 \left|c_{13}^2 c_{12}^2 e^{i2(\alpha_1-\delta_{\rm CP})} + \sqrt{1+\frac{\Delta m_{21}^2}{m_0^2}}\, c_{13}^2 s_{12}^2 e^{i2(\alpha_2-\delta_{\rm CP})} + \sqrt{1+\frac{\Delta m_{3\ell}^2}{m_0^2}}\, s_{13}^2\right| \\ \text{IO:} & m_0 \left|\sqrt{1-\frac{\Delta m_{3\ell}^2+\Delta m_{21}^2}{m_0^2}}\, c_{13}^2 c_{12}^2 e^{i2(\alpha_1-\delta_{\rm CP})} + \sqrt{1-\frac{\Delta m_{3\ell}^2}{m_0^2}}\, c_{13}^2 s_{12}^2 e^{i2(\alpha_2-\delta_{\rm CP})} + s_{13}^2\right| \end{cases} \end{aligned} \tag{94}$$

which, unlike Eq. (93), depends also on the CP violating phases. Finally, as discussed in the first lecture, neutrino masses have also interesting cosmological effects and cosmological data mostly give information on the sum of the neutrino masses, $\sum_i m_i$, while they have very little to say on their mixing structure and on the ordering of the mass states.

Correlated information on these three probes of the neutrino mass scale can be obtained by mapping the results from the global analysis of oscillations presented previously and from the expressions above one finds that the correlations are different for NO and IO. We show in Fig. 13 the present status of this exercise. Also, the relatively large width of the regions in the right panel are due to the unknown Majorana phases. Thus, in principle, from a positive determination of two of these probes, information can be obtained on the the mass ordering [83, 84] and on the value the Majorana phases.

3.2 Beyond the 3ν paradigm: Light sterile neutrinos

Besides the huge success of three-flavour oscillations described above, there are some anomalies which cannot be explained within the 3ν framework and which might point towards the existence of additional neutrino states with masses at the eV scale. In brief:

- the LSND experiment [85] reported evidence for $\bar\nu_\mu \to \bar\nu_e$ transitions with $E/L \sim 1\ \text{eV}^2$, where E and L are the neutrino energy and the distance between source and detector.
- this effect has also been searched for by the MiniBooNE experiment [86], which reports a yet unexplained event excess in the low-energy region of the electron neutrino and anti-neutrino event spectra. No significant excess is found at higher neutrino energies. Interpreting the data in terms of oscillations, parameter values consistent with the ones from LSND are obtained, but the test is not definitive;
- radioactive source experiments at the Gallium solar neutrino experiments both in SAGE and GALLEX/GNO have obtained an event rate which is somewhat lower than expected. If not due to uncertainties in the interaction cross section, this effect can be explained by the hypothesis of ν_e disappearance due to oscillations with $\Delta m^2 \gtrsim 1\ \text{eV}^2$ ("Gallium anomaly") [87, 88];
- new calculations of the neutrino flux emitted by nuclear reactors [89, 90] predict a neutrino rate which is a few percent higher than observed in short-baseline ($L \lesssim 100$ m) reactor experiments. If not due to systematic or theoretical uncertainties, a decrease rate at those distances can be explained by assuming $\bar\nu_e$ disappearance due to oscillations with $\Delta m^2 \sim 1\ \text{eV}^2$ ("reactor anomaly") [91]. This reactor anomaly is under study both by the experimental community – with a set of follow-up measurements performed at SBL both at reactors and accelerators – , and by the theory community for improvements of the reactor flux calculations.

As mentioned in the first lecture, whatever the extension of the SM we want to consider it must contain only three light active neutrinos. Therefore if we need more than three light massive states we must add sterile neutrinos to the particle content of the model.

The most immediate question as these anomalies were reported was whether they could all be consistently described in combination with the rest of the neutrino data – in particular with the negative

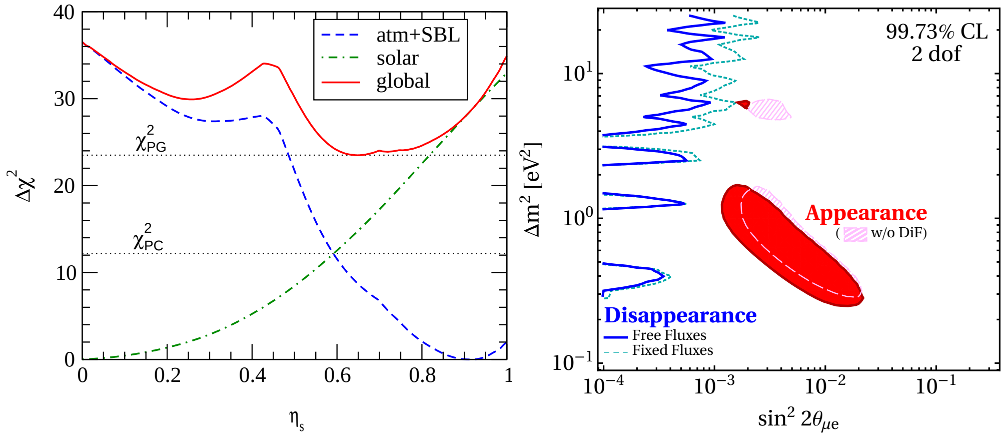


Fig. 14: *Left*: Status of the 2+2 oscillation scenarios from Ref. [93] ($\eta_S = \sum_i |U_{is}|^2$ where i runs over the two massive states mostly relevant for solar neutrino oscillations). In the figure also shown are the values of $\chi^2_{\rm PC}$ and $\chi^2_{\rm PG}$ relevant for parameter consistency test and parameter goodness of fit respectively. *Right*: Present status of 3+1 oscillation scenarios from Ref. [94].

results on disappearance of ν_μ at short distances – if one adds those additional sterile states. Quantitatively one can start by adding a fourth massive neutrino state to the spectrum, and perform a global data analysis to answer this question. Although the answer is always the same the physical reason behind it depends on ordering assumed for the states. In brief, there are six possible four-neutrino schemes which can in principle accommodate the results of solar+KamLAND and atmospheric+LBL neutrino experiments as well as the SBL result. They can be divided in two classes: (2+2) and (3+1). In the (3+1) schemes, there is a group of three close-by neutrino masses (as on the 3ν schemes described in the previous section) that is separated from the fourth one by a gap of the order of 1 eV, which is responsible for the SBL oscillations. In (2+2) schemes, there are two pairs of close masses (one pair responsible for solar results and the other for atmospheric [92]) separated by the $\mathcal{O}(\text{eV})$ gap. The main difference between these two classes is the following: if a (2+2)-spectrum is realized in nature, the transition into the sterile neutrino is a solution of either the solar or the atmospheric neutrino problem, or the sterile neutrino takes part in both. Consequently a (2+2)-spectrum is easier to test because the required mixing of sterile neutrinos in either solar and/or atmospheric oscillations would modify their effective matter potential in the Sun and in the Earth and giving distinctive effects in the solar and/or atmospheric neutrino observables. Those distinctive effects were not observed so oscillations into sterile neutrinos did not describe well either solar or atmospheric data. Consequently as soon as the early 2000's 2+2 spectra could be ruled out already beyond 3-4 σ as seen in the left panel in Fig.14 taken from Ref. [93]. On the contrary, for a (3+1)-spectrum (and more generally for a $3+N$-spectrum with an arbitrary N number of sterile states), the sterile neutrino(s) could be only slightly mixed with the active ones and mainly provide a description of the SBL results. In this case the oscillation probabilities for experiments working at $E/L \sim 1$ eV^2 take a simple form:

$$P_{\alpha\alpha} = 1 - \sin^2 2\theta_{\alpha\alpha} \sin^2 \Delta\,, \qquad P_{\mu e} = \sin^2 2\theta_{\mu e} \sin^2 \Delta\,, \tag{95}$$

where $\Delta \equiv \Delta m^2_{41} L/4E$ and one can define effective mixing angles

$$\sin^2 2\theta_{\alpha\alpha} \equiv 4|U_{\alpha 4}|^2(1 - |U_{\alpha 4}|^2)\,, \qquad \sin^2 2\theta_{\mu e} \equiv 4|U_{\mu 4}|^2 |U_{e4}|^2\,. \tag{96}$$

In here $\alpha = e, \mu$ and $U_{\alpha 4}$ are the elements of the lepton mixing matrix describing the mixing of the 4th neutrino mass state with the electron and muon flavour. In this scenario there is no sensitivity to CP violation in the the Δ driven oscillations, so the relations above are valid for both neutrinos and antineutrinos. At linear order in the mixing elements one can derive a relation between the amplitudes of appearance and disappearance probabilities:

$$4\sin^2 2\theta_{\mu e} \approx \sin^2 2\theta_{ee} \sin^2 2\theta_{\mu\mu} \,. \tag{97}$$

This relation implies a constraint between the possible results in disappearance and appearance experiments. Consequently it is not trivial to find a consistent description to all the SBL anomalies. Over the years, different groups have performed a variety of such global analysis leading to quantitative different conclusions on the statistical quality of the global fit (see for example Refs. [94–99], see also Refs. [100, 101] for recent reviews on the subject). Generically the results of the global analysis show that there is significant tension between groups of different data sets – in particular between appearance and disappearance results – and Eq. (97) makes it difficult to obtain a good global fit as illustrated in the right panel in Fig.14 taken from Ref. [94] which concluded that 3+1 scenario is excluded at 4.7σ level.

A straightforward question to ask is whether the situation improves if more neutrino states at the eV scale are introduced. Simplest extension is the introduction of 2 states with eV scale mass splittings, ν_4 and ν_5. The ordering of the states can be such that Δm^2_{41} and Δm^2_{51} are both positive (“3+2”) or one of them is negative (“1+3+1”). From the point of view of the description of the data the most important new qualitative feature in that now non-zero CP violation at $E/L \sim \text{eV}^2$ is possibly observable [97, 102–104]. This allows some additional freedom in fitting neutrino versus anti-neutrino data from LSND and MiniBooNE together. However, it still holds that a non-zero $\nu_\mu \to \nu_e$ appearance at SBL necessarily predicts SBL disappearance for both ν_e and ν_μ. So, generically, the tension between appearance and disappearance results remains, thought differences in the methodology of statistical quantification of the degree of agreement/disagreement in these scenarios can lead to different conclusions on whether they can provide a successful description of all the data [94, 100, 101].

At present there is an active experimental program to further test these anomalies but the results are still inconclusive.

Cosmological observations can provide complementary information on the number of relativistic neutrino states in thermal equilibrium in the early Universe and on the sum of their masses which sets further constrains on light sterile neutrinos scenarios.

3.3 Beyond the 3ν paradigm: Non-standard interactions

Another extension of the 3ν flavour transitions scenario is that of non-standard neutrino interactions (NSI) with matter. In particular, neutral current NSI’s, which can impact the coherent scattering of neutrinos in matter. They can be parametrized by effective four-fermion operators of the form

$$\mathcal{L}_{\text{NSI}} = -2\sqrt{2} G_F \varepsilon^{fP}_{\alpha\beta} (\bar\nu_\alpha \gamma^\mu L \nu_\beta)(\bar f \gamma_\mu P f) \,, \tag{98}$$

where $f = e, u, d$ is a charged fermion, $P = (L, R)$ and $\varepsilon^{fP}_{\alpha\beta}$ are dimensionless parameters encoding the deviation from standard interactions. These operators contribute to the effective matter potential in the Hamiltonian describing the evolution of the neutrino flavour state:

$$H_{\text{mat}} = \sqrt{2} G_F N_e(x) \begin{pmatrix} 1+\epsilon_{ee} & \epsilon_{e\mu} & \epsilon_{e\tau} \\ \epsilon^*_{e\mu} & \epsilon_{\mu\mu} & \epsilon_{\mu\tau} \\ \epsilon^*_{e\tau} & \epsilon^*_{\mu\tau} & \epsilon_{\tau\tau} \end{pmatrix} , \quad \text{with} \quad \epsilon_{\alpha\beta}(x) = \sum_{f=e,u,d} \frac{N_f(x)}{N_e(x)} \epsilon^{f,V}_{\alpha\beta} \,, \tag{99}$$

with $N_f(x)$ being the density of fermion f along the neutrino path and $\epsilon^{f,V}_{\alpha\beta} = \epsilon^{f,L}_{\alpha\beta} + \epsilon^{f,R}_{\alpha\beta}$. The “1” in the ee entry in Eq. (99) corresponds to the SM matter potential. Therefore, the effective NSI parameters entering oscillations, $\epsilon_{\alpha\beta}$, may depend on x and will be generally different for neutrinos crossing the Earth or the solar medium and as such can be constrained by the global analysis of neutrino oscillation data.

The task becomes troubled by an intrinsic degeneracy in the Hamiltonian governing neutrino oscillations which is introduced by the NSI-induced matter potential. In general, CPT implies that neutrino evolution is invariant if the relevant Hamiltonian is transformed as $H \to -H^*$. In vacuum this transformation can be realized by changing the oscillation parameters as

$$\Delta m^2_{31} \to -\Delta m^2_{31} + \Delta m^2_{21} = -\Delta m^2_{32}\,, \quad \sin\theta_{12} \leftrightarrow \cos\theta_{12}\,, \quad \delta_{\rm CP} \to \pi - \delta_{\rm CP}\,. \tag{100}$$

In the standard 3ν oscillation scenario, this symmetry is broken by the standard matter potential, and this allows for the determination of the octant of θ_{12} and (in principle) of the sign of Δm^2_{31}. However, in the presence of NSI, the symmetry can be restored if in addition to the transformation Eq. (100), NSI parameters are transformed as

$$(\varepsilon_{ee} - \varepsilon_{\mu\mu}) \to -(\varepsilon_{ee} - \varepsilon_{\mu\mu}) - 2\,, \quad (\varepsilon_{\tau\tau} - \varepsilon_{\mu\mu}) \to -(\varepsilon_{\tau\tau} - \varepsilon_{\mu\mu})\,, \quad \varepsilon_{\alpha\beta} \to -\varepsilon^*_{\alpha\beta} \qquad (\alpha \neq \beta)\,. \tag{101}$$

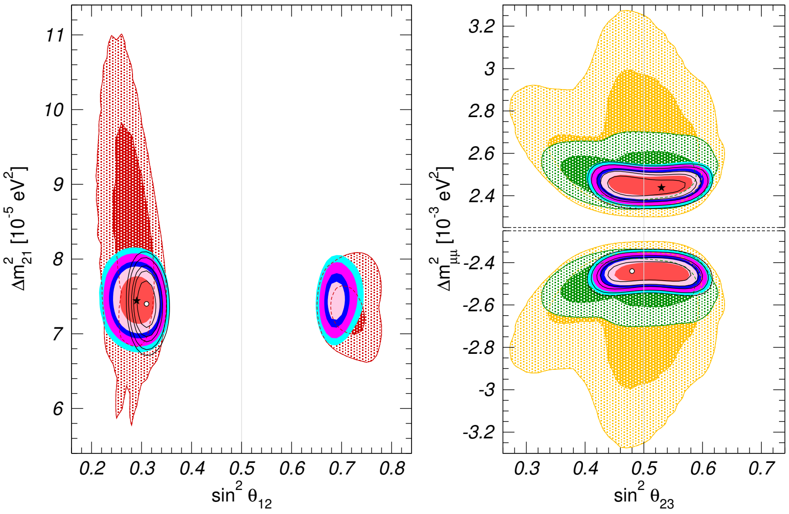


Fig. 15: Two-dimensional projections of the allowed regions onto different vacuum parameters (on the right $\Delta m^2_{\mu\mu} \simeq \Delta m^2_{31}$) after marginalizing over the matter potential parameters and the not displayed oscillation parameters. The solid colored regions correspond to the global analysis of all oscillation data, and show the 1σ, 90%, 2σ, 99% and 3σ CL allowed regions; the best fit point is marked with a star. The black void regions correspond to the analysis with the standard matter potential (*i.e.*, without NSI) and its best fit point is marked with an empty dot. For comparison, in the left panel we show in red the 90% and 3σ allowed regions including only solar and KamLAND results, while in the right panels we show in green the 90% and 3σ allowed regions excluding solar and KamLAND data, and in yellow the corresponding ones excluding also IceCube and reactor data.

This degeneracy can be seen in Fig. 15 where I show the two-dimensional projections of the allowed regions onto different sets of oscillation parameters from the global analysis in Ref. [105] in the presence of this generalized matter potential, Eq. (99). These regions are obtained after marginalizing over the not displayed vacuum parameters as well as the NSI couplings. For comparison its also shown as black-contour void regions the corresponding results with the standard matter potential, *i.e.*, in the absence of NSI.

From the figure we read the following:

• The determination of the oscillation parameters discussed in the previous section is robust under the presence of NSI as large as allowed by the oscillation data itself with the exception of the octant of θ_{12}. This result relies on the complementarity and synergies between the different data sets, which allows to constrain those regions of the parameter space where cancellations between standard and non-standard effects occur in a particular data set.

- A solution with $\theta_{12} > 45°$ still provides a good fit. This is the *so-called* LMA Dark (LMA-D) solution and it was first found in Ref. [106]. It is is a consequence of the intrinsic degeneracy in the Hamiltonian described above. Eq. (100) shows that this degeneracy implies a change in the octant of θ_{12} (as manifest in the LMA-D). As such it cannot be ruled out by oscillation data only. Scattering data, in particular from the finally-observed coherent scattering in nuclei [107] disfavoured it at more then 3σ for NSI coupling neutrinos with either up or down quarks [108]. But it is still allowed for more general NSI couplings [105, 109].

The results of the oscillation analysis show that LMA-D requires large $\varepsilon_{ee} - \varepsilon_{\mu\mu} \sim \mathcal{O}(2)$ which are therefore still allowed. But for all other couplings the same global analysis sets strong constrains on $\varepsilon_{\alpha\beta}$ yielding the most restrictive bounds on the NSI parameters, in particular those involving τ flavour.

3.4 Some implications

The need of new physics and its scale

As we discussed in the first lecture, the SM is a gauge theory based on the gauge symmetry $SU(3)_{\rm C} \times SU(2)_{\rm L} \times U(1)_{\rm Y}$ spontaneously broken to $SU(3)_{\rm C} \times U(1)_{\rm EM}$ by the the vacuum expectations value (VEV), v, of the a Higgs doublet field ϕ with three fermion generations which reside in chiral representations of the gauge group as required by the interactions. No right-handed neutrino is included in the model since neutrinos are neutral.

In the SM, fermion masses arise from the Yukawa interactions, Eq. (8). But mo Yukawa interaction can be written that would give mass to the neutrino because no right-handed neutrino field exists in the model. We also argue that neutrino masses could not arise from loop corrections or from non-perturbative effects on the basis of the global symmetries of the model. More precisely, the SM, presents the accidental global symmetry in Eq. (7) which implies that total lepton number $L = L_e + L_\mu + L_\tau$ is a global symmetry of the SM. Therefore any term form from loop corrections within this model must conserve total lepton number.

But with the SM particle content the only mass term (that is, the only operator involving a left-handed and a right-handed fermion field) for the neutrino which could be generated would be of the form

$$\left(\bar{L}_{Li}\tilde{\phi}\right)\left(\phi^{+} L^{C}_{Lj}\right) + \text{h.c.}, \tag{102}$$

($L^C_{Li} = C\bar{L}^T_{Li}$) which violates $G^{\rm global}_{\rm SM}$ (in particular in violates total lepton number). Therefore it cannot be generated by SM loop corrections. Also, it cannot be generated by non-perturbative effects.

In other words, the SM predicts that neutrinos are precisely massless and consequently, there is neither mixing nor CP violation in the leptonic sector. Thus the simplest and most straightforward lesson of the experimental evidence for neutrino masses is also the most striking one: *there is new physics beyond the SM*. This has been the first experimental result that is inconsistent with the SM.

Furthermore the determined ranges of neutrino masses and leptonic mixing raise two main questions:

- Why are neutrinos so light?, which is directly related to issue of the origin of neutrino mass.
- Why is lepton mixing so different from quark mixing?, which is related to the flavour puzzle.

A possible way to address these questions it to realize that if the SM is not a complete picture of Nature, then new physics (NP) is expected to appear at some higher energies. In this case the SM is an effective low energy theory valid up to the scale $\Lambda_{\rm NP}$ which characterizes the NP. In this approach, the gauge group, the fermionic spectrum, and the pattern of spontaneous symmetry breaking are still valid ingredients to describe Nature at energies $E \ll \Lambda_{\rm NP}$. The difference between the SM as a complete description of Nature and as a low energy effective theory is that in the latter case we must consider also non-renormalizable (dim>4) terms in the Lagrangian whose effect will be suppressed by powers $1/\Lambda_{\rm NP}^{\rm dim-4}$. In this approach the largest effects at low energy are expected to come from dim$=5$ operators

There is a single set of dimension-five terms that is made of SM fields and is consistent with the gauge symmetry given by

$$\mathcal{O}_5 = \frac{c_{5ij}}{2\Lambda_{\rm NP}} \left(\bar{L}_{Li}\tilde{\phi}\right)\left(\tilde{\phi}^T L^C_{Lj}\right) + {\rm h.c.}, \tag{103}$$

which violates total lepton number by two units and leads, upon spontaneous symmetry breaking, to:

$$-L_{M_\nu} = \frac{c_{5ij}}{4}\frac{v^2}{\Lambda_{\rm NP}}\overline{\nu^c}_i\nu_j + {\rm h.c.}\ . \tag{104}$$

Comparing with Eqs. (13) (85) we see that this is a Majorana neutrino mass with:

$$(M_\nu)_{ij} = \frac{c_{5ij}}{2}\frac{v^2}{\Lambda_{\rm NP}}. \tag{105}$$

Equation (105) arises in a generic extension of the SM which means that neutrino masses are very likely to appear if there is NP. Furthermore comparing Eq. (105) and Eq. (9) we find that the scale of neutrino masses is suppressed by $v/\Lambda_{\rm NP}$ when compared to the scale of charged fermion masses providing an explanation not only for the existence of neutrino masses but also for their smallness. Finally, Eq. (105) breaks not only total lepton number but also the lepton flavor symmetry $U(1)_e \times U(1)_\mu \times U(1)_\tau$. Therefore we should expect lepton mixing and CP violation.

Given the relation (105), $m_\nu \sim v^2/\Lambda_{\rm NP}$, it is straightforward to use the measured neutrino masses to estimate the scale of NP that is relevant to their generation. In particular, if there is no quasi-degeneracy in the neutrino masses, the heaviest of the active neutrino masses can be estimated, $m_h = m_3 \sim \sqrt{\Delta m^2_{31}} \approx 0.05$ eV (in the case of inverted hierarchy the implied scale is $m_h = m_2 \sim \sqrt{|\Delta m^2_{31}|} \approx 0.05$ eV). It follows that the scale in the non-renormalizable term (103) is given by

$$\Lambda_{\rm NP} \sim v^2/m_h \approx 10^{15}\ {\rm GeV}. \tag{106}$$

We should clarify two points regarding Eq. (106):

1. There could be some level of degeneracy between the neutrino masses that are relevant to the atmospheric neutrino oscillations. In such a case Eq. (106) becomes an upper bound on the scale of NP.

2. It could be that the $c_{5\alpha\beta}$ couplings of Eq. (103) are much smaller than one. In such a case, again, Eq. (106) becomes an upper bound on the scale of NP.

The estimate Eq. (106) is very exciting. First, the upper bound on the scale of NP is well below the Planck scale. This means that there is a new scale in Nature which is intermediate between the two known scales, the Planck scale $m_{\rm Pl} \sim 10^{19}$ GeV and the electroweak breaking scale, $v \sim 10^2$ GeV. Second, the scale $\Lambda_{\rm NP} \sim 10^{15}$ GeV is intriguingly close to the scale of gauge coupling unification.

In simple renormalizable realizations of NP this dimension-5 operator can be generated by the tree-level exchange of three types of new particles (see Fig. 16):

- Type-I and Type-III see–saw : One adds at least two fermionic singlets (Type-I) or triplets (Type-III) of mass M and Yukawa couplings λ. The neutrino masses are as Eq. (105) with $\Lambda_{\rm NP} = M$ and $c_5 \sim \lambda^2$.

- Type-II see–saw: One adds an $SU(2)_L$ Higgs triplet Δ of mass M which couples to the SM $SU(2)_L$ leptons with coupling f, with a neutral component and scalar doublet-triplet mixing μ term in the scalar potential. The neutrino masses are as Eq. (105) with $\Lambda_{\rm NP} = M^2/\mu$ and $c_5 \sim f$.

Of course, neutrinos could be conventional Dirac particles described as in Eq. (84) and we would be left in the darkness on the reason of the smallness of the neutrino mass.

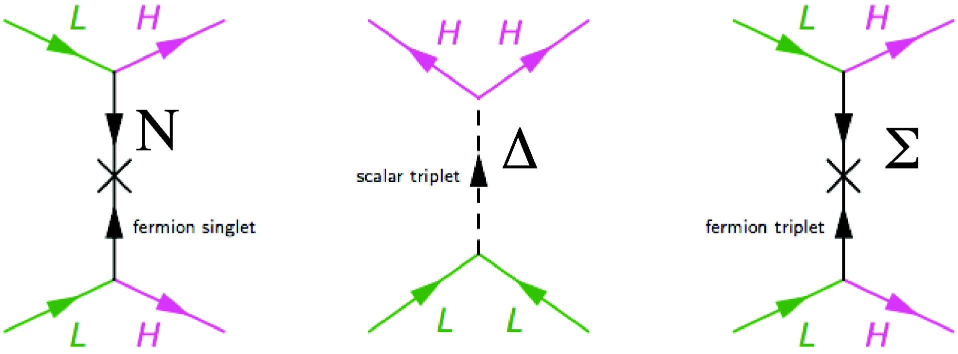


Fig. 16: Tree level diagrams for the Type-I,II and III see–saw, leading to the dim-5 operator for neutrino mass after integrating out the intermediate state

The possibility of leptogenesis

An interesting consequence of neutrinos acquiring their mass via the generic scenario described above is the possibility of explaining the cosmic matter-antimatter asymmetry via the process of leptogenesis [110] in the early Universe.

From what we see and measure, the Universe is made of particles and not of antiparticles. This fact can be quantified in terms of the difference between the density of baryons and antibaryons normalized to the density of photons:

$$Y_B = \frac{n_B - n_{\bar{B}}}{n_\gamma} \sim \frac{n_B}{n_\gamma} \tag{107}$$

From the Big-Bang nucleosynthesis and from the precise data on measurements of the cosmic microwave background, we know that this asymmetry is tiny:

$$Y_B \approx 5 \times 10^{-10} \tag{108}$$

In a seminal paper, Sakharov [111] established the three conditions that any particle physics theory should verify to be able to generate this asymmetry

- Total baryon number B must be violated,
- C and CP must be violated,
- The process which violate these symmetries must occur out of thermal equilibrium.

In principle the SM verifies these conditions because $B+L$ are violated by non-perturbative effects, CP is violated by the CP phase of the CKM quark mixing matrix, and there is departure from thermal equilibrium at the electroweak phase transition provided it is a first order transition. However within the present bounds of the Higgs mass the electroweak phase transition is not strong first order and furthermore the CKM CP violation is too suppressed. As a consequence $Y_{B,SM} \ll 10^{-10}$.

Leptogenesis [110] is the possible origin of such a small asymmetry related to neutrino physics. In a possible realization of leptogenesis, $L \neq 0$ is generated in the Early Universe by the decay of one of the heavy right-handed neutrinos of the type-I see-saw mechanism with CP being violated in the decay. In this case we have:

- Total lepton number is violated by the Majorana mass term of the right-handed neutrinos.
- Due to the interference between the tree-level and one-loop diagrams shown in Fig. 17 the decay rates of the right-handed neutrino into leptons and anti-leptons can be different, so C and CP can be violated

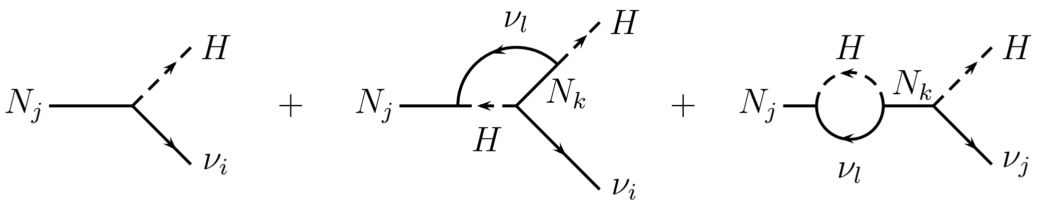


Fig. 17: The tree-level and one-loop diagrams of right-handed neutrino decay into leptons and Higgs.

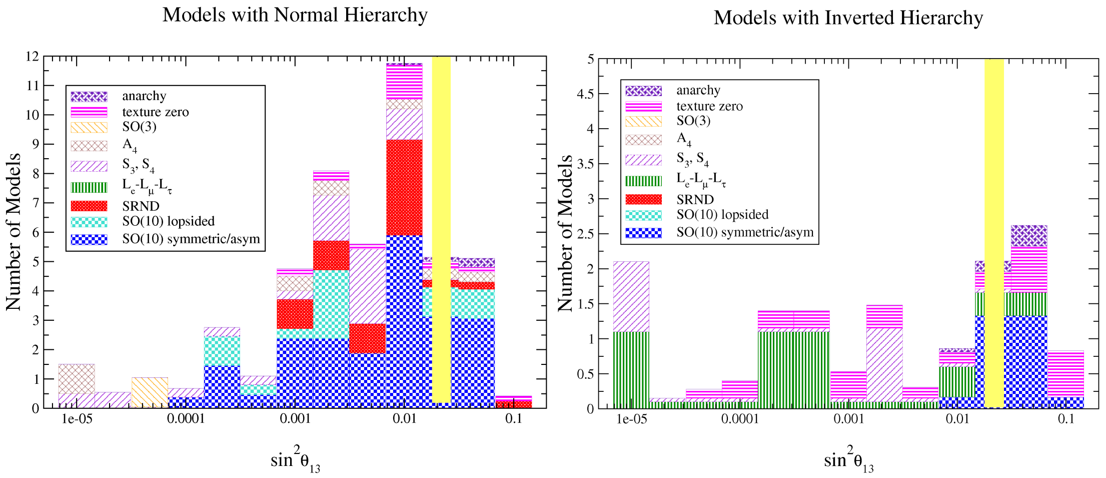


Fig. 18: Compilation in Ref. [113] of the prediction of the value of θ_{13} in several flavour models compared with the present determination.

– The decay can be be out of equilibrium if $\Gamma_{\nu_R} \ll$ Universe expansion rate.

Therefore we have all the conditions to generate total lepton number L in the early Universe.

Non perturbative effects known as *sphaleron* [112] processes transform the lepton asymmetry into a baryon asymmetry and below the electroweak phase transition a net baryon asymmetry is generated $\Delta B \simeq -\frac{\Delta \mathrm{L}}{2}$ (the exact coefficient relating ΔB to ΔL is model dependent.)

The details of the leptogenesis scenario are model dependent and much work has been done in the framework of specific neutrino models. Generically the resulting asymmetry depends on the size of the CP violating phases, the mass of the lightest heavy neutrino and the light neutrino masses. It has been shown that with the present bounds of the neutrino masses and mixing a right-handed neutrino of about 10^{10} GeV can account for the cosmic baryon asymmetry from its out-of-equilibrium decay.

Implications for flavour models

The relevance of the precise determination of the leptonic mixing matrix to address the flavour puzzle is illustrated in Fig. 18 where I show the compilation in Ref. [113] of the predictions of the expected values of θ_{13} is 63 types of flavour models in 2006. As seen from the figure only about 10% of the models survived the precise determination of θ_{13} in 2012.

Among those which did not survive the test of the precise determination of the mixing parameters were the models predicting bimaximal mixing ($\theta_{12} = \theta_{23} = 45°$, $\theta_{13} = 0$), tri-bimaximal mixing ($\theta_{12} = 35.2°$ $\theta_{23} = 45°$, $\theta_{13} = 0$), and the golden ratio ($\theta_{12} = 31.7°$ $\theta_{23} = 45°$, $\theta_{13} = 0$). Generically these structures appear in models with flavour symmetries with the smallest symmetry groups A_4, S_4 and A_5. Consequently either the group has to be enlarged, or corrections to the mixing have to be obtained from other sectors. Generically these attempts lead to new *sum rules* relating the leptonic flavour parameters among themselves and with those of quarks. Relations which can be testable with enough experimental precision. In this respect the next frontier is the precise determination of the ordering of the states.

Neutrino mass models for collider signatures

One may notice that even in the particularly simple forms of NP of the three type of see-saw realizations represented in Fig. 16, the full theory contains very different high–energy particle contents but they lead to the same low energy operator $\mathcal{O}_5$ which contains only 9 parameters and that are everything we can measure at neutrino oscillation experiments. This simple example illustrates the limitation of the "bottom-up" approach in deriving model independent implications of the presently observed neutrino masses and mixing. This is the challenge of performing measurements at a much lower scale than that of the NP.

Alternatively one can go "top-down" by studying the low energy effective neutrino masses and mixing induced by specific high energy models as sketched in the discussion about flavour models above.

The bottom line of this discussion is that in order to advance further in the understanding of the dynamics underlying neutrino masses in a model independent approach we need more (and more precise) data. Furthermore synergy among different types of observations such as charge lepton flavour experiments and collider experiments are probably going to be fundamental in this advance. In this respect I will finish by discussing a possible framework in which this connection between neutrino physics and collider signatures arises.

Generically, at low energies the Lagrangian of the full theory can be expanded as

$$\mathcal{L} = \mathcal{L}_{SM} + \frac{c_5}{\Lambda_{LN}}\mathcal{O}_5 + \sum_i \frac{c_{6,i}}{\Lambda^2_{FL}}\mathcal{O}_{6,i} + \dots \tag{109}$$

where $\mathcal{O}_5$ is Weinberg's operator responsible for neutrino masses given above, and $\mathcal{O}_{6,i}$ are flavour-changing, but lepton number conserving, dimension-6 operators. In writing Eq. (109) we have explicitly denoted Λ_{LN} as the NP scale for lepton number breaking and Λ_{FL} the NP scale for lepton flavour breaking. In this context attractive testable scenarios are those for which it is possible to relate the mass of the new states $M \sim \Lambda_{FL} \sim \mathcal{O}$ (TeV) but still keep $\Lambda_{LN} \gg \Lambda_{FL}$ to explain the smallness of the neutrino mass.

Furthermore to relate the flavour structure of the signals at collider, or low energy charged lepton flavour experiments with that derived from the neutrino sector one would need some connection between the coefficients c_5 and c_6. This is precisely provided by the assumption of minimal lepton flavour violation (MLFV) of the NP. Indeed these conditions are automatically fulfilled by the simplest Type-II see–saw model if a light double-triplet mixing μ is assumed. For LHC phenomenology this leads to the interesting possibility of the production of the triplet scalar states with all their decay modes determined by the neutrino mass parameters which has been therefore extensively searched for at LHC. The possibility of constructing and observing MLFV scenarios of Type-I and Type-III see-saws was explored in Refs. [114–116]

References

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

[2] C.L. Cowan *et al.*, Detection of the free neutrino: A confirmation, *Science*, **124**(3212):103–104, 1956, doi:10.1126/science.124.3212.103.

[3] M. Goldhaber, L. Grodzins, and A.W. Sunyar, Helicity of neutrinos, *Phys. Rev.*, **109**(3):1015–1017, 1958, doi:10.1103/PhysRev.109.1015.

[4] B. Pontecorvo, Neutrino experiments and the problem of conservation of leptonic charge, *Sov. Phys. JETP*, **26**:984–988, 1968, https://inspirehep.net/literature/51319.

[5] V.N. Gribov and B. Pontecorvo, Neutrino astronomy and lepton charge, *Phys. Lett. B*, **28**(7):493–496, 1969, doi:10.1016/0370-2693(69)90525-5.

[6] J.N. Bahcall, *Neutrino astrophysics*, Cambridge Univ. Press, 1989, ISBN:9780521379755.

[7] R.N. Mohapatra and P.B. Pal, *Massive neutrinos in physics and astrophysics.*, World Scientific, Singapore, 3. ed., 2004, doi:10.1142/5024.

[8] C.W. Kim and A. Pevsner, *Neutrinos in physics and astrophysics*, Harwood Academic Publishers, Chur, 1993, ISBN:9783718605675.

[9] B. Kayser, F. Gibrat-Debu, and F. Perrier, *The physics of massive neutrinos*, World Scientific, Singapore, 1989, doi:10.1142/0655.

[10] C. Giunti and C.W. Kim, *Fundamentals of neutrino physics and astrophysics*, Oxford Univ. Press, 2007, doi:10.1093/acprof:oso/9780198508717.001.0001.

[11] M.C. Gonzalez-Garcia and Y. Nir, Neutrino masses and mixing: Evidence and implications, *Rev. Mod. Phys.*, **75**(2):345–402, 2003, doi:10.1103/RevModPhys.75.345.

[12] M.C. Gonzalez-Garcia and M. Maltoni, Phenomenology with massive neutrinos, *Phys. Rept.*, **460**(1-3):1–129, 2008, doi:10.1016/j.physrep.2007.12.004.

[13] M.C. Gonzalez-Garcia and M. Yokoyama, Neutrino mases, mixing, and oscillations, *Prog. Theor. Exp. Phys.*, **2020**(083C01):285–311, 2019, doi:10.1093/ptep/ptaa104.

[14] M. Tanabashi *et al.*, Review of Particle Physics, *Phys. Rev. D*, **98**(3):030001, 2018, doi:10.1103/PhysRevD.98.030001.

[15] P. Minkowski, $\mu \to e\gamma$ at a rate of one out of 10^9 muon decays?, *Phys. Lett. B*, **67**(4):421–428, 1977, doi:10.1016/0370-2693(77)90435-X.

[16] P. Ramond, The family group in grand unified theories, in *Int. Symp. Fundamentals of Quantum Theory and Quantum Field Theory, Palm Coast, Florida, 25 February–2 March, 1979*, pp. 265–280, 1979, arXiv:hep-ph/9809459.

[17] M. Gell-Mann, P. Ramond, and R. Slansky, Complex spinors and unified theories, in *Supergravity Workshop Stony Brook, New York, 27–28 September, 1979*, pp. 315–321, 1979, arXiv:1306.4669.

[18] T. Yanagida, Horizontal gauge symmetry and masses of neutrinos, in *Proc. Workshop on the Unified Theories and the Baryon Number in the Universe: Tsukuba, Japan, 13–14 February, 1979*, pp. 95–99, 1979, https://neutrino.kek.jp/seesaw/KEK-79-18-Yanagida.pdf.

[19] R.N. Mohapatra and G. Senjanovic, Neutrino mass and spontaneous parity nonconservation, *Phys. Rev. Lett.*, **44**(14):912–915, 1980, doi:10.1103/PhysRevLett.44.912.

[20] Z. Maki, M. Nakagawa, and S. Sakata, Remarks on the unified model of elementary particles, *Prog. Theor. Phys.*, **28**(5):870–880, 1962, doi:10.1143/PTP.28.870.

[21] N. Cabibbo, Unitary symmetry and leptonic decays, *Phys. Rev. Lett.*, **10**(12):531–533, 1963, doi:10.1103/PhysRevLett.10.531.

[22] M. Kobayashi and T. Maskawa, CP Violation in the renormalizable theory of weak interaction, *Prog. Theor. Phys.*, **49**(2):652–657, 1973, doi:10.1143/PTP.49.652.

[23] M. Aker *et al.*, An improved upper limit on the neutrino mass from a direct kinematic method by KATRIN, *Phys.Rev.Lett.*, **123**(22):221802, 2019, doi:10.1103/PhysRevLett.123.221802.

[24] C. Weinheimer *et al.*, The Mainz neutrino mass experiment, *PoS*, **hep2001**:192, 2001, doi:10.22323/1.007.0192.

[25] J. Bonn *et al.*, The Mainz neutrino mass experiment, *Nucl. Phys. B Proc. Suppl.*, **91**(1-3):273–279, 2001, doi:10.1016/S0920-5632(00)00951-8.

[26] V. M. Lobashev *et al.*, Direct search for neutrino mass and anomaly in the tritium beta-spectrum: Status of 'Troitsk neutrino mass' experiment, *Nucl. Phys. B Proc. Suppl.*, **91**(1-3):280–286, 2001, doi:10.1016/S0920-5632(00)00952-X.

[27] J. Schechter and J. W. F. Valle, Neutrinoless double beta decay in SU(2) x U(1) theories, *Phys. Rev. D*, **25**(11):2951, 1982, doi:10.1103/PhysRevD.25.2951.

[28] A. Gando *et al.*, Search for Majorana neutrinos near the inverted mass hierarchy region with KamLAND-Zen, *Phys. Rev. Lett.*, **117**(8):082503, 2016, doi:10.1103/PhysRevLett.117.082503, Erratum: doi:10.1103/PhysRevLett.117.109903.

[29] J. Engel and J. Menendez, Status and future of nuclear matrix elements for neutrinoless double-beta decay: A review, *Rept. Prog. Phys.*, **80**(4):046301, 2017, doi:10.1088/1361-6633/aa5bc5.

[30] E.Kh. Akhmedov, Quantum mechanics aspects and subtleties of neutrino oscillations, in *Proc. Int. Conf. History of the Neutrino: 1930-2018 Paris, France, 5–7 September, 2018*, 2019, arXiv:1901.05232.

[31] L. Wolfenstein, Neutrino oscillations in matter, *Phys. Rev. D*, **17**(9):2369–2374, 1978, doi:10.1103/PhysRevD.17.2369.

[32] A. Halprin, Neutrino oscillations in nonuniform matter, *Phys. Rev. D*, **34**(11):3462–3466, 1986, doi:10.1103/PhysRevD.34.3462.

[33] P.D. Mannheim, Derivation of the formalism for neutrino matter oscillations from the neutrino relativistic field equations, *Phys. Rev. D*, **37**(7):1935, 1988, doi:10.1103/PhysRevD.37.1935.

[34] A.J. Baltz and J. Weneser, Matter oscillations: Neutrino transformation in the sun and regeneration in the earth, *Phys. Rev. D*, **37**(12):3364, 1988, doi:10.1103/PhysRevD.37.3364.

[35] S.P. Mikheyev and A. Yu. Smirnov, Resonance amplification of oscillations in matter and spectroscopy of solar neutrinos, *Sov. J. Nucl. Phys.*, **42**:913–917, 1985, https://inspirehep.net/literature/228623.

[36] T.K. Kuo and J. Pantaleone, Neutrino oscillations in matter, *Rev. Mod. Phys.*, **61**(4):937, 1989, doi:10.1103/RevModPhys.61.937.

[37] S.J. Parke, Nonadiabatic level crossing in resonant neutrino oscillations, *Phys. Rev. Lett.*, **57**(10):1275–1278, 1986, doi:10.1103/PhysRevLett.57.1275.

[38] W.C. Haxton, Adiabatic conversion of solar neutrinos, *Phys. Rev. Lett.*, **57**(10):1271–1274, 1986, doi:10.1103/PhysRevLett.57.1271.

[39] S.T. Petcov, On the nonadiabatic neutrino oscillations in matter, *Phys. Lett. B*, **191**(3):299–303, 1987, doi:10.1016/0370-2693(87)90259-0.

[40] M. Agostini *et al.*, Comprehensive measurement of *pp*-chain solar neutrinos, *Nature*, **562**(7728):505–510, 2018, doi:10.1038/s41586-018-0624-y.

[41] N. Vinyoles *et al.*, A new generation of standard solar models, *Astrophys. J.*, **835**(2):202, 2017, doi:10.3847/1538-4357/835/2/202.

[42] J. Bergström *et al.*, Updated determination of the solar neutrino fluxes from solar neutrino data, *JHEP*, **03**:132, 2016, doi:10.1007/JHEP03(2016)132.

[43] J.N. Bahcall, N.A. Bahcall, and G. Shaviv, Present status of the theoretical predictions for the Cl-36 solar neutrino experiment, *Phys. Rev. Lett.*, **20**(21):1209–1212, 1968, doi:10.1103/PhysRevLett.20.1209.

[44] R. Davis, Jr., D.S. Harmer, and K.C. Hoffman, Search for neutrinos from the sun, *Phys. Rev. Lett.*, **20**(21):1205–1209, 1968, doi:10.1103/PhysRevLett.20.1205.

[45] B.T. Cleveland *et al.*, Measurement of the solar electron neutrino flux with the Homestake chlorine detector, *Astrophys. J.*, **496**(1):505–526, 1998, doi:10.1086/305343.

[46] 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**:181–193, 2002, doi:10.1134/1.1506424.

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

[48] M. Altmann *et al.*, Complete results for five years of GNO solar neutrino observations, *Phys. Lett.*, **B616**:174–190, 2005, doi:10.1016/j.physletb.2005.04.068.

[49] K.S. Hirata *et al.*, Real time, directional measurement of ^{8}B solar neutrinos in the Kamiokande-II detector, *Phys. Rev.*, **D44**:2241, 1991, doi:10.1103/PhysRevD.44.2241, Erratum: *Phys. Rev.*, **D45**:2170, 1992, doi:10.1103/PhysRevD.45.2170.

[50] Y. Fukuda *et al.*, Measurements of the solar neutrino flux from Super-Kamiokande's first 300 days, *Phys. Rev. Lett.*, **81**:1158–1162, 1998, doi:10.1103/PhysRevLett.81.1158, Erratum: *Phys. Rev. Lett.*, **81**:4279, 1998, doi:10.1103/PhysRevLett.81.4279.

[51] S. Fukuda *et al.*, Constraints on neutrino oscillations using 1258 days of Super-Kamiokande solar neutrino data, *Phys. Rev. Lett.*, **86**:5656–5660, 2001, doi:10.1103/PhysRevLett.86.5656.

[52] S. Abe *et al.*, Precision measurement of neutrino oscillation parameters with KamLAND, *Phys. Rev. Lett.*, **100**:221803, 2008, doi:10.1103/PhysRevLett.100.221803.

[53] M. Honda *et al.*, Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model, *Phys. Rev.*, **D92**(2):023004, 2015, doi:10.1103/PhysRevD.92.023004.

[54] G.D. Barr *et al.*, Uncertainties in atmospheric neutrino fluxes, *Phys. Rev.*, **D74**:094009, 2006, doi:10.1103/PhysRevD.74.094009.

[55] G. Battistoni *et al.*, The FLUKA atmospheric neutrino flux calculation, *Astropart. Phys.*, **19**:269–290, 2003, doi:10.1016/S0927-6505(02)00246-3, Erratum: *Astropart. Phys.*, **19**:291, 2003, doi:10.1016/S0927-6505(03)00107-5.

[56] J. Evans *et al.*, Uncertainties in atmospheric muon-neutrino fluxes arising from cosmic-ray primaries, *Phys. Rev.*, **D95**(2):023012, 2017, doi:10.1103/PhysRevD.95.023012.

[57] F. Reines *et al.*, Evidence for high-energy cosmic ray neutrino interactions, *Phys. Rev. Lett.*, **15**:429–433, 1965, doi:10.1103/PhysRevLett.15.429.

[58] C.V. Achar *et al.*, Detection of muons produced by cosmic ray neutrinos deep underground, *Phys. Lett.*, **18**:196–199, 1965, doi:10.1016/0031-9163(65)90712-2.

[59] Y. Fukuda *et al.*, Evidence for oscillation of atmospheric neutrinos, *Phys. Rev. Lett.*, **81**:1562–1567, 1998, doi:10.1103/PhysRevLett.81.1562.

[60] M.G. Aartsen *et al.*, Measurement of atmospheric neutrino oscillations at 6–56 GeV with IceCube DeepCore, *Phys. Rev. Lett.*, **120**(7):071801, 2018, doi:10.1103/PhysRevLett.120.071801.

[61] S.H. Ahn *et al.*, Detection of accelerator produced neutrinos at a distance of 250 km, *Phys. Lett.*, **B511**:178–184, 2001, doi:10.1016/S0370-2693(01)00647-5.

[62] D.G. Michael *et al.*, The magnetized steel and scintillator calorimeters of the MINOS experiment, *Nucl. Instrum. Meth.*, **A596**:190–228, 2008, doi:10.1016/j.nima.2008.08.003.

[63] M.A. Acero *et al.*, First measurement of neutrino oscillation parameters using neutrinos and antineutrinos by NOvA, *Phys. Rev. Lett.*, **123**:151803, 2019, doi:10.1103/PhysRevLett.123.151803.

[64] M. Apollonio *et al.*, Search for neutrino oscillations on a long baseline at the CHOOZ nuclear power station, *Eur.Phys.J.*, **C27**:331–374, 2003, doi:10.1140/epjc/s2002-01127-9.

[65] F. Boehm *et al.*, Final results from the Palo Verde neutrino oscillation experiment, *Phys. Rev.*, **D64**:112001, 2001, doi:10.1103/PhysRevD.64.112001.

[66] D. Adey *et al.*, Measurement of the electron antineutrino oscillation with 1958 days of operation at Daya Bay, *Phys. Rev. Lett.*, **121**(24):241805, 2018, doi:10.1103/PhysRevLett.121.241805.

[67] G. Bak *et al.*, Measurement of reactor antineutrino oscillation amplitude and frequency at RENO, *Phys. Rev. Lett.*, **121**(20):201801, 2018, doi:10.1103/PhysRevLett.121.201801.

[68] H. de Kerret *et al.*, Double Chooz θ_{13} measurement via total neutron capture detection, *Nature Phys.*, **16**(5):558–564, 2019, doi:10.1038/s41567-020-0831-y.

[69] N. Cabibbo, Time reversal violation in neutrino oscillation, *Phys. Lett.*, **72B**:333–335, 1978, doi:10.1016/0370-2693(78)90132-6.

[70] S.M. Bilenky, J. Hosek, and S.T. Petcov, On oscillations of neutrinos with Dirac and Majorana masses, *Phys. Lett.*, **94B**:495–498, 1980, doi:10.1016/0370-2693(80)90927-2.

[71] V.D. Barger, K. Whisnant, and R.J.N. Phillips, CP nonconservation in three neutrino oscillations, *Phys. Rev. Lett.*, **45**:2084, 1980, doi:10.1103/PhysRevLett.45.2084.

[72] P. Langacker *et al.*, Implications of the Mikheev–Smirnov–Wolfenstein (MSW) mechanism of amplification of neutrino oscillations in matter, *Nucl. Phys.*, **B282**:589–609, 1987, doi:10.1016/0550-3213(87)90699-7.

[73] I. Esteban *et al.*, Global analysis of three-flavour neutrino oscillations: synergies and tensions in the determination of $\theta_{23}, \delta_{\rm CP}$, and the mass ordering, *JHEP*, **01**:106, 2019, doi:10.1007/JHEP01(2019)106.

[74] I. Esteban *et al.*, NuFIT4.1 at NuFit webpage, `http://www.nu-fit.org`, last accessed 22 March 2022.

[75] F. Capozzi *et al.*, Current unknowns in the three neutrino framework, *Prog. Part. Nucl. Phys.*, **102**:48–72, 2018, doi:10.1016/j.ppnp.2018.05.005.

[76] P.F. de Salas *et al.*, Status of neutrino oscillations 2018: 3σ hint for normal mass ordering and improved CP sensitivity, *Phys. Lett.*, **B782**:633–640, 2018, doi:10.1016/j.physletb.2018.06.019.

[77] K. Abe *et al.*, Atmospheric neutrino oscillation analysis with external constraints in Super-Kamiokande I-IV, *Phys. Rev.*, **D97**(7):072001, 2018, doi:10.1103/PhysRevD.97.072001.

[78] M. Friend, Updated results from the T2K experiment with 3.13×10^{21} protons on target, January 2019, KEK/J-PARC Physics seminar, 10 January, 2019, https://t2k.org/docs/talk/335/2019kekseminar.

[79] C. Jarlskog, Commutator of the quark mass matrices in the Standard Electroweak Model and a measure of maximal CP nonconservation, *Phys. Rev. Lett.*, **55**:1039, 1985, doi:10.1103/PhysRevLett.55.1039.

[80] M.C. Gonzalez-Garcia, M. Maltoni, and T. Schwetz, Updated fit to three neutrino mixing: status of leptonic CP violation, *JHEP*, **11**:052, 2014, doi:10.1007/JHEP11(2014)052.

[81] Y. Farzan and A. Yu. Smirnov, Leptonic unitarity triangle and CP violation, *Phys. Rev.*, **D65**:113001, 2002, doi:10.1103/PhysRevD.65.113001.

[82] A. Dueck, S.T. Petcov, and W Rodejohann, On leptonic unitary triangles and boomerangs, *Phys. Rev.*, **D82**:013005, 2010, doi:10.1103/PhysRevD.82.013005.

[83] G.L. Fogli *et al.*, Observables sensitive to absolute neutrino masses: Constraints and correlations from world neutrino data, *Phys. Rev.*, **D70**:113003, 2004, doi:10.1103/PhysRevD.70.113003.

[84] S. Pascoli, S. T. Petcov, and T. Schwetz, The absolute neutrino mass scale, neutrino mass spectrum, Majorana CP-violation and neutrinoless double-beta decay, *Nucl. Phys.*, **B734**:24–49, 2006, doi:10.1016/j.nuclphysb.2005.11.003.

[85] 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**:112007, 2001, doi:10.1103/PhysRevD.64.112007.

[86] A.A. Aguilar-Arevalo *et al.*, A combined $\nu_\mu \to \nu_e$ and $\bar\nu_\mu \to \bar\nu_e$ oscillation analysis of the MiniBooNE excesses, *arXiv:1207.4809*, 2012, arXiv:1207.4809.

[87] M.A. Acero, C. Giunti, and M. Laveder, Limits on ν_e and $\bar{\nu}_e$ disappearance from gallium and reactor experiments, *Phys. Rev.*, **D78**:073009, 2008, doi:10.1103/PhysRevD.78.073009.

[88] C. Giunti and M. Laveder, Statistical significance of the gallium anomaly, *Phys. Rev.*, **C83**:065504, 2011, doi:10.1103/PhysRevC.83.065504.

[89] Th.A. Mueller *et al.*, Improved predictions of reactor antineutrino spectra, *Phys. Rev.*, **C83**:054615, 2011, doi:10.1103/PhysRevC.83.054615.

[90] P. Huber, Determination of anti-neutrino spectra from nuclear reactors, *Phys. Rev.*, **C84**:024617, 2011, doi:10.1103/PhysRevC.84.024617, Erratum: *Phys. Rev.*, **C85**:029901, 2012, doi:10.1103/PhysRevC.85.029901.

[91] G. Mention *et al.*, Reactor antineutrino anomaly, *Phys. Rev.*, **D83**:073006, 2011, doi:10.1103/PhysRevD.83.073006.

[92] J.J. Gomez-Cadenas and M.C. Gonzalez-Garcia, Future ν_τ oscillation experiments and present data, *Z. Phys.*, **C71**:443–454, 1996, doi:10.1007/BF02907002.

[93] M. Maltoni *et al.*, Ruling out four neutrino oscillation interpretations of the LSND anomaly?, *Nucl. Phys.*, **B643**:321–338, 2002, doi:10.1016/S0550-3213(02)00747-2.

[94] M. Dentler *et al.*, Updated global analysis of neutrino oscillations in the presence of eV-scale sterile neutrinos, *JHEP*, **08**:010, 2018, doi:10.1007/JHEP08(2018)010.

[95] C. Giunti and M. Laveder, Status of 3+1 neutrino mixing, *Phys. Rev.*, **D84**:093006, 2011, doi:10.1103/PhysRevD.84.093006.

[96] J.M. Conrad *et al.*, Sterile neutrino fits to short baseline neutrino oscillation measurements, *Adv. High Energy Phys.*, **2013**:163897, 2013, doi:10.1155/2013/163897.

[97] J. Kopp *et al.*, Sterile neutrino oscillations: The global picture, *JHEP*, **05**:050, 2013, doi:10.1007/JHEP05(2013)050.

[98] G.H. Collin *et al.*, First constraints on the complete neutrino mixing matrix with a sterile neutrino, *Phys. Rev. Lett.*, **117**(22):221801, 2016, doi:10.1103/PhysRevLett.117.221801.

[99] S. Gariazzo *et al.*, Updated global 3+1 analysis of short-baseline neutrino oscillations, *JHEP*, **06**:135, 2017, doi:10.1007/JHEP06(2017)135.

[100] A. Diaz *et al.*, Where are we with light sterile neutrinos?, *Phys. Rept.*, **884**:1–59, 2019, doi:10.1016/j.physrep.2020.08.005.

[101] S. Böser *et al.*, Status of light sterile neutrino searches, *Prog. Part. Nucl. Phys.*, **111**:103736, 2020, doi:10.1016/j.ppnp.2019.103736.

[102] G. Karagiorgi *et al.*, Leptonic CP violation studies at MiniBooNE in the (3+2) sterile neutrino oscillation hypothesis, *Phys. Rev.*, **D75**:013011, 2007, doi:10.1103/PhysRevD.75.013011, Erratum: *Phys. Rev.*, **D80**:099902, 2009, doi:10.1103/PhysRevD.80.099902.

[103] M. Maltoni and T. Schwetz, Sterile neutrino oscillations after first MiniBooNE results, *Phys. Rev.*, **D76**:093005, 2007, doi:10.1103/PhysRevD.76.093005.

[104] C. Giunti and M. Laveder, 3+1 and 3+2 sterile neutrino fits, *Phys. Rev.*, **D84**:073008, 2011, doi:10.1103/PhysRevD.84.073008.

[105] I. Esteban *et al.*, Updated constraints on non-standard interactions from global analysis of oscillation data, *JHEP*, **08**:180, 2018, doi:10.1007/JHEP08(2018)180.

[106] O.G. Miranda, M.A. Tortola, and J.W.F. Valle, Are solar neutrino oscillations robust?, *JHEP*, **10**:008, 2006, doi:10.1088/1126-6708/2006/10/008.

[107] D. Akimov *et al.*, Observation of coherent elastic neutrino-nucleus scattering, *Science*, **357**(6356):1123–1126, 2017, doi:10.1126/science.aao0990.

[108] P. Coloma *et al.*, COHERENT enlightenment of the neutrino dark side, *Phys. Rev.*, **D96**(11):115007, 2017, doi:10.1103/PhysRevD.96.115007.

[109] P. Coloma *et al.*, Improved global fit to non-standard neutrino interactions using COHERENT energy and timing data, *JHEP*, **02**:023, 2020, doi:10.1007/JHEP02(2020)023, Addendum: *JHEP*, **12**:071, 2020, doi:10.1007/JHEP12(2020)071.

[110] M. Fukugita and T. Yanagida, Baryogenesis without grand unification, *Phys. Lett.*, **B174**:45–47, 1986, doi:10.1016/0370-2693(86)91126-3.

[111] A.D. Sakharov, Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe, *Pisma Zh. Eksp. Teor. Fiz.*, **5**:32–35, 1967, English transl. reprinted in *Sov.Phys.Usp.*, **34**(5):392, 1991, doi:10.1070/PU1991v034n05ABEH002497.

[112] V.A. Kuzmin, V.A. Rubakov, and M.E. Shaposhnikov, On the anomalous electroweak baryon number nonconservation in the early universe, *Phys. Lett.*, **155B**:36, 1985, doi:10.1016/0370-2693(85)91028-7.

[113] C.H. Albright and M.-C. Chen, Model predictions for neutrino oscillation parameters, *Phys. Rev.*, **D74**:113006, 2006, doi:10.1103/PhysRevD.74.113006.

[114] M.B. Gavela *et al.*, Minimal flavour seesaw models, *JHEP*, **09**:038, 2009, doi:10.1088/1126-6708/2009/09/038.

[115] O.J.P. Eboli, J. Gonzalez-Fraile, and M.C. Gonzalez-Garcia, Neutrino masses at LHC: Minimal lepton flavour violation in Type-III see-saw, *JHEP*, **12**:009, 2011, doi:10.1007/JHEP12(2011)009.

[116] N.R. Agostinho, O.J.P. Eboli, and M.C. Gonzalez-Garcia, LHC Run I bounds on minimal lepton flavour violation in Type-III see-saw: A case study, *JHEP*, **11**:118, 2017, doi:10.1007/JHEP11(2017)118.

Proceedings of the 2019 European School of High-Energy Physics, St Petersburg, Russia, 14–27 September 2019, edited by C. Duhr and M. Mulders, CERN-2021-005 (CERN, Geneva, 2021)

Cosmology and dark matter

V.A. Rubakov
Institute for Nuclear Research of the Russian Academy of Sciences, Moscow, Russia
Department of Particle Physics and Cosmology, Moscow State University, Moscow, Russia

Abstract
Cosmology and astroparticle physics give the strongest possible evidence for the incompleteness of the Standard Model of particle physics. Leaving aside the mysterious dark energy, which may or may not be just the cosmological constant, two properties of the Universe cannot be explained by the Standard Model: dark matter and matter-antimatter asymmetry. Dark matter particles may well be discovered in foreseeable future; this issue is under intense experimental investigation. Theoretical hypotheses on the nature of the dark matter particles are numerous, so we concentrate on several well motivated candidates, such as weakly interacting massive particles, axions and sterile neutrinos, and also give examples of less motivated and more elusive candidates such as fuzzy dark matter. This gives an idea of the spectrum of conceivable dark matter candidates, while certainly not exhausting it. We then consider the matter-antimatter asymmetry and discuss whether it may result from physics at 100 GeV–TeV scale. Finally, we turn to the earliest epoch of the cosmological evolution. Although the latter topic does not appear immediately related to contemporary particle physics, it is of great interest due to its fundamental nature. We emphasize that the cosmological data, notably, on cosmic microwave background anisotropies, unequivocally show that the well understood hot stage was not the earliest one. The best guess for the earlier stage is inflation, which is consistent with everything known to date; however, there are alternative scenarios. We discuss the ways to study the earliest epoch, with emphasis on future cosmological observations.

Keywords
Cosmology; Dark matter; Dark energy; Axions; Baryon asymmetry; Lectures.

1 Introduction

It is a commonplace by now that cosmology and astroparticle physics, on the one side, and particle physics, on the other, are deeply interrelated. Indeed, the gross properties of the Universe—the existence of dark matter and the very presence of conventional, baryonic matter—call for the extension of the Standard Model of particle physics. A fascinating possibility is that the physics behind these phenomena is within reach of current or future terrestrial experiments. The experimental programs in these directions are currently intensely pursued.

Another aspect of cosmology, which currently does not appear directly related to terrestrial particle physics experiments, is the earliest epoch of the evolution of the Universe. On the one hand, there is no doubt that the usual hot epoch was preceded by another, much less conventional stage. This knowledge comes from the study of inhomogeneities in the Universe through the measurements of cosmic microwave background (CMB) anisotropies, as well as matter distribution (galaxies, clusters of galaxies, voids) in the present and recent Universe. On the the other hand, we know only rather general properties of the cosmological perturbations, which, we are convinced, were generated before the hot epoch. For this reason, we cannot be sure about the earliest epoch; the best guess is inflation, but alternatives to inflation have not yet been ruled out. It is conceivable that future cosmological observations will be

http://doi.org/10.23730/CYRSP-2021-005.129

able to disentangle between different hypotheses; it is amazing that the study of the Universe at large will possibly reveal the properties of the very early epoch characterized by enormous energy density and evolution rate.

Cosmology and astroparticle physics is a large area of research, so we will be unable to cover it to any level of completeness. On the dark matter side, the number of proposals for dark matter objects invented by theorists in more than 30 years is enormous, so we do not attempt even to list them. Instead, we concentrate on a few hypotheses which may or may not have to do with reality. Namely, we study reasonably well motivated candidates—weakly interacting massive particles (WIMPs), axions, sterile neutrinos—and also discuss more exotic possibilities. On the baryon asymmetry side, we focus on scenarios for its generation which employ physics accessible by terrestrial experiments. A particular mechanism of this sort is the electroweak baryogenesis. The last part of these lectures deals with the earliest cosmology—inflation and its alternatives.

To end up this introduction, we point out that most of the topics we discuss are studied, in one or another way, in books [1]. There are of course numerous reviews, some of which will be referred to in appropriate places.

2 Homogeneous and isotropic Universe

2.1 Friedmann–Lemaître–Robertson–Walker metric

When talking about the Universe, we will always mean its visible part. The visible part is, almost for sure, a small, and maybe even tiny patch of a huge space; for the time being (at least) we cannot tell what is outside the part we observe. At large scales the (visible part of the) Universe is *homogeneous and isotropic*: all regions of the Universe are the same, and no direction is preferred. Homogeneous and isotropic three-dimensional spaces can be of three types. These are three-sphere, flat (Euclidean) space and three-hyperboloid.

A basic property of our Universe is that it expands: the space stretches out. This is encoded in the space-time metric (Friedmann–Lemaître–Robertson–Walker, FLRW)

$$ds^2 = dt^2 - a^2(t)d\mathbf{x}^2 \ , \tag{1}$$

where $d\mathbf{x}^2$ is the distance on unit three-sphere or Euclidean space or hyperboloid, $a(t)$ is the scale factor. Observationally, the three-dimensional space is Euclidean (flat) to good approximation (see, however, Ref. [2] where it is claimed that Planck lensing data prefers a closed Universe), so we will treat $d\mathbf{x}^2 = \delta_{ij}dx^i dx^j$, $i, j = 1, 2, 3$, as line interval in three-dimensional Euclidean space.

The coordinates $\mathbf{x}$ are comoving. This means that they label positions of free, static particles in space (one has to check that world lines of free static particles obey $\mathbf{x} = \text{const}$; this is indeed the case). As an example, distant galaxies stay at fixed $\mathbf{x}$ (modulo peculiar motions, if any). In our expanding Universe, the scale factor $a(t)$ increases in time, so the distance between free masses of fixed spatial coordinates $\mathbf{x}$ grows, $dl^2 = a^2(t)d\mathbf{x}^2$. The galaxies run away from each other.

Since the space stretches out, so does the wavelength of a photon; the photon experiences redshift. If the wavelength at emission (say, by distant star) is λ_e, then the wavelength we measure is

$$\lambda_0 = (1+z)\lambda_\mathrm{e} \ , \quad \text{where} \quad z = \frac{a(t_0)}{a(t_\mathrm{e})} - 1 \ .$$

Here t_e is the time at emission, and z is redshift. Hereafter we denote by subscript 0 the quantities measured at the present time. We sometimes set $a_0 \equiv a(t_0) = 1$ and put ourselves at the origin of coordinate frame, then $|\mathbf{x}|$ is the *present* distance to a point with coordinates $\mathbf{x}$. We also call this the *comoving distance*.

Clearly, the further from us is the source, the longer it takes for light, seen by us today, to travel, i.e., the larger $t_0 - t_\mathrm{e}$. High redshift sources are far away from us both in space and in time. For not so distant sources, we have $t_0 - t_\mathrm{e} = r$, where r is the physical distance to the source[1]. For $z \ll 1$ we thus have the Hubble law,

$$z = H_0 r \ . \tag{2}$$

$H_0 \equiv H(t_0)$ is the Hubble constant, i.e., the present value of the Hubble parameter

$$H(t) = \frac{\dot{a}(t)}{a(t)} \ .$$

The value of the Hubble constant is a subject of some controversy. While the redshift of an object can be measured with high precision (λ_e is the wavelength of a photon emitted by an excited atom; one identifies a series of emission lines, thus determining λ_e, and measures their actual wavelengths λ_0, both with spectroscopic precision; absorption lines are used as well), absolute distances to astrophysical sources have considerable systematic uncertainty. The precise value of H_0 will not be important for our semi-quantitative discussions; we quote here the value found by the Planck collaboration [3],

$$H_0 = (67.7 \pm 0.4) \ \frac{\text{km/s}}{\text{Mpc}} \approx (14.4 \cdot 10^9 \ \text{yrs})^{-1} \ . \tag{3}$$

Here Mpc is the length unit used in cosmology and astrophysics,

$$1 \ \text{Mpc} \approx 3 \cdot 10^6 \ \text{light years} \approx 3 \cdot 10^{24} \ \text{cm} \ .$$

The funny unit used in the first expression in Eq. (3) has to do with (somewhat misleading) interpretation of redshift as Doppler effect: galaxies run away from us at velocity $v = z$. To account for uncertainties in H_0 one writes for the present value of the Hubble parameter

$$H_0 = h \cdot 100 \ \frac{\text{km/s}}{\text{Mpc}} \ . \tag{4}$$

Thus $h \approx 0.7$. We will use this value in estimates.

Concerning length scales characteristic of various objects, we quote the following:

- sizes of visible parts of dwarf galaxies are of order 1 kpc and even smaller;
- sizes of visible parts of galaxies like ours are of order 10 kpc;
- dark halos of galaxies extend to distances of order 100 kpc and larger;
- clusters of galaxies have sizes of order $1 - 3$ Mpc;
- the homogeneity scale[2] today is of order 200 Mpc;
- the size of the visible Universe is 14 Gpc.

2.2 Cosmic microwave background

One of the fundamental discoveries of 1960's was cosmic microwave background (CMB). These are photons with black-body spectrum of temperature

$$T_0 = 2.7255 \pm 0.0006 \ \text{K} \ . \tag{5}$$

Measurements of this spectrum are quite precise and show no deviation from the Planck spectrum (although some deviations are predicted, see Ref. [4] for review). The energy density of CMB photons is given by the Stefan–Boltzmann formula

$$\rho_{\gamma,0} = \frac{\pi^2}{15} T_0^4 = 2.7 \cdot 10^{-10} \ \frac{\text{GeV}}{\text{cm}^3} \ . \tag{6}$$

[1] Hereafter we use the natural units, with the speed of light, Planck and Boltzmann constants equal to 1, $c = \hbar = k_\mathrm{B} = 1$. Then Newton's gravity constant is $G = M_\mathrm{Pl}^{-2}$, where $M_\mathrm{Pl} = 1.2 \cdot 10^{19}$ GeV is the Planck mass.

[2] Regions of this size and larger look all the same, while smaller regions differ from each other; they contain different numbers of galaxies.

while the number density of CMB photons is $n_{\gamma,0} = 410 \ \ \text{cm}^{-3}$.

The discovery of CMB has shown that the Universe was hot at early times, and cooled down due to expansion. As we pointed out, the wavelength of a photon increases in time as $a(t)$, so the energies and hence temperature of photons scale as

$$\omega(t) \propto a^{-1}(t) \ , \qquad T(t) = \frac{a_0}{a(t)} T_0 = (1+z) T_0 \ .$$

Importantly, the energy density of CMB photons scales as

$$\rho_\gamma \propto T^4 \propto a^{-4} \ .$$

This is in contrast with the scaling of energy density (mass density) of non-relativistic particles (baryons, dark matter)

$$\rho_{\text{M}} \propto a^{-3} \ ,$$

which is obtained by simply noting that the mass in comoving volume remains constant.

2.3 Friedmann equation

The expansion of the spatially flat Universe is governed by the Friedmann equation,

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3M_{\text{Pl}}^2} \rho \ , \tag{7}$$

where ρ is the *total* energy density in the Universe. This is nothing but the (00)-component of the Einstein equations of General Relativity, $R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi T_{\mu\nu}$, specified to spatially flat FLRW metric and homogeneous and isotropic matter.

One conventionally defines the parameter (critical density),

$$\rho_{\text{c}} = \frac{3}{8\pi} M_{\text{Pl}}^2 H_0^2 \approx 5 \cdot 10^{-6} \ \frac{\text{GeV}}{\text{cm}^3} \ . \tag{8}$$

It is equal to the sum of all forms of energy density in the *present* Universe.

2.4 Present composition of the Universe

The *present* composition of the Universe is characterized by the parameters

$$\Omega_\lambda = \frac{\rho_{\lambda,0}}{\rho_{\text{c}}} \ .$$

where λ labels various forms of energy: relativistic matter ($\lambda = rad$), non-relativistic matter ($\lambda = M$), dark energy ($\lambda = \Lambda$). Clearly, Eq. (7) gives

$$\sum_\lambda \Omega_\lambda = 1 \ .$$

Let us quote the numerical values:

$$\Omega_{\text{rad}} = 8.6 \cdot 10^{-5} \ , \tag{9a}$$
$$\Omega_{\text{M}} = 0.31 \ , \tag{9b}$$
$$\Omega_\Lambda = 0.69 \ . \tag{9c}$$

The value in Eq. (9a) is calculated for the unrealistic case in which *all neutrinos are relativistic today*, so the radiation component even at present consists of CMB photons and three neutrino species. This prescription is convenient for studying the early Universe, since the energy density of relativistic neutrinos scales in the same way as that of photons,

$$\rho_\nu \propto T^4 \propto a^{-4} \ ,$$

and at temperatures above neutrino masses (but below 1 MeV) we have

$$\rho_\nu == \Omega_\nu \rho_{\mathrm{c}} \left(\frac{a_0}{a}\right)^4 \ .$$

Non-relativistic matter consists of baryons and dark matter. Their contributions are [3]

$$\Omega_{\mathrm{B}} = 0.049 \ , \tag{10a}$$

$$\Omega_{\mathrm{DM}} = 0.26 \ . \tag{10b}$$

As we pointed out above, energy densities of various species evolve as follows:

– radiation (photons and neutrinos at temperatures above neutrino mass):

$$\rho_{\mathrm{rad}}(t) = \left(\frac{a_0}{a(t)}\right)^4 \rho_{\mathrm{rad},0} = (1+z)^4 \, \Omega_{\mathrm{rad}} \rho_{\mathrm{c}} \ . \tag{11}$$

– Non-relativistic matter:

$$\rho_{\mathrm{M}}(t) = \left(\frac{a_0}{a(t)}\right)^3 \rho_{\mathrm{M},0} = (1+z)^3 \, \Omega_{\mathrm{M}} \rho_{\mathrm{c}} \ . \tag{12}$$

– The dark energy density does not change in time, or changes very slowly. In what follows we take it constant in time,

$$\rho_\Lambda = \Omega_\Lambda \rho_c = \text{const} \ . \tag{13}$$

This assumption is not at all innocent. It means that dark energy is assumed to be a cosmological constant. However, even slight dependence of ρ_Λ on time would mean that we are dealing with something different from the cosmological constant. In that case the dark energy density would be associated with some field; there are various theoretical proposals concerning the properties of such a field. Present data is consistent with time-independent ρ_Λ, but the precision of this statement is not yet very high. It is extremely important to study the time-(in)dependence of ρ_Λ with high precision; several experiments are aimed at that.

2.5 Cosmological epochs

The Friedmann equation (7) is now written as

$$\begin{aligned} H^2(t) &= \frac{8\pi}{3M_{\mathrm{Pl}}^2}[\rho_\Lambda + \rho_{\mathrm{M}}(t) + \rho_{\mathrm{rad}}(t)] \\ &= H_0^2 \left[\Omega_\Lambda + \Omega_{\mathrm{M}} \left(\frac{a_0}{a(t)}\right)^3 + \Omega_{\mathrm{rad}} \left(\frac{a_0}{a(t)}\right)^4\right] \end{aligned}$$

This shows that the dominant term in the right hand side at early times (small $a(t)$) was ρ_{rad}, i.e., the expansion was dominated by ultrarelativistic particles (radiation). This is radiation domination epoch.

Then the term $\rho_{\rm M}$ took over, and matter dominated epoch began. The redshift at radiation–matter equality, when the energy densities of radiation and matter were equal, is

$$1 + z_{\rm eq} = \frac{a_0}{a(t_{\rm eq})} = \frac{\Omega_{\rm M}}{\Omega_{\rm rad}} \approx 3500 \; ,$$

and using the Friedmann equation one finds the age of the Universe at equality

$$t_{\rm eq} \approx 50\,000 \text{ years} \; .$$

The present Universe is at the end of the transition from matter domination to Λ-domination: the dark energy density ρ_Λ will completely dominate over non-relativistic matter in future.

So, we have the following sequence of the regimes of evolution:

$$\cdots \Longrightarrow \text{Radiation domination} \Longrightarrow \text{Matter domination} \Longrightarrow \Lambda\text{–domination} \; . \tag{14}$$

Dots here denote some cosmological epoch preceding the hot stage. We discuss this point later on.

2.6 Radiation domination

2.6.1 Expansion law

The evolution of the scale factor at radiation domination is obtained by using $\rho_{\rm rad} \propto a^{-4}$ in the Friedmann equation (7):

$$\frac{\dot a}{a} = \frac{\text{const}}{a^2} \; .$$

This gives

$$a(t) = \text{const} \cdot \sqrt{t} \; . \tag{15}$$

The constant here does not have physical significance, as one can re-scale the coordinates $\mathbf{x}$ at one moment of time, thus changing the normalization of a.

There are several properties that immediately follow from the result Eq. (15). First, the expansion *decelerates*:

$$\ddot a < 0 \; .$$

Second, time $t = 0$ is the Big Bang singularity (assuming, for the sake of argument, that the Universe starts right from radiation domination epoch). The expansion rate

$$H(t) = \frac{1}{2t}$$

diverges as $t \to 0$, and so does the energy density $\rho(t) \propto H^2(t)$ and temperature $T \propto \rho^{1/4}$. This is "classical" singularity (singularity in classical General Relativity) which, one expects, is resolved in one or another way in complete quantum gravity theory. One usually assumes (although this is not necessarily correct) that the classical expansion begins just after the Planck epoch, when $\rho \sim M_{\rm Pl}^4$, $H \sim M_{\rm Pl}$, etc.

2.6.2 Particle horizon

The third observation has to do with the causal structure of space-time in the Hot Big Bang Theory (theory that assumes that the evolution starts from the singularity directly into radiation domination—no dots in Eq. (14)). Consider signals emitted right after the Big Bang singularity and travelling at the speed of light. The light cone obeys $ds = 0$, and hence $a(t)dx = dt$. So, the coordinate distance that a signal travels from the Big Bang to time t is

$$x = \int_0^t \frac{dt}{a(t)} \equiv \eta \; . \tag{16}$$

In the radiation dominated Universe

$$\eta = \text{const} \cdot \sqrt{t} \ .$$

The physical distance from the emission point to the position of the signal is

$$l_{\rm H}(t) = a(t)x = a(t) \int_0^t \frac{dt}{a(t)} \ . \tag{17}$$

This physical distance is finite; it is the size of a causally connected region at time t. It is called the horizon size (more precisely, the size of particle horizon). In other words, an observer at time t can have information only on the part of the Universe whose physical size at that time is $l_{\rm H}(t)$. At radiation domination, one has

$$l_{\rm H}(t) = 2t \ .$$

Note that this horizon size is of the order of the Hubble size,

$$l_{\rm H}(t) \sim H^{-1}(t) \ . \tag{18}$$

The notion of horizon is straightforwardly extended to the matter dominated epoch and to the present time: relation Eq. (17) is of general nature, while the scale factor $a(t)$ has to be calculated anew. To give an idea of numbers, the horizon size at the present epoch is

$$l_{\rm H}(t_0) \approx 14 \text{ Gpc} \simeq 4 \cdot 10^{28} \text{ cm} \ .$$

2.6.3 Energy density

At radiation domination, cosmic plasma is almost always in thermal equilibrium, and interactions between particles are almost always weak. So, the plasma properties are determined by thermodynamics of a gas of free relativistic particles. At different times, the number of relativistic species that contribute to the energy density is different. As an example, at temperatures above 1 MeV, but below 100 MeV, relativistic are photons, three types of neutrinos, electrons and positrons; while at temperatures of about 200 GeV all Standard Model particles are relativistic. In most cases, one can neglect chemical potentials, i.e., consider cosmic plasma symmetric under interchange of particles with antiparticles (chemical potential of photons is zero, since photons can be created in processes like $\rm e^-e^- \to e^-e^-\gamma$; since particle and its antiparticle can annihilate into photons, e.g., $\rm e^+e^- \to \gamma\gamma$, chemical potentials of particles and antiparticles are equal in modulus and opposite in sign, e.g., $\mu_{\rm e^+} = -\mu_{\rm e^-}$; in symmetric plasma $\mu_{\rm e^+} = -\mu_{\rm e^-} = 0$). Then the Stefan–Boltzmann law gives for the energy density

$$\rho_{\rm rad} = \frac{\pi^2}{30} g_* T^4 \ , \tag{19}$$

where g_* is the effective number of degrees of freedom,

$$g_* = \sum_{\rm bosons} g_i + \frac{7}{8} \sum_{\rm fermions} g_i \ ,$$

g_i is the number of spin states of a particle i, the factor 7/8 is due to Fermi-statistics. The parameter g_* depends on temperature, and hence on time: as the temperature decreases below the mass of a particle, this particle drops out from the sum here. The formula (19) enables one to write the Friedmann equation (7) as

$$H = \frac{T^2}{M_{\rm Pl}^*} \ , \qquad M_{\rm Pl}^* = \frac{M_{\rm Pl}}{1.66\sqrt{g_*}} \ . \tag{20}$$

We use this simple result in what follows.

2.6.4 Entropy

The cosmological expansion is slow, which implies conservation of entropy (modulo quite exotic scenarios with large entropy generation). The entropy density of a free relativistic gas in thermal equilibrium is given by

$$s = \frac{2\pi^2}{45} g_* T^3 \ .$$

The conservation of entropy means that the entropy density scales *exactly* as a^{-3},

$$sa^3 = \text{const} \ , \tag{21}$$

while temperature scales *approximately* as a^{-1} (this is because g_* depends on time). We note for future reference that the effective number of degrees of freedom in the Standard Model at $T \gtrsim 100$ GeV is

$$g_*(100 \text{ GeV}) \approx 100 \ .$$

The present entropy density in the Universe, still with the prescription that neutrinos are relativistic, is

$$s_0 \approx 3000 \text{ cm}^{-3} \ . \tag{22}$$

The precise meaning of this number is that at high temperatures (when there is thermal equilibrium), the entropy density is $s(t) = (a_0/a(t))^3 s_0$.

Notion of entropy is convenient, in particular, for characterizing asymmetries which can exist if there are conserved quantum numbers, such as the baryon number after baryogenesis. The density of a conserved number also scales as a^{-3}, so the time independent characteristic of, say, the baryon abundance is the baryon-to-entropy ratio

$$\Delta_\text{B} = \frac{n_\text{B}}{s} \ .$$

At late times, one can use another parameter, baryon-to-photon ratio

$$\eta_\text{B} = \frac{n_\text{B}}{n_\gamma} \ , \tag{23}$$

where n_γ is photon number density. It is related to Δ_B by a numerical factor, but this factor depends on time through g_* and stays constant only after e^+e^--annihilation, i.e., at $T \lesssim 0.5$ MeV. Numerically,

$$\Delta_\text{B} = 0.14 \, \eta_{\text{B},0} = 0.86 \cdot 10^{-10} \ . \tag{24}$$

In what follows we discuss the ways to obtain this number from observations.

2.7 Matter domination

At matter domination, we have $\rho \propto a^{-3}$, and the Friedmann equation (7) gives

$$a(t) = \text{const} \cdot t^{2/3}$$

Qualitatively, matter domination is similar to radiation domination: expansion is decelerated, the size of the particle horizon is of order of the Hubble size, $l_\text{H}(t) \sim H^{-1}(t) \sim t$. An important difference between radiation and matter dominated epochs is that inhomogeneities in energy density ("scalar perturbations") grow rapidly at matter domination and slowly at radiation domination. Thus, matter domination is the epoch of structure formation in the Universe.

2.8 Dark energy domination

The expansion of the Universe is accelerated today. Within General Relativity this is attributed to dark energy. We know very little about this "substance": we know its energy density, Eq. (9c), and also know that this energy density changes in time very slowly, if at all. The latter fact is quantified in the following way. Let us denote by p the effective pressure, i.e., spatial component of the energy-momentum tensor in locally-Lorentz frame $T_{\mu\nu} = \text{diag}(\rho, p, p, p)$. Then covariant conservation of the energy-momentum in an expanding Universe gives for any fraction that does not interact with other fractions

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + p)$$

(note that relativistic and non-relativistic matter have $p = \rho/3$ and $p = 0$, respectively, so this equation gives for them $\rho \propto a^{-4}$ and $\rho \propto a^{-3}$, as it should). A simple parametrization of time-dependent dark energy is $p_\Lambda = w_\Lambda \rho_\Lambda$ with time-independent w_Λ. The combination of cosmological data gives [3]

$$w_\Lambda \approx -1.03 \pm 0.03 \; . \tag{25}$$

Thus, with reasonable precision one has $p_\Lambda = -\rho_\Lambda$, which corresponds to a time-independent dark energy density.

The solution to the Friedmann equation (7) with constant $\rho = \rho_\Lambda$ is

$$a(t) = \mathrm{e}^{\mathrm{H_\Lambda t}} \; ,$$

where $H_\Lambda = (8\pi\rho_\Lambda/3M_{\text{Pl}}^2)^{1/2} = \text{const}$. This gives an accelerated expansion, $\ddot{a} > 0$, unlike at radiation or matter domination. The transition from decelerated (matter dominated) to accelerated expansion (dark energy dominated) has been confirmed quite some time ago by combined observational data, see Fig. 1, which shows the dependence on redshift of the quantity $H(z)/(1+z) = \dot{a}(t)/a_0$.

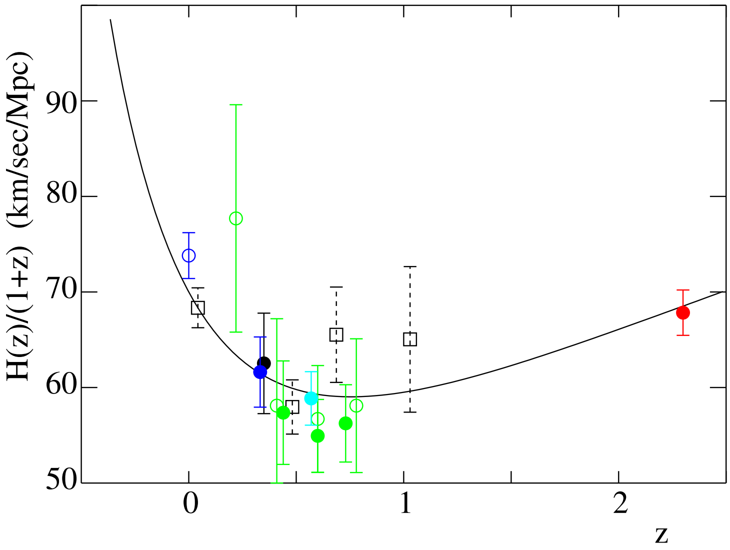


Fig. 1: Observational data on the time derivative of the scale factor as function of redshift z [5]. The change of the behavior from decreasing to increasing, as z decreases, means the change from decelerated to accelerated expansion. The theoretical curve corresponds to spatially flat Universe with $h = 0.7$ and $\Omega_\Lambda = 0.73$.

In the case of the cosmological constant, the energy-momentum tensor is proportional to the metric, and in a locally-Lorentz frame it reads

$$T_{\mu\nu} = \rho_\Lambda \eta_{\mu\nu} \; ,$$

where $\eta_{\mu\nu}$ is the Minkowski tensor. Hence $w_\Lambda = -1$. One can view this as the characteristic of the vacuum, whose energy-momentum tensor must be Lorentz-covariant. As we pointed out above, any deviation from $w = -1$ would mean that we are dealing with something other than vacuum energy density.

The problem with dark energy is that its present value is extremely small by particle physics standards,

$$\rho_{\mathrm{DE}} \approx 4\ \mathrm{GeV/m}^3 = (2 \times 10^{-3}\mathrm{eV})^4 \,.$$

In fact, there are two hard problems. One is that the dark energy density is zero to an excellent approximation. Another is that it is non-zero nevertheless, and one has to understand its energy scale. We are not going to discuss these points anymore, and only emphasize that we are not aware of a compelling mechanism that solves any of the two cosmological constant problems (with possible exception of anthropic argument by Weinberg and Linde [6, 7]).

3 Cornerstones of thermal history

3.1 Recombination = photon last scattering

Going back in time, we reach so high temperatures that the usual matter (electrons and protons with rather small admixture of light nuclei, mainly ^{4}He) is in the plasma phase. In plasma, photons interact with electrons due to the Thomson scattering and protons have Coulomb interaction with electrons. These interactions are strong enough to keep photons, electrons and protons in thermal equilibrium. When the temperature drops to

$$T_{\mathrm{rec}} \approx 3000\ \mathrm{K}\,, \qquad z_{\mathrm{rec}} \approx 1090\,,$$

almost all electrons "recombine" with protons into neutral hydrogen atoms (helium recombined earlier). The number density of atoms at that time is quite small, 250 cm^{-3}, so from that time on, the Universe is transparent to photons[3]. Thus, T_{rec} is the *photon last scattering temperature*. At that time the age of the Universe is $t_{\mathrm{rec}} \approx 380$ thousand years (for comparison, its present age is about 13.8 billion years).

CMB photons give us (literally!) the photographic picture of the Universe at the photon last scattering epoch. The last scattering epoch lasted considerably shorter than the then Hubble time $H^{-1}(t_{\mathrm{rec}}) \sim t_{\mathrm{rec}}$; to a meaningful (although rather crude) approximation, recombination occurred instantaneously. This is important, since in the opposite case of long recombination, the photographic picture would be strongly washed out.

This photographic picture is shown in Fig. 2. Here brighter (darker) regions correspond to higher (lower) temperatures. The relative temperature fluctuation is of the order of $\delta T/T = 10^{-4} - 10^{-5}$, so the 380 thousand year old Universe was much more homogeneous than today.

One performs Fourier decomposition of the temperature fluctuations, i.e., decomposition in spherical harmonics:

$$\frac{\delta T}{T}(\theta,\varphi) = \sum_{l,m} a_{lm} Y_{lm}(\theta,\varphi)\,.$$

Here a_{lm} are independent Gaussian random variables (no non-Gaussianities have been found so far) with $\langle a_{lm} a^*_{l'm'}\rangle \propto \delta_{ll'}\delta_{mm'}$ and $\langle a^*_{lm} a_{lm}\rangle = C_l$. The multipoles C_l, or, equivalently,

$$D_l = \frac{l(l+1)}{2\pi} C_l$$

are the main quantities of interest. The larger l, the smaller the angular scales, hence the shorter the wavelengths of density perturbations producing the temperature anisotropy.

[3] Modulo effects of re-ionization that occurred much later and affected a small fraction of CMB photons.

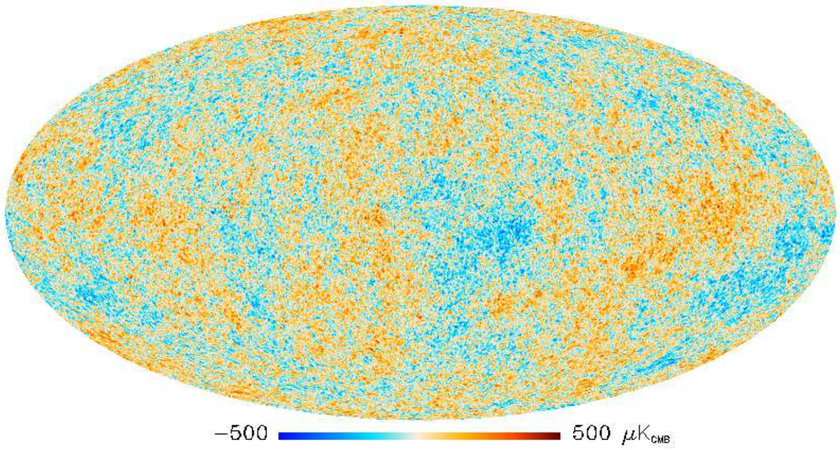


Fig. 2: CMB sky as seen by Planck.

It is worth noting that averaging here is understood in terms of an ensemble of Universes, while we have just one Universe. So, there is inevitable uncertainty in C_l, called cosmic variance. For given l, one has $(2l+1)$ quantities a_{lm}, $m = 0, \pm 1, \ldots, \pm l$, so the uncertainty is $\Delta C_l / C_l \simeq 1/\sqrt{2l}$.

CMB temperature multipoles are shown in Fig. 3 (error bars there are due to cosmic variance, not the measurement errors). Also measured are CMB polarization multipoles and temperature-polarization cross-correlation multipoles. There is a lot of physics behind these quantities, which has to do with:

- primordial perturbations: the perturbations that are built in already at the beginning of the hot cosmological epoch, see Section 11;
- development of sound waves in the cosmic plasma from the early hot stage to recombination; gravitational potentials due to dark matter at recombination (which are sensitive to the composition of cosmic medium);
- propagation of photons after recombination (which is sensitive to expansion history of the Universe and structure formation).

Clearly, CMB measurements are a major source of the cosmological information. We come back to CMB in due course.

3.2 Big Bang nucleosynthesis

As we go back further in time, we arrive at a temperature in the Universe in the MeV range. The epoch characterized by temperatures 1 MeV–30 keV is the epoch of Big Bang nucleosynthesis. That epoch starts at a temperature of 1 MeV, when the age of the Universe is 1 s. At temperatures above 1 MeV, there are rapid weak processes like

$$\mathrm{e}^- + \mathrm{p} \longleftrightarrow \mathrm{n} + \nu_e \; . \tag{26}$$

These processes keep neutrons and protons in chemical equilibrium; the ratio of their number densities is determined by the Boltzmann factor, $n_{\mathrm{n}}/n_{\mathrm{p}} = \exp\left(-\frac{m_{\mathrm{n}} - m_{\mathrm{p}}}{T}\right)$. At $T_{\mathrm{n}} \approx 1$ MeV neutron-proton transitions in Eq. (26) switch off, and neutron-proton ratio is frozen out at the value

$$\frac{n_{\mathrm{e}}}{n_{\mathrm{p}}} = \mathrm{e}^{-\frac{m_{\mathrm{n}} - m_{\mathrm{p}}}{T_{\mathrm{n}}}} \; .$$

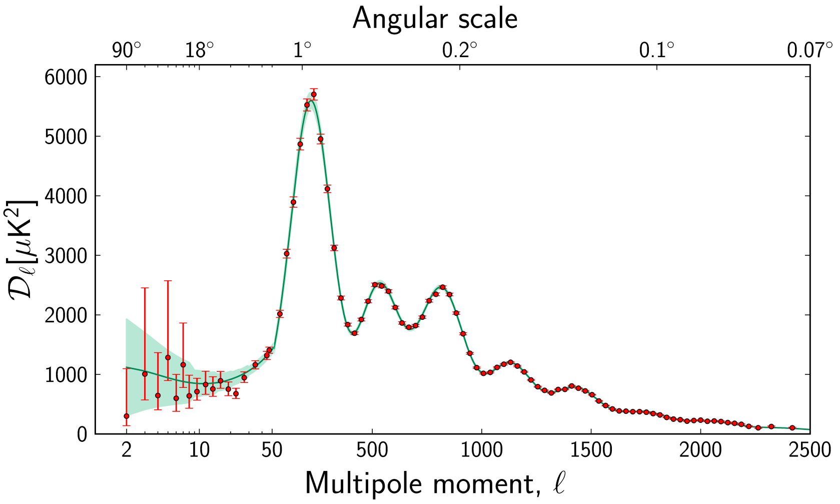


Fig. 3: Multipoles D_l as measured by Planck.

Interestingly, $m_{\rm n} - m_{\rm p} \sim T_{\rm n}$, so the neutron-proton ratio at neutron freeze-out and later was neither equal to 1, nor very small. If it were equal to 1, protons would in the end combine with neutrons into ^{4}He, and there would remain no hydrogen in the Universe. On the other hand, for very small $n_{\rm n}/n_{\rm p}$, too few light nuclei would be formed, and we would not have any observable remnants of the Big Bang nucleosynthesis (BBN) epoch. In either case the Universe would be quite different from what it actually is. It is worth noting that the approximate relation $m_{\rm n} - m_{\rm p} \sim T_{\rm n}$ is a coincidence: $m_{\rm n} - m_{\rm p}$ is determined by light quark masses and electromagnetic coupling, while $T_{\rm n}$ is determined by the strength of weak interactions (the rates of the processes in Eq. (26)) and gravity (the expansion of the Universe). This is one of numerous coincidences we encounter in cosmology.

At temperatures 100–30 keV, neutrons combined with protons into light elements in thermonuclear reactions

$$\begin{array}{rcl} p+n & \to & {}^2{\rm H}+\gamma\ , \\ {}^2\ {\rm H}+p & \to & {}^3\ {\rm He}+\gamma\ , \\ {}^3{\rm He}+{}^2\ {\rm H} & \to & {}^4\ {\rm He}+p\ , \end{array} \qquad (27)$$

etc., up to ^{7}Li. The abundances of light elements have been measured, see Fig. 4. The only parameter relevant for calculating these abundances (assuming negligible neutrino-antineutrino asymmetry) is the baryon-to-photon ratio $\eta_{\rm B} \equiv \eta$, see Eq. (23), which determines the number density of baryons. Comparison of the Big Bang nucleosynthesis theory with the observational determination of the composition of cosmic medium enables one to determine $\eta_{\rm B}$ and check the overall consistency of the BBN picture. It is even more reassuring that a completely independent measurement of $\eta_{\rm B}$ that makes use of the CMB temperature fluctuations is in excellent agreement with BBN. Thus, BBN gives us confidence that we understand the Universe at $T \sim 1$ MeV, $t \sim 1$ s. In particular, we are convinced that the cosmological expansion was governed by general relativity.

3.3 Neutrino decoupling

Another class of processes of interest at temperatures in the MeV range is neutrino production, annihilation and scattering,

$$\nu_\alpha + \bar{\nu}_\alpha \longleftrightarrow {\rm e}^+ + {\rm e}^-$$

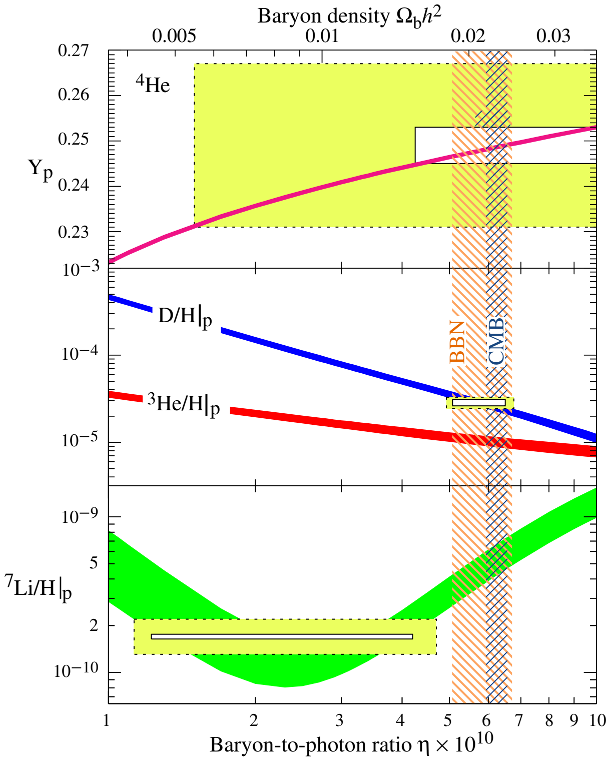


Fig. 4: Abundances of light elements, measured (boxes; larger boxes include systematic uncertainties) and calculated as functions of baryon-to-photon ratio η [8]. The determination of $\eta \equiv \eta_B$ from BBN (vertical range marked BBN) is in excellent agreement with the determination from the analysis of CMB temperature fluctuations (vertical range marked CMB).

and crossing processes. Here the subscript α labels neutrino flavors. These processes switch off at $T \sim$ 2–3 MeV, depending on the neutrino flavor. Since then neutrinos do not interact with the cosmic medium other than gravitationally, but they do affect the properties of the CMB and the distribution of galaxies through their gravitational interactions. Thus, observational data can be used to establish, albeit somewhat indirectly, the existence of relic neutrinos and set limits on neutrino masses. We quote here the limit reported by the Planck collaboration [3]

$$\sum m_\nu < 0.12 \text{ eV} ,$$

where the sum runs over the three neutrino species. Other analyses give somewhat weaker limits. Also, the data can be used to determine the effective number of neutrino species that counts the number of relativistic degrees of freedom [3]:

$$N_{\nu,\text{eff}} = 2.99 \pm 0.17 ,$$

which is consistent with the Standard Model value $N_\nu = 3$. We see that cosmology *requires* relic neutrinos.

4 Dark matter: evidence

Unlike dark energy, dark matter experiences the same gravitational force as the baryonic matter. Dark matter is discussed in numerous reviews, see, e.g., Refs. [9–12]. It consists presumably of new stable massive particles. These make clumps of mass which constitute most of the mass of galaxies and clusters of galaxies. Dark matter is characterized by the mass-to-entropy ratio,

$$\left(\frac{\rho_{\rm DM}}{s}\right)_0 = \frac{\Omega_{\rm DM}\rho_c}{s_0} \approx \frac{0.26 \cdot 5 \cdot 10^{-6}\ {\rm GeV} \cdot {\rm cm}^{-3}}{3000\ {\rm cm}^{-3}} = 4 \cdot 10^{-10}\ {\rm GeV}\ . \qquad (28)$$

This ratio is constant in time since the freeze-out of the dark matter density: both, the number density of dark matter particles $n_{\rm DM}$ (and hence their mass density $\rho_{\rm DM} = m_{\rm DM} n_{\rm DM}$) and the entropy density, decrease exactly as a^{-3}.

There are various ways of measuring the contribution of non-baryonic dark matter to the total energy density of various objects and the Universe as a whole.

4.1 Dark matter in galaxies

Dark matter exists in galaxies. Its distribution is measured by the observation of rotation velocities of distant stars and gas clouds around a galaxy, Fig. 5. If the mass was concentrated in a luminous central part of a galaxy, the velocities of objects away from the central part would decrease with the distance r to the center as $v \propto r^{-1/2}$—this immediately follows from Newton's second law

$$\frac{v^2}{r} = G\frac{M(r)}{r^2}\ .$$

In reality, rotation curves are typically flat up to distances exceeding the size of the bright part by a factor of 10 or so. The fact that dark matter halos are so large is explained by the defining property of dark matter particles: they do not lose their energies by emitting photons, and, in general, interact with conventional matter very weakly.

4.2 Dark matter in clusters of galaxies

Dark matter makes most of the mass of the largest gravitationally bound objects—clusters of galaxies. There are various methods to determine the gravitating mass of a cluster, and mass distribution in a

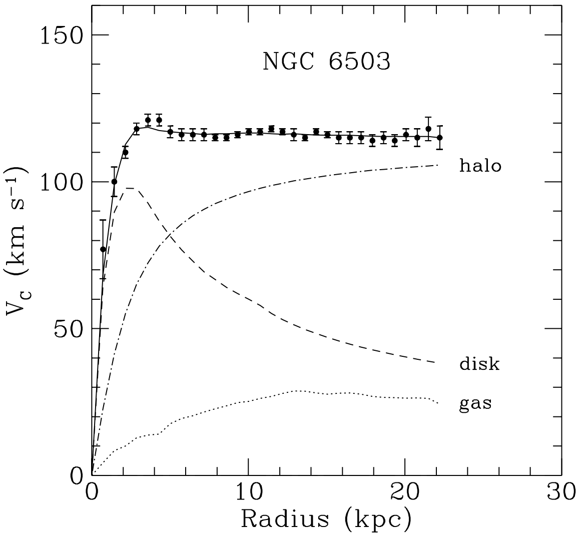


Fig. 5: Rotation velocities of hydrogen gas clouds around a galaxy NGC 6503 [13]. Lines show the contributions of the three main components that produce the gravitational potential. The main contribution at large distances is due to dark matter, labeled "halo".

cluster, which give consistent results. These include measurements of rotational velocities of galaxies in a cluster (original Zwicky argument that goes back to 1930's), measurements of temperature of hot gas (which actually makes most of the baryonic matter in clusters), observations of gravitational lensing of extended light sources (galaxies) behind the cluster, see Fig. 6. All these determinations show that baryons (independently measured through their X-ray emission) make less than 1/4 of total mass in clusters. The rest is dark matter.

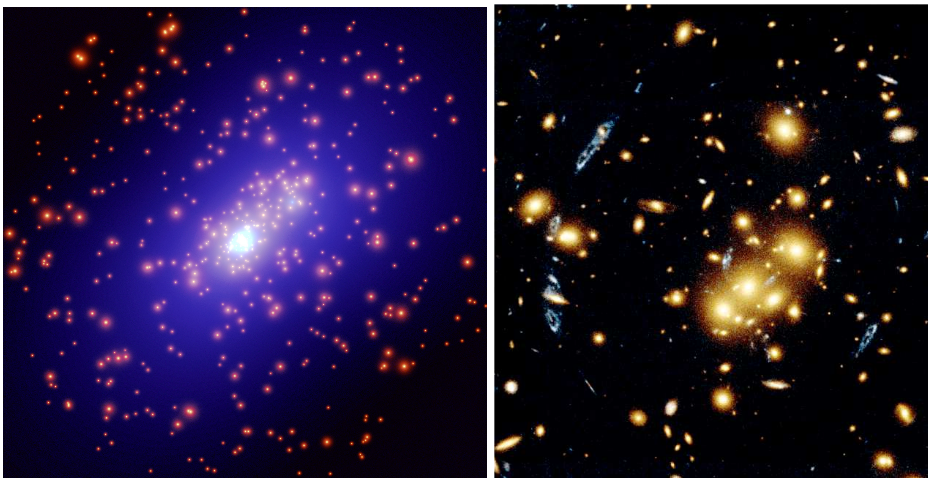


Fig. 6: Cluster of galaxies CL0024 + 1654 [14], acting as gravitational lens. Right panel: cluster in visible light. Round yellow spots are galaxies in the cluster. Elongated blue strips are images of one and the same galaxy behind the cluster. Left panel: reconstructed distribution of gravitating mass in the cluster; brighter regions have larger mass density.

Concerning galaxies and clusters of galaxies, we note that there are attempts to attribute the properties of rotation curves and other phenomena, which are usually considered as evidence for dark matter, to a modification of gravity, and in this way get rid of dark matter altogether. There are several strong arguments that rule out this idea. One argument has to do with the Bullet Cluster, Fig. 7. Shown are two galaxy clusters that passed through each other. The dark matter and galaxies do not experience friction and thus do not lose their velocities. On the contrary baryons in hot, X-ray emitting gas do experience friction and hence get slowed down and lag behind the dark matter and galaxies. In this way the baryons (which are mainly in hot gas) and dark matter are separated in space. Since the baryonic mass and gravitational potentials are not concentric, one cannot attribute gravitational potentials solely to baryons, even assuming the modification of Newton's gravity law. As a remark, the fact that dark matter moves after cluster a collision considerably faster than baryonic gas means that elastic scattering between dark matter particles is weak. Quantitatively, the limit on the dark matter elastic scattering cross section is

$$\sigma^{(\mathrm{el})}_{\mathrm{DM-DM}} < 1 \cdot 10^{-24}\ \mathrm{cm}^2\ . \tag{29}$$

This limit is not particularly strong, but it does rule out part of the parameter space of strongly interacting massive particle (SIMP) dark matter models, see Section 5.2.

4.3 Dark matter imprint in CMB

The composition of the Universe strongly affects the CMB angular anisotropy and polarization. Before recombination, the energy density perturbation is a sum of the perturbation in the baryon-electron-photon component and the dark matter component,

$$\delta\rho = \delta\rho_{\mathrm{B}} + \delta\rho_{\mathrm{DM}}$$

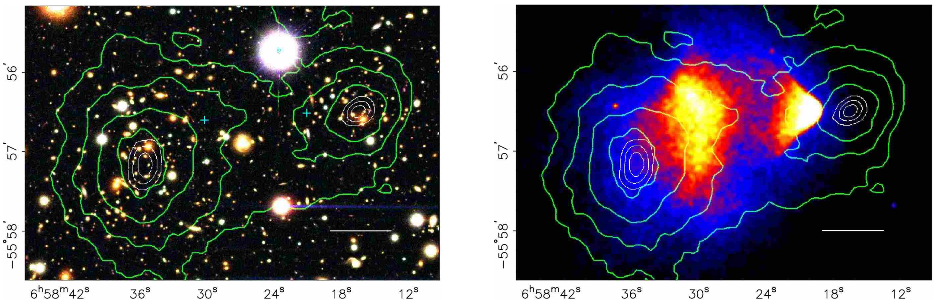


Fig. 7: Observation [15] of the Bulet Cluster 1E0657–558 at $z = 0.296$. Closed lines show the gravitational potential produced mainly by dark matter and measured through gravitational lensing. Bright regions in the right panel show X-ray emission of hot baryon gas, which makes most of the baryonic matter in the clusters. The length of the white interval is 200 kpc in the comoving frame.

(we simplify things here, as there is also perturbation in the gravitational potentials induced by the density perturbation). The tightly coupled baryon-electron-photon plasma has high pressure (due to the photon component with $p_\gamma = \rho_\gamma/3$), so the density perturbations in this fraction undergo acoustic oscillations: every Fourier mode oscillates in time as

$$\delta\rho_{\rm B}(\mathbf{k}, t) = A(\mathbf{k})\cos\left(\int_0^t v_s \frac{k}{a(t)} dt\right) , \tag{30}$$

where $\mathbf{k}$ is the comoving momentum (and $\mathbf{k}/a(t)$ is the physical momentum which gets redshifted), $v_s \approx 1/\sqrt{3}$ is the sound speed, and $A(\mathbf{k})$ is the amplitude that varies slowly with k (in statistical sense: $\delta\rho(\mathbf{k})$ is the Gaussian random field). We comment in Section 11 on the fact that the phase of cosine in Eq. (30) is well defined. On the contrary, dark matter is pressureless, so its perturbation is almost independent of time,

$$\delta\rho_{\rm DM} \approx \delta\rho_{\rm DM}(\mathbf{k}) ,$$

where $\delta\rho_{\rm DM}(\mathbf{k})$ slowly varies with k. At recombination time t_r, the energy density perturbation is a sum

$$\delta\rho(\mathbf{k}, t_r) = A(\mathbf{k})\cos\left(\int_0^{t_r} v_s \frac{k}{a(t)} dt\right) + \delta\rho_{\rm DM}(\mathbf{k}) . \tag{31}$$

The first term here oscillates *as a function of* k, while the second term is a smooth, non-oscillating function of k.

Now, the behavior of $\delta\rho(t_r)$ as function of the spatial momentum k translates into the behavior of the CMB temperature fluctuation δT as function of the multipole number l. δT at a given point in space at the recombination epoch is proportional to $\delta\rho$ (here we again simplify things, this time quite considerably). We see CMB coming from a *photon last scattering sphere*; a smaller angular scale in this photographic picture corresponds to a smaller spatial scale at the recombination epoch, hence a larger multipole l corresponds to a higher three-momentum k. Thus, the oscillatory formula (31) translates into the oscillatory behavior in Fig. 3. Both, the oscillatory part of the temperature angular spectrum (which is due to the first, baryonic term in Eq. (31)) and the smooth part (due to the second, dark matter term in Eq. (31)), are clearly visible in Fig. 3. The detailed analysis of this angular spectrum enables one to determine the baryon content and the dark matter content in the Universe, $\Omega_{\rm B}$ and $\Omega_{\rm DM}$ quoted in Eq. (10).

4.4 Dark matter and structure formation

Dark matter is crucial for our existence, for the following reason: As we discussed above, the density perturbations in the baryon-electron-photon plasma before recombination do not grow because of high pressure; instead, they oscillate with a time-independent amplitudes. Hence, in a Universe without dark matter, density perturbations in the baryonic component would start to grow only after baryons decouple from photons, i.e., after recombination. The mechanism of the growth is qualitatively simple: an overdense region gravitationally attracts surrounding matter; this matter falls into the overdense region, and the density contrast increases. In the expanding, matter dominated Universe this gravitational instability results in the density contrast growing like $(\delta\rho/\rho)(t) \propto a(t)$. Hence, in a Universe without dark matter, the growth factor for baryon density perturbations would be at most[4]

$$\frac{a(t_0)}{a(t_{\rm rec})} = 1 + z_{\rm rec} = \frac{T_{\rm rec}}{T_0} \approx 10^3 \; . \qquad (32)$$

The initial amplitude of density perturbations is very well known from the CMB anisotropy measurements, $(\delta\rho/\rho)_i = 5 \cdot 10^{-5}$. Hence, a Universe without dark matter would still be nearly homogeneous: the density contrast would be in the range of a few per cent. No structure would have been formed, no galaxies, no life. No structure would be formed in future either, as the accelerated expansion due to dark energy will soon terminate the growth of perturbations.

Since dark matter particles decoupled from the plasma much earlier than baryons, the perturbations in dark matter started to grow much earlier. The corresponding growth factor is larger than Eq. (32), so that the dark matter density contrast at galactic and sub-galactic scales becomes of order one, the perturbations enter a non-linear regime, collapse and form dense dark matter clumps at $z = 5 - 10$. Baryons fall into potential wells formed by dark matter, so dark matter and baryon perturbations work together soon after recombination. Galaxies get formed in the regions where dark matter was overdense originally. For this picture to hold, dark matter particles must be non-relativistic early enough, as relativistic particles fly through gravitational wells instead of being trapped there. This means, in particular, that neutrinos cannot constitute a considerable part of dark matter.

4.5 Digression – Standard ruler: BAO

Before recombination, the sound speed in the baryon-electron-photon component is about $v_s \approx 1/\sqrt{3}$. After recombination, baryons (atoms) decouple from photons, the sound speed in the baryon component is practically zero, and baryons no longer move in space. This leads to a feature in the spatial distribution of matter (galaxies) which is known as Baryon Acoustic Oscillations (BAO). It is worth noting that similar phenomenon was described by A.D. Sakharov [16] back in 1965, but in the context of a cold cosmological model (Sakharov's paper was written before the discovery of CMB).

The physics behind BAO is illustrated in Fig. 8. Suppose there is an overdense region in the very early Universe (in the beginning of the hot epoch). Importantly, the initial conditions for the baryon-electron-photon component and dark matter are the same: overdensity exists in both of them in the same place in space (this is the property of adiabatic scalar perturbations; CMB measurements ensure that primordial perturbations are indeed adiabatic). This initial condition is shown in the left panel of Fig. 8. Before recombination, dark matter perturbation stays in the same place, while the perturbation in baryon-electron-photon component moves away with the sound speed. If the initial perturbation is spherically symmetric, then the sound wave is spherical, as shown in the right panel. At recombination, the baryon perturbation is frozen in, and the whole picture expands merely due to the cosmological expansion. The comoving distance between the dark matter overdensity and the baryon overdensity shell is the comoving sound horizon at recombination

$$r_s = \int_0^{t_r} v_s \frac{k}{a(t)} dt$$

(this is precisely the argument of cosine in Eq. (31)); its present value is $r_s \simeq 150$ Mpc (we set $a_0 = 1$ here), and the value at redshift z is $r_s/(1+z)$.

[4] Because of the presence of dark energy, the growth factor is even somewhat smaller.

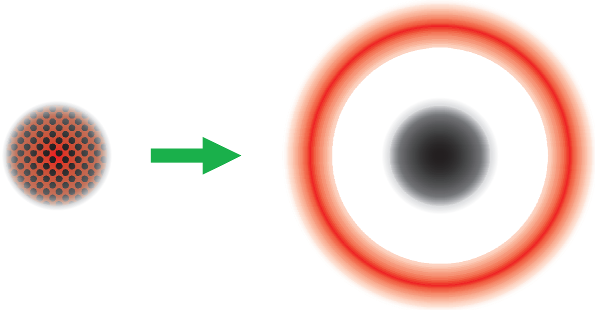


Fig. 8: Schematic picture of Baryon Acoustic Oscillations. Dark regions show dark matter overdensity, less dark (red) regions are the ones with baryon overdensity. Left: initial condition. Right: at recombination and later.

Due to BAO, there is correlation between the matter densities (dark matter plus baryons) separated by the comoving distance r_s. It shows up as a feature in the galaxy-galaxy correlation function $\xi(s)$, where s is the comoving distance. This bump in the correlation function was detected in Ref. [17], see Fig. 9. Clearly, BAO serves as a standard ruler at various redshifts, which can be used to study the evolution of the Universe in a not so distant past.

Currently, BAO is a very powerful tool of observational cosmology. It is used in particular to study time (in)dependence of dark energy.

The bump in the spatial correlation function translates into oscillations in momentum space, hence the name.

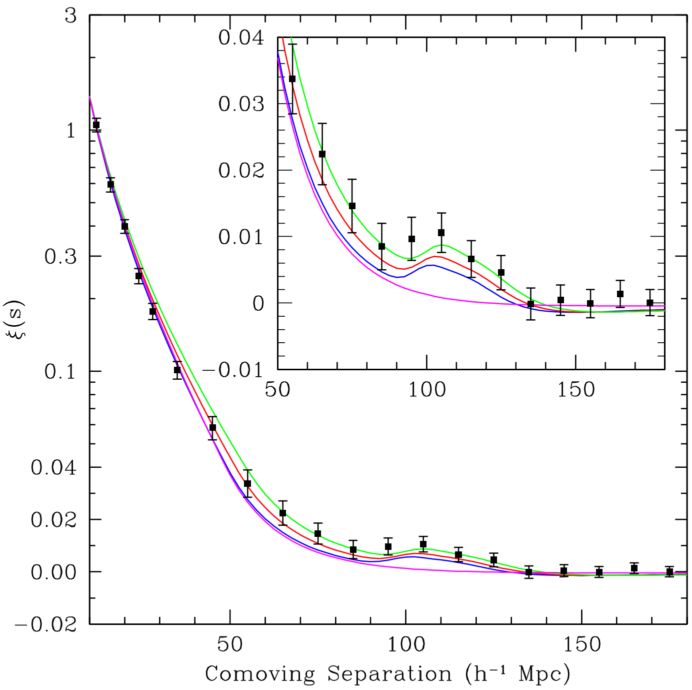


Fig. 9: The first detection of BAO: the correlation function $\xi\,(s)$ determined by the analysis of the Sloan Digital Sky Survey (SDSS) data on the distribution of distant galaxies. Solid lines show the predictions of various cosmological models. Green, red and blue lines correspond to $\Omega_M h^2 = 0.12, 0.13, 0.14$, respectively, with $\Omega_{\rm B} h^2 = 0.024$, $n_s = 0.98$ in all cases. The magenta line corresponds to an unrealistic Universe without baryons. The parameter h is defined in Eq. (4); numerically, $h_0 \approx 0.7$.

5 Astrophysics: more hints on dark matter properties

Important information on dark matter properties is obtained by theoretical analysis of structure formation and its comparison with observational data. Indeed, as we discussed above, dark matter plays the key role in structure formation, so the properties of galaxies and their distribution in space potentially tell us a lot about dark matter.

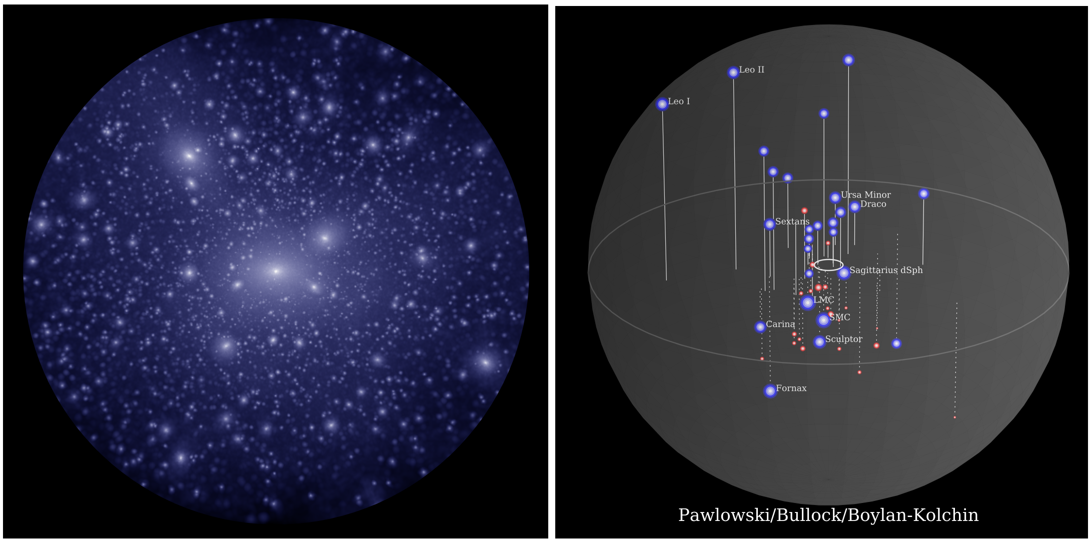
Currently, theoretical studies are made mostly via numerical simulations, many of which ignore effects due to baryons (dark-matter-only). Thus, these simulations give the dark matter distribution. To compare it with observed structures, one often assumes that baryons trace dark matter, with the qualification that baryons are capable of losing their kinetic energy and forming more compact structures inside dark matter halos. In other words, a simulated dark matter collapsed clump of mass, characteristic of a galaxy, is associated with a visible galaxy, while heavier dark matter clumps are interpreted as clusters of galaxies, etc.

Currently, the most popular dark matter scenario is cold dark matter, CDM. It consists of particles whose velocities are negligible at all stages of structure formation, and whose non-gravitational interactions with themselves and with baryons are negligible too (from the viewpoint of structure formation). The CDM numerical simulations (plus the above assumption concerning baryons) are in very good agreement with observations *at relatively large spatial scales*. This is an important result that implies interesting limits on dark matter properties, which we discuss below.

However, there are astrophysical phenomena at shorter scales that may or may not hint towards something different from weakly interacting CDM. The situation is inconclusive yet, but it is worth keeping these phenomena in mind, which we now discuss in turn.

5.1 Missing satellite problem: astrophysics vs warm dark matter

It has long been known that CDM-only simulations produce a lot of small mass halos, $M \lesssim 10^9 M_\odot$ where $M_\odot$ is the Solar mass. Galaxies like the Milky Way have masses $(10^{11} - 10^{12}) M_\odot$, so we are talking about dwarf galaxies. As an example, the left panel of Fig. 10 shows the simulated dark matter distribution in a ball of radius 250 kpc around a galaxy similar to the Milky Way. Assuming that baryons trace dark matter, one observes that there must be hundreds of satellite galaxies there. The actual Milky Way satellites are shown in the right panel of Fig. 10; clearly the number of satellites is a lot smaller. This is the missing satellite problem.



Pawlowski/Bullock/Boylan-Kolchin

Fig. 10: Left: CDM-only simulation of 250 kpc vicinity of a galaxy like the Milky Way; right: actual distribution of satellite galaxies in 250 kpc vicinity of the Milky Way [12].

It is conceivable that this problem has astrophysical solution within the CDM model. One point is that the number of observed faint satellite galaxies around the Milky Way is not that small any longer: while a few years ago this number was about 20, it is currently about 60, and this is not a complete sample because of a limited detection efficiency—the expectation [18] for a complete sample is 150–300 with masses exceeding $10^8 M_\odot$. Another property is that dark matter halos of mass $M < 10^9 M_\odot$ appear fairly inefficient in forming a luminous component[5]—this has been suggested by simulations that include numerous effects due to baryons [19, 20]. Thus, if the CDM model is correct, and the missing satellite problem has an astrophysical solution, there must be a large number of ultra-faint dwarf galaxies with masses $(10^8 - 10^9)M_\odot$ and even larger number of non-luminous dark matter halos with $M \lesssim 10^8 M_\odot$ in the vicinity of the Milky Way. It will be possible to check this prediction in near future, notably, with the Large Synoptic Survey Telescope, LSST [22].

An alternative, particle physics solution to the missing satellite problem is *warm dark matter*, WDM. A reasonably well motivated WDM candidate is a sterile neutrino, which we discuss in Section 8. Another popular candidate is a light gravitino. In WDM case, dark matter particles decouple from the kinetic equilibrium with the baryon-photon component when they are relativistic. Let us assume for definiteness that they are in *kinetic* equilibrium with the cosmic plasma at a temperature T_f when their number density freezes out (there is no *chemical* equilibrium at $T = T_\mathrm{f}$, otherwise the dark matter would be overabundant). After the kinetic equilibrium breaks down at a temperature $T_\mathrm{d} \le T_\mathrm{f}$, the spatial momenta decrease as a^{-1}, i.e., the momenta are of the order T all the time after decoupling. When dark matter particles are relativistic, the density perturbations do not grow: relativistic particles escape from the gravitational potentials, so they do not experience the gravitational instability; in fact, the density perturbations, and hence the gravitational wells, get smeared out instead of getting deeper. WDM particles become non-relativistic at $T \sim m$, where m is their mass. Only after that the WDM perturbations start to grow. Before becoming non-relativistic, WDM particles travel the distance of the order of the horizon size; the WDM perturbations therefore are suppressed at those scales. The horizon size at the time t_nr when $T \sim m$ is of order

$$l_\mathrm{H}(t_\mathrm{nr}) \simeq H^{-1}(T \sim m) = \frac{M^*_\mathrm{Pl}}{T^2} \sim \frac{M^*_\mathrm{Pl}}{m^2} \; .$$

Due to the expansion of the Universe, the corresponding length at present is

$$l_0 = l_\mathrm{H}(t_\mathrm{nr}) \frac{a_0}{a(t_\mathrm{nr})} \sim l_\mathrm{H}(t_\mathrm{nr}) \frac{T}{T_0} \sim \frac{M_\mathrm{Pl}}{m T_0} \; , \tag{33}$$

where we neglected the (rather weak) dependence on g_*. Hence, in WDM scenario, the structures of comoving sizes smaller than l_0 are less abundant as compared to CDM. Let us point out that l_0 refers to the size of the perturbation in the linear regime; in other words, this is the size of the region from which matter collapses into a compact object.

To solve the missing satellite problem, one requires that the mass of dark matter which was originally distributed over the volume of comoving size l_0, and collapsed later on, is of order of the mass of the satellite galaxy,

$$\frac{4\pi}{3} l_0^3 \, \Omega_\mathrm{DM} \rho_c \sim M_\mathrm{dwarf} \; .$$

With $M_\mathrm{dwarf} \sim 10^8 M_\odot$ we find $l_0 \sim 100$ kpc, and Eq. (33) gives the estimate for the mass of a dark matter particle

$$\mathrm{WDM} \; : \quad m_\mathrm{DM} = 3 \; - \; 10 \; \mathrm{keV} \; . \tag{34}$$

On the other hand, this effect is absent, i.e., dark matter is cold, for

$$\mathrm{CDM} \; : \quad m_\mathrm{DM} \gtrsim 10 \; \mathrm{keV} \; . \tag{35}$$

[5] Another effect, important for satellite galaxies close to the center of the Milky Way, is the tidal force due to the gravitational potential produced by the disk of the host galaxy [21].

Let us recall that these estimates apply to particles that are initially in kinetic equilibrium with the cosmic plasma. They do *not* apply in the opposite case; an example is axion dark matter, which is cold despite of very small axion mass.

Reversing the argument, one obtains a limit on the mass of the WDM particle which decouples in the kinetic equilibrium [18],

$$m \gtrsim 4\,\mathrm{keV}\ . \tag{36}$$

5.1.1 Digression: phase space bound

In fact there are other ways to obtain limits on m. One has to do with the phase space density: the maximum value of the coarse grained phase space density

$$f(p,x)_{\text{coarse grained}} = \left(\frac{dN}{d^3p\,d^3x}\right)_{\text{coarse grained}}$$

does not decrease in the course of the evolution (here N is the number of particles). Indeed, the Liouville theorem tells that the microscopic phase space density is time-independent. What happens in the course of evolution is that particles penetrate initially unoccupied regions of phase space, see Fig. 11. While the maximum value of the microscopic phase space density remains constant in time, the maximum value of the coarse grained phase space density (average over phase space volume shown by dashed line in Fig. 11) decreases.

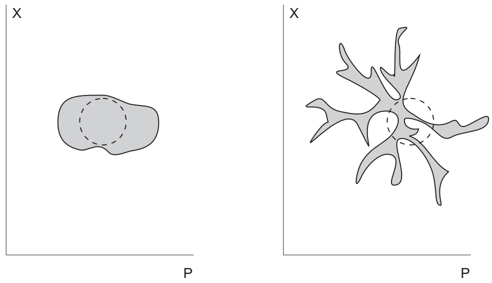


Fig. 11: Sketch of the behavior of an ensemble of particles in phase space. As the ensemble evolves, an initial compact distribution (left panel) becomes less compact.

The initial phase space density of particles in kinetic equilibrium is

$$f_i = \frac{A}{(2\pi)^3}\frac{1}{\mathrm{e}^{p/T}+1}\ ,$$

where we consider fermions for definiteness. The parameter A is determined by requiring that the number density n takes the prescribed value, so that

$$n_0 = \frac{\Omega_{\mathrm{DM}}\rho_{\mathrm{c}}}{m}\ .$$

We find

$$n = \int f_i d^3p = A\cdot\frac{3\zeta(3)}{4\pi^2}T^3\ ,$$

where $\zeta(3)\approx 1.2$. So, the maximum of the initial phase space density is

$$f_{i,\,\max} = \frac{n}{12\pi\zeta(3)T^3} = \frac{\Omega_{\mathrm{DM}}\rho_c}{12\pi\zeta(3)mT_{0\,\mathrm{eff}}^3}\ ,$$

where $T_{0\,\mathrm{eff}}$ depends on the decoupling temperature and is somewhat lower than the present photon temperature.

On the other hand, one can measure a quantity

$$Q = \frac{\rho_{\mathrm{DM,\,gal}}}{\langle v_{\mathrm{gal}}^2/3\rangle^{3/2}}$$

where $\rho_{\mathrm{DM,\,gal}}$ is the mass density (say, in a central part of a dwarf galaxy), $\langle v_{\mathrm{gal}}^2\rangle$ is the average velocity squared, and hence $\langle v_{\mathrm{gal}}^2/3\rangle$ is the average velocity squared along the line of sight (of stars, and hence dark matter particles, in a virialized galaxy). Since $v_{\mathrm{gal}} = p_{\mathrm{gal}}/m$ and $\rho_{\mathrm{DM,\,gal}} = mn_{\mathrm{gal}}$, one obtains an estimate for the phase space density of dark matter particles in a dwarf galaxy,

$$f \simeq \frac{n_{\mathrm{gal}}}{\langle p_{\mathrm{gal}}^2\rangle^{3/2}} = \frac{Q}{3^{3/2}m^4} \; .$$

One requires that

$$f < f_{i\,\mathrm{max}}$$

and obtains the bound on the mass of the dark matter particle

$$m \gtrsim 3 \cdot \left(\frac{Q}{\Omega_{\mathrm{DM}}\rho_{\mathrm{c}}}\right)^{1/3} T_{0\,\mathrm{eff}} \; .$$

The values of Q measured in compact dwarf galaxies are in the range

$$Q \sim (5\cdot 10^{-3} - 2\cdot 10^{-2}) \cdot \frac{M_\odot/\mathrm{pc}^3}{(\mathrm{km/s})^3}$$

while for the relic that decouples at $T = (1-100)$ MeV one has $T_{0\,\mathrm{eff}} = 2.0$ K. This gives [23, 24]

$$m \gtrsim 6 \text{ keV} \; .$$

Accidentally, this bound is similar to Eq. (36). We note that the bounds coming from the phase space density considerations are called bounds of Tremain–Gunn type.

We also note that similar (in fact, slightly stronger but less robust) bounds are obtained by the study of Lyman-α forest, see, e.g., Ref [25].

5.2 Other hints, SIMP and fuzzy DM

There are two other issues that may or may not be problematic for CDM. One is the "core-cusp problem": CDM-only simulations show singular mass density profiles (cusps) in the centers of galaxies, $\rho_{\mathrm{DM}}(r) \propto r^{-1}$, while observations imply enhanced but smooth profiles (cores). Another is the "too-big-to-fail" problem, which currently means that the densities in large satellite galaxies ($M \sim 10^{10}\; M_\odot$), predicted by CDM-only simulations are systematically higher than the observed mass densities [12].

The astrophysical solutions to these problems again have to do with baryons (supernovae feedback, etc.), and also with interactions of satellite galaxies with large host galaxy, the Milky Way, see, e.g., Refs. [12, 26] for discussion. On the particle physics side, WDM may again help out. Two other particle physics solutions are strongly interacting massive particles (SIMP) as dark matter, and fuzzy dark matter.

The idea of SIMP [27] is that dark matter is cold, but elastic scattering of dark matter particles smoothes out the cuspy mass distribution in galactic centers. Elastic scattering can also lead to a decrease of the dark matter density and thus alleviate the too-big-to-fail problem. To give an idea of the elastic scattering cross section, we take mass density of dark matter of order $\rho_{\mathrm{DM}} \sim 1\;\mathrm{GeV/cm}^3$ and require that the mean free path of a dark matter particle is of order $l \sim 1$ kpc (typical values, by order of magnitude, both for centers of large galaxies and for dwarf galaxies),

$$1 \sim l\sigma^{(\mathrm{el})} n_{\mathrm{DM}} = l\sigma^{(\mathrm{el})}\frac{\rho_{\mathrm{DM}}}{m} \; ,$$

and obtain

$$\frac{\sigma^{(\mathrm{el})}}{m} \sim \frac{1}{l\rho_{\mathrm{DM}}} \sim 10^{-24}\frac{\mathrm{cm}^2}{\mathrm{GeV}} \ .$$

This is a very large cross section by particle physics standards, and, in view of Eq. (29), the dark matter particle must be fairly light, $m \lesssim 1$ GeV. The large elastic cross section may be due to t-channel exchange of a light mediator with $m_{med} \sim 10-100$ MeV. This mediator must decay into e^+e^-, $\gamma\gamma$, etc., otherwise it would be dark matter itself. All these features make the SIMP scenario interesting from the viewpoint of collider (search in Z-decays) and "beyond collider" experiments, such as SHiP.

Yet another proposal is fuzzy dark matter consisting of very light bosons,

$$m \sim (10^{-21} - 10^{-22}) \ \mathrm{eV} \ .$$

The mechanism of their production must ensure that all of them are born with zero momenta, i.e., these particles form a scalar condensate. An oversimplified picture is that the de Broglie wavelength of these particles at velocities typical for galactic centers and dwarf galaxies, $v \sim 10$ km/s, is about 1 kpc:

$$\frac{2\pi}{mv} \sim 1 \ \mathrm{kpc} \ .$$

Detailed discussion of advantages of fuzzy dark matter is given, e.g., in Ref. [28]. A way to constrain this scenario is again to study the Lyman-α forest; current constraints [29] are at the level of $2 \cdot 10^{-21}$ eV. Interestingly, effects of fuzzy dark matter may in the future be detected by the pulsar timing method [30].

From a particle physics viewpoint, fuzzy dark matter particles may emerge as pseudo-Nambu–Goldstone bosons, similar to axions. We discuss axions later, and here we borrow the main ideas. The axion-like Lagrangian for the pseudo-Nambu–Goldstone scalar field θ reads

$$L = \frac{F^2}{2}(\partial\theta)^2 - \mu^4(1-\cos^2\theta) \approx \frac{F^2}{2}(\partial\theta)^2 - \frac{\mu^4}{2}\theta^2 \ ,$$

where F is the expectation value of a field that spontaneously breaks approximate $U(1)$ symmetry, and μ is the parameter of the explicit violation of this symmetry. Then the mass of the axion-like particle is

$$m = \frac{\mu^2}{F} \ .$$

The mechanism that creates the scalar condensate is misalignment. The initial value of θ is an arbitrary number between $-\pi$ and π, so that $\theta_i \sim 1$. The field starts to oscillate when the expansion rate becomes small enough, $H \sim m$. The calculation of the present mass density is a simplified version of the axion calculation that we give in Section 7; one finds that $\Omega_{\mathrm{DM}} \sim 0.25$ is obtained for $m = 10^{-22}$ eV if

$$F \sim 10^{17} \ \mathrm{GeV} \ .$$

This is in the ballpark of GUT/string scales, which is intriguing.

5.3 Summary of DM astrophysics

Let us summarize the astrophysics of dark matter.

- Cold dark matter describes remarkably well the distribution and properties of structures in the Universe at relatively large scales, from galaxies like the Milky Way or somewhat smaller ($M \gtrsim 10^{11}M_\odot$), to larger structures like clusters of galaxies, filaments, etc.; also, CDM is remarkably consistent with CMB data which probe even larger scales.
- Currently, data and simulations at shorter scales are inconclusive: they may or may not show that there are "anomalies", the features that contradict the CDM model.

- It will become clear fairly soon whether these “anomalies” are real or not. The progress will come from refined simulations with all effects of baryons included, and from new instruments, notably LSST.
- If the “anomalies” are real, we will have to give up CDM, and, responding to the data, will narrow down the set of dark matter models (WDM, or SIMP, or fuzzy dark matter, or something else). This will have a profound effect on the strategy of search for dark matter particles.
- If the “anomalies” are not there, astrophysics will have to deliver the confirmation of CDM model by the discoveries of relatively light ultra-faint dwarf galaxies ($M = (10^8 - 10^9)M_\odot$) and dark objects of even smaller mass.

All this makes astrophysics a powerful tool of studying dark matter and directing particle physics in its search for dark matter particles.

6 Thermal WIMP

6.1 WIMP abundance: annihilation cross section

Thermal WIMP (weakly interacting massive particle) is a scenario featuring a simple mechanism of the dark matter generation in the early Universe. The WIMP is a *cold* dark matter candidate. Because of its simplicity and robustness, it has been considered by many as the most likely one.

Let us not go into all the details of the (fairly straightforward) calculation of the thermal WIMP abundance. These details are given in several textbooks, and also presented in proceedings of similar Schools, see, e.g., Ref. [31]. Instead, we give the main assumptions behind this mechanism and describe the main steps of the calculation.

One assumes that there exists a heavy stable neutral particle χ, and that χ particles can only be destroyed or created in the cosmic plasma via their pair-annihilation or creation, with the annihilation products being the particles of the Standard Model[6]. We note that there is a version of WIMP model in which the χ particle is not truly neutral, i.e., it does not coincide with its own antiparticle. In that case one assumes that the production and destruction occurs only via $\chi - \bar{\chi}$ annihilation, and there is no asymmetry between χ and $\bar{\chi}$ in the cosmic plasma, $n_\chi = n_{\bar{\chi}}$. The calculation in the $\chi - \bar{\chi}$ model is identical to the case of a truly neutral particle, so we consider the latter case only.

One also assumes that the χ particles are not strongly coupled, but the $\chi - \chi$ annihilation cross section is sufficiently large, so the χ particles are in complete thermal equilibrium at high temperatures. The latter assumption is justified in the end of the calculation. The thermal equilibrium means, in particular, that the abundance of χ particles is given by the standard Bose–Einstein or Fermi–Dirac distribution formula.

The cosmological behaviour of χ particles is as follows: At high temperatures, $T \gg m_\chi$, the number density of χ particles is high, $n_\chi(T) \sim T^3$. As the temperature drops below m_χ, the equilibrium number density decreases,

$$n_\chi^{(\mathrm{eq})} \propto \mathrm{e}^{-\frac{m_\chi}{T}} , \tag{37}$$

At some “freeze-out” temperature T_f the number density becomes so small, that χ particles can no longer meet each other during the Hubble time, and their annihilation terminates[7]. After that the number density of survived χ particles decreases as a^{-3}, and these relic particles form the CDM. The freeze-out

[6] The latter assumption can be relaxed: decay products of χ particles may be new particles which sufficiently strongly interact with the Standard Model particles and in the end disappear from the cosmic plasma. Also, destruction and creation of χ particles may occur via co-annihilation with their nearly degenerate partners and inverse pair creation processes; this occurs in a class of supersymmetric models where χ is the lightest supersymmetric particle and its partner is the next-to-lightest supersymmetric particle.

[7] This is a slightly oversimplified picture, which, however, gives a correct estimate, modulo a factor of order 1 in the argument of the logarithm.

temperature $T_{\rm f}$ is obtained by equating the mean free time of χ particle with respect to the annihilation,

$$\tau_{\rm ann}(T_{\rm f}) = (\sigma_0(T_{\rm f}) n_\chi(T_{\rm f}))^{-1}$$

to the Hubble time (see Eq. (20))

$$H^{-1}(T_{\rm f}) = \frac{M_{\rm Pl}^*}{T_{\rm f}^2} \; .$$

Here we introduced the weighted annihilation cross section

$$\sigma_0(T) = \langle \sigma_{\rm ann} v \rangle_T \; ,$$

where v is the relative velocity of χ particles (in the non-relativistic regime relevant here we have $v \ll 1$), and we average over the thermal ensemble.

Thus, freeze-out occurs when

$$\sigma_0(T_{\rm f}) n_\chi(T_{\rm f}) = \frac{T_{\rm f}^2}{M_{\rm Pl}^*} \; .$$

Because of exponential decay of $n_\chi^{(eq)}$ with temperature, Eq. (37), the freeze-out temperature is smaller than the mass by a logarithmic factor only,

$$T_{\rm f} \approx \frac{m_\chi}{\ln(M_{\rm Pl}^* m_\chi \sigma_0)} \; . \tag{38}$$

Note that due to the large logarithm, χ particles are indeed non-relativistic at freeze-out: their velocity squared is of order of

$$v^2(T_{\rm f}) \simeq 0.1 \; .$$

At freeze-out, the number density is

$$n_\chi(T_{\rm f}) = \frac{T_{\rm f}^2}{M_{\rm Pl}^* \sigma_0(T_{\rm f})} \; , \tag{39}$$

Note that this density is inversely proportional to the annihilation cross section (modulo logarithm). The reason is that for a higher annihilation cross section, the creation-annihilation processes are longer in equilibrium, and less χ particles survive. Up to a numerical factor of order 1, the number-to-entropy ratio at freeze-out is

$$\frac{n_\chi}{s} \simeq \frac{1}{g_*(T_{\rm f}) M_{\rm Pl}^* T_{\rm f} \sigma_0(T_{\rm f})} \; . \tag{40}$$

This ratio stays constant until the present time, so the present number density of χ particles is $n_{\chi,0} = s_0 \cdot (n_\chi/s)_{\rm freeze-out}$, and the mass-to-entropy ratio is

$$\frac{\rho_{\chi,0}}{s_0} = \frac{m_\chi n_{\chi,0}}{s_0} \simeq \frac{\ln(M_{\rm Pl}^* m_\chi \sigma_0)}{g_*(T_{\rm f}) M_{\rm Pl}^* \sigma_0(T_{\rm f})} \simeq \frac{\ln(M_{\rm Pl}^* m_\chi \sigma_0)}{\sqrt{g_*(T_{\rm f})} M_{\rm Pl} \sigma_0(T_{\rm f})} \; ,$$

where we made use of (38). This formula is remarkable. The mass density depends mostly on one parameter, the annihilation cross section σ_0. The dependence on the mass of χ particle is through the logarithm and through $g_*(T_{\rm f})$; it is very mild. Plugging in $g_*(T_{\rm f}) \sim 100$, as well as a numerical factor omitted in Eq. (40), and comparing with (28) we obtain the estimate

$$\sigma_0(T_{\rm f}) \equiv \langle \sigma v \rangle (T_{\rm f}) = 1 \cdot 10^{-36}\ {\rm cm}^2 = 1\ {\rm pb} \; . \tag{41}$$

This is a weak scale cross section, which tells us that the relevant energy scale is 100 GeV–TeV. We note in passing that the estimate (41) is quite precise and robust.

The annihilation cross section can be parameterized as $\sigma_0 = \alpha^2/M^2$ where α is some coupling constant, and M is a mass scale responsible for the annihilation processes[8] (which may be higher than m_χ). This parametrization is suggested by the picture of χ pair-annihilation via the exchange by another particle of mass M. With $\alpha \sim 10^{-2}$, the estimate for the mass scale is roughly $M \sim 1$ TeV. Thus, with mild assumptions, we find that the WIMP dark matter may naturally originate from the TeV-scale physics. In fact, what we have found can be understood as an approximate equality between the cosmological parameter, the mass-to-entropy ratio of dark matter, and the particle physics parameters,

$$\text{mass-to-entropy} \simeq \frac{1}{M_{\rm Pl}} \left(\frac{\rm TeV}{\alpha_W} \right)^2 .$$

Both are of order 10^{-10} GeV, and it is very tempting to think that this "WIMP miracle" is not a mere coincidence. For long time the above argument has been—and still is—a strong motivation for WIMP searches.

6.2 WIMP candidates: "minimal" and supersymmetry; direct searches

6.2.1 "Minimal" WIMP

Even though the name—*weakly* interacting massive particle—suggests that this particle participates in the Standard Model weak interactions, in most theoretical models this is not so. An exception is the "minimal" WIMP [32]. This is a member of the electroweak multiplet with zero electric charge and zero coupling to the Z boson (couplings to the photon and Z would yield to too strong interactions with the Standard Model particles which are forbidden by direct searches). This is possible for vector-like 5-plet (weak isospin 2) with zero weak hypercharge. Another, albeit fine-tuned option is vector-like triplet (weak isospin 1) with zero weak hypercharge. Particles in vector-like representations may have "hard" masses (not given by Englert–Brout–Higgs mechanism). The right annihilation cross section in Eq. (41) is obtained for masses of these particles

$$\text{5-plet}: \quad m_5 = 9.6 \text{ TeV} , \qquad \text{3-plet}: \quad m_3 = 3 \text{ TeV} .$$

These particles are on the verge of being ruled out by direct searches.

6.2.2 Neutralino

A well motivated WIMP candidate is the neutralino of supersymmetric extensions of the Standard Model. The situation with the neutralino is rather tense, however. One point is that the pair-annihilation of neutralinos often occurs in p-wave, rather than s-wave. This gives the suppression factor in σ_0, proportional to $v^2 \sim 0.1$. Hence, neutralinos tend to be overproduced for large part of the parameter space of MSSM and other supersymmetry (SUSY) models.

Another point is the null results of the direct searches for WIMPs in underground laboratories. The idea of direct search is that WIMPs orbiting around the center of our Galaxy with velocity of order 10^{-3} sometimes hit a nucleus in a detector and deposit small energy in it. The relevant parameters for these searches are WIMP-nucleon elastic scattering cross section and WIMP mass. One distinguishes spin-independent and spin-dependent scattering. In the former case, the WIMP-*nucleus* cross section is proportional to A^2, where A is the number of nucleons in the nucleus (this is an effect of coherent scattering), while in the latter case the cross section is proportional to $J(J+1)$ where J is the spin of the nucleus.

To illustrate the progress in direct searches, we show in Fig. 12 the situation for neutralinos and the direct searches as of 1999, Ref. [33], while Fig. 13 shows the best current limits on the spin-independent cross section [34].

[8] For s-wave annihilation, σ_0 is independent of the particle velocity, and hence temperature; if the annihilation is in p-wave, there is an additional suppression by $v^2(T_{\rm f}) \sim 0.1$.

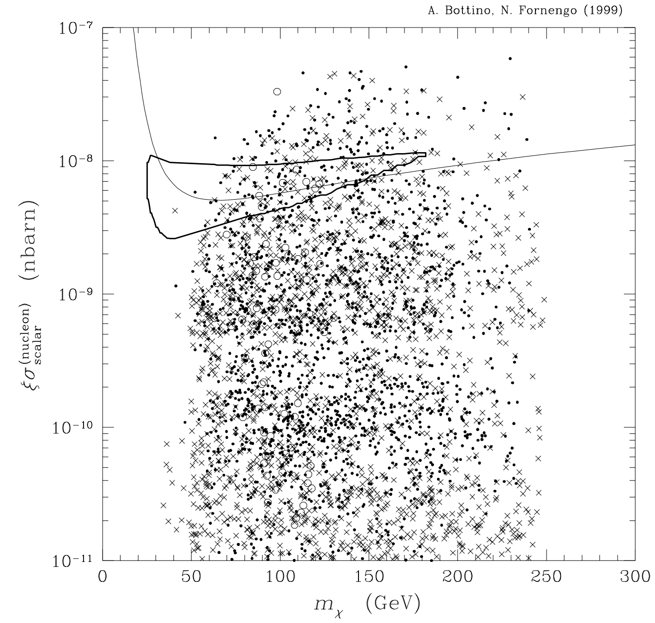


Fig. 12: The situation with neutralinos and the direct searches in 1999 [33]. Shown are theoretical predictions (crosses and dots) and direct detection limits (open solid line; closed solid line is DAMA hint). Vertical axis: spin-independent cross section of elastic WIMP-nucleus scattering per nucleon; parameter ξ takes value 1 for a spin-1/2 neutralino; note that 10^{-10} nbarn $= 10^{-43}$ cm^2.

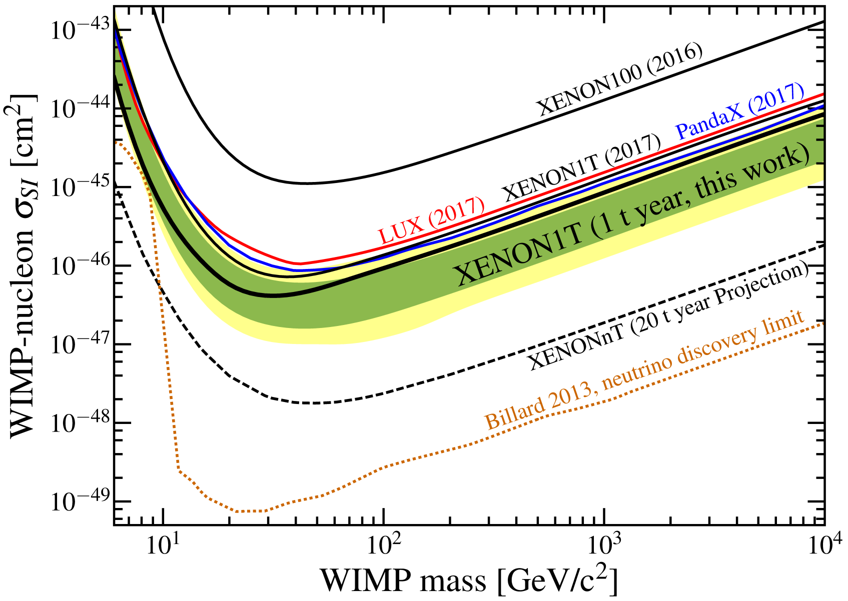


Fig. 13: Current results of direct searches for WIMPs: best limits on spin-independent WIMP-nucleus cross section per nucleon come from the XENON-1T experiment [34]. Note that the cross section 10^{-10} nbarn $= 10^{-43}$ cm^2 in the lower part of Fig. 12 is in the upper part of this figure.

Figure 14 shows both current limits (solid lines) and projected sensitivities of future dark matter detection experiments, again for spin-independent interactions [10]. We see that, on the one hand, the progress in experimental search is truly remarkable, and, on the other, the null results of the searches are becoming alarming. The null results of direct (and also indirect, see below) searches are particularly worrying in view of the null results of SUSY searches at the Large Hadron Collider (LHC).

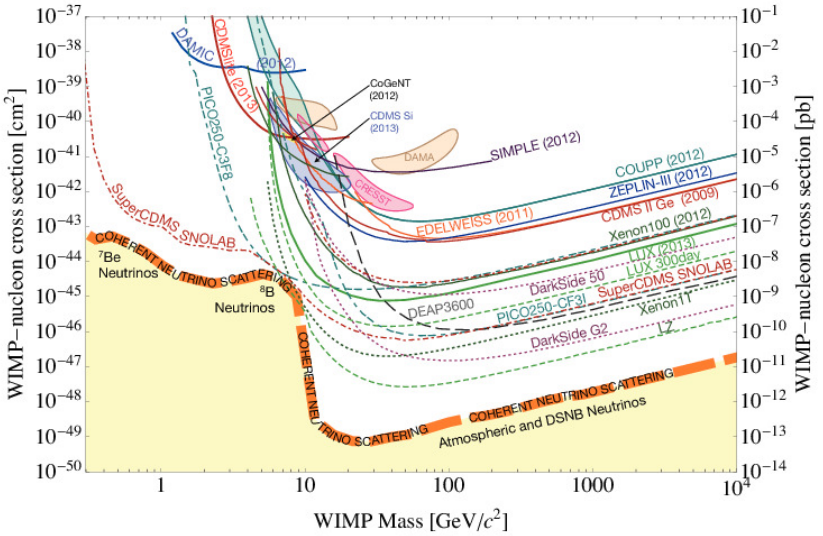


Fig. 14: Current limits and projected sensitivities of direct searches for WIMPs (spin-independent WIMP-nucleus cross section per nucleon). Yellow band in the lower part is the "neutrino floor", at which interactions of cosmic neutrinos become an important background.

6.3 Ad hoc WIMP candidates; indirect searches and the LHC

In view of the strong direct detection limits and null results of the SUSY searches at the LHC, it makes sense to consider less motivated, ad hoc WIMP candidates. The simplest assumption is that the WIMP is not nearly degenerate with any other new particle, so that the calculation of its abundance outlined above applies, and that there is one particle that mediates its pair-annihilation. This mediator can be either a Standard Model particle or a new one; we give examples of both cases. The models of this sort are often called simplified. We emphasize that the two examples of simplified models which we are going to discuss do not exhaust all possible WIMPs and mediators. Some of the models that we leave aside are actually consistent with both cosmology (they give the right value of $\Omega_{\rm DM}$) and experimental limits. The study of numerous simplified models is given, e.g., in Ref. [11].

With this reservation, it is fair to say that many simplified models are either already ruled out or will be ruled out soon. As one illustration, we consider the "Higgs portal", a set of models where the only field which interacts directly with WIMPs is the Englert–Brout–Higgs field. The lowest dimension Higgs-WIMP interaction terms in the cases of a spin-0 WIMP χ and spin-1/2 WIMP ψ are

$$\lambda^H_\chi \chi^* \chi H^\dagger H \; , \qquad \frac{\lambda^H_\psi}{\Lambda} \bar{\psi}\psi H^\dagger H \; ,$$

where H is the EBH field. Here λ^H_χ, λ^H_ψ are dimensionless parameters, while Λ has the dimension of mass. In both cases χ (ψ) is a Standard Model singlet with zero weak hypercharge; it has a "hard" mass $m_{\chi(\psi)}$. Since the vacuum expectation value of the EBH field H is non-zero, the above terms induce trilinear WIMP-WIMP-Higgs interactions responsible for s-channel WIMP annihilation via the Higgs exchange. It is this annihilation that is relevant in the early Universe. The trouble is that almost the entire parameter space of the Higgs portal is ruled out by direct searches. This is illustrated in Fig. 15,

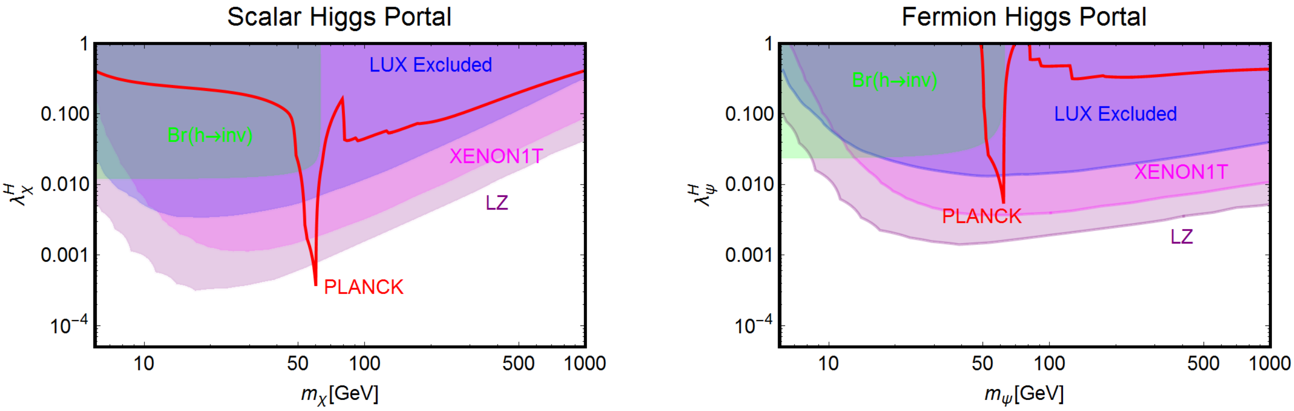


Fig. 15: Predictions from dark matter abundance (red solid lines labeled "PLANCK") and direct detection limits (shadows) in the Higgs portal models [11]. Left panel: spin-0 WIMP; right panel: spin-1/2 WIMP.

Ref. [11]. Another illustration is the Z'-portal. One assumes that both WIMP (say, spin-1/2 particle ψ) and Standard Model fermions interact with a new vector boson Z':

$$g_\psi \bar{\psi}(V_\psi - A_\psi \gamma^5)\psi Z' + \sum_f g_f \bar{f}(V_f - A_f \gamma^5) f \, Z' \; , \qquad (42)$$

where sum runs over all Standard Model fermions (an important role is played by quarks). The coupling constants g_ψ, g_f are often chosen to be of order 0.5, as suggested by GUTs. Almost all parameter space of Z'-portal models with $V_\psi \neq 0$ is also ruled out by direct searches [11], as shown in Fig. 16.

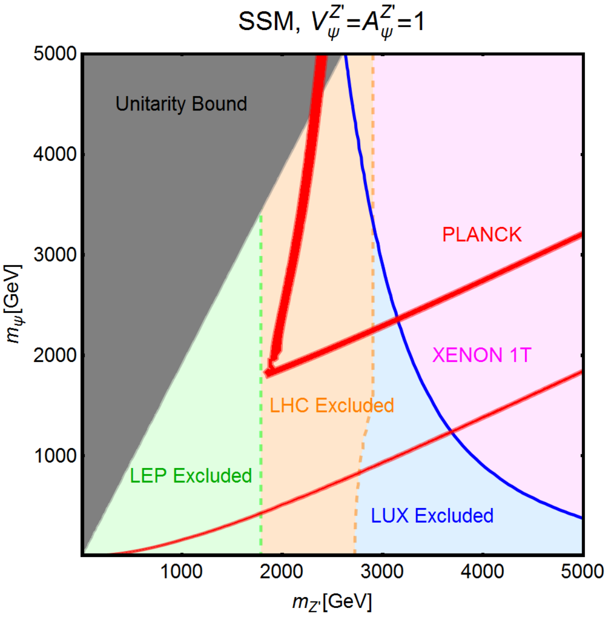


Fig. 16: Same as in Fig. 15 but for the Z'-portal in Eq. (42) with $g_\psi = g_f = 0.65$ and $V_\psi = A_\psi = V_f = A_f = 1$.

The situation is better in models with axial-vector interactions of a new vector boson (we still call it Z') with both the Standard Model particles and WIMPs,

$$V_\psi = V_f = 0 \; .$$

In that case, the interaction of WIMPs with a nucleons is spin-dependent, the elastic WIMP-nucleus cross section is not enhanced by A^2, so the direct detection limits are not as strong as in the case of a spin-independent interaction. An important player here is the LHC, whose limits are the most stringent [11], see Fig. 17. We see from Fig. 17 that models with $M_{Z'} \gtrsim 2.8$ TeV are capable of producing the correct abundance of dark matter and at the same time are not ruled out experimentally.

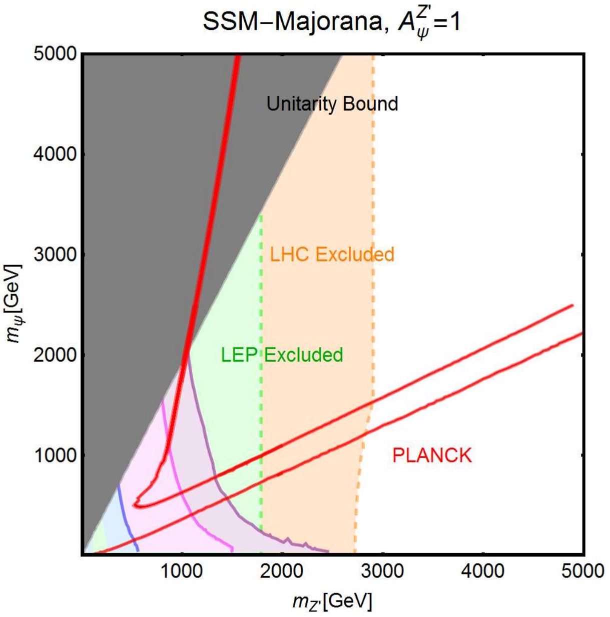


Fig. 17: Same as in Fig. 16 but for an axial-vector Z'-portal in Eq. (42) with and $V_\psi = V_f = 0$, $A_\psi = A_f = 1$.

Another way of comparing current sensitivities of direct and LHC searches is given in Figs. 18 and 19. The plots (compiled by the ATLAS collaboration) refer to the model in Eq. (42) with a vector boson Z' and coupling constants with quarks $g_{\rm q}$, leptons $g_{\rm l}$ and WIMPs $g_\psi \equiv g_\chi$ whose values are written in the figures. Figure 18 shows the limits in the vector case, $A_\psi = A_f = 0$, $V_\psi = V_f = 1$, while Fig. 19 refers to the axial-vector case $A_\psi = A_f = 1$, $V_\psi = V_f = 0$. Clearly, the direct searches are more sensitive than the LHC in the vector case (spin-independent WIMP-nucleon elastic cross section), while the LHC wins in the axial-vector case (spin-dependent elastic cross section). Overall, the LHC has become an important source of limits on WIMPs.

Besides direct and LHC searches for cosmic and collider-produced WIMPs, respectively, important ways to address WIMPs are indirect searches. One approach is to search for high energy γ-rays which are produced in annihilations of WIMPs in various cosmic sources, from dwarf galaxies, to the Galactic center, to clusters of galaxies, and also a diffuse γ-ray flux coming from the entire Universe. This approach is particularly relevant if the WIMP annihilation proceeds in s-wave: in that case the non-relativistic annihilation rate is determined by Eq. (41), which is velocity-independent (modulo a possible Sommerfeld enhancement, see Ref. [35] for a detailed discussion). On the contrary, for p-wave annihilation the rate σv is proportional to v^2, and since the velocities in the sources are small ($v^2 \lesssim 10^{-6}$ as compared to $v^2 \simeq 0.1$ relevant to Eq. (41)), the annihilation cross section is strongly suppressed in the present Universe. Thus, meaningful limits are obtained by γ-ray observatories for WIMPs annihilating in s-wave. The current situation and future prospects are illustrated in Fig. 20, Ref. [10]. The assumption that enters this compilation is that the major WIMP annihilation channel is $b\bar{b}$. Clearly, already existing instruments, and to even larger extent future experiments are sensitive to a wide class of WIMP models.

Indirect searches for dark matter WIMPs include the search for neutrinos coming from the centers of the Earth and Sun (WIMPs may concentrate and annihilate there), see, e.g., Ref. [36], positrons and

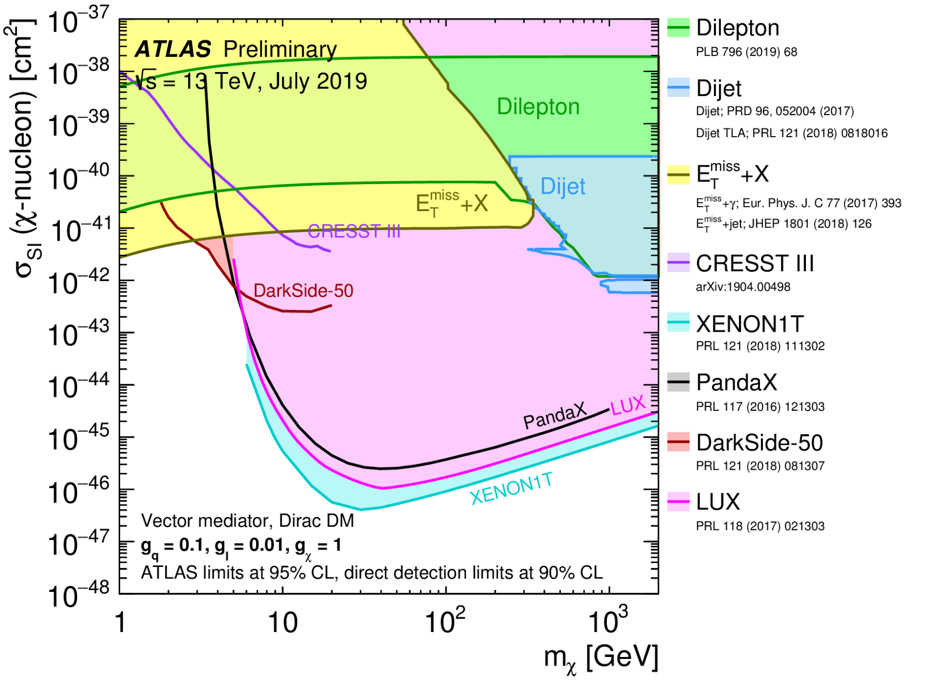


Fig. 18: LHC and direct detection limits in the case of spin-independent WIMP-nucleon elastic cross section.

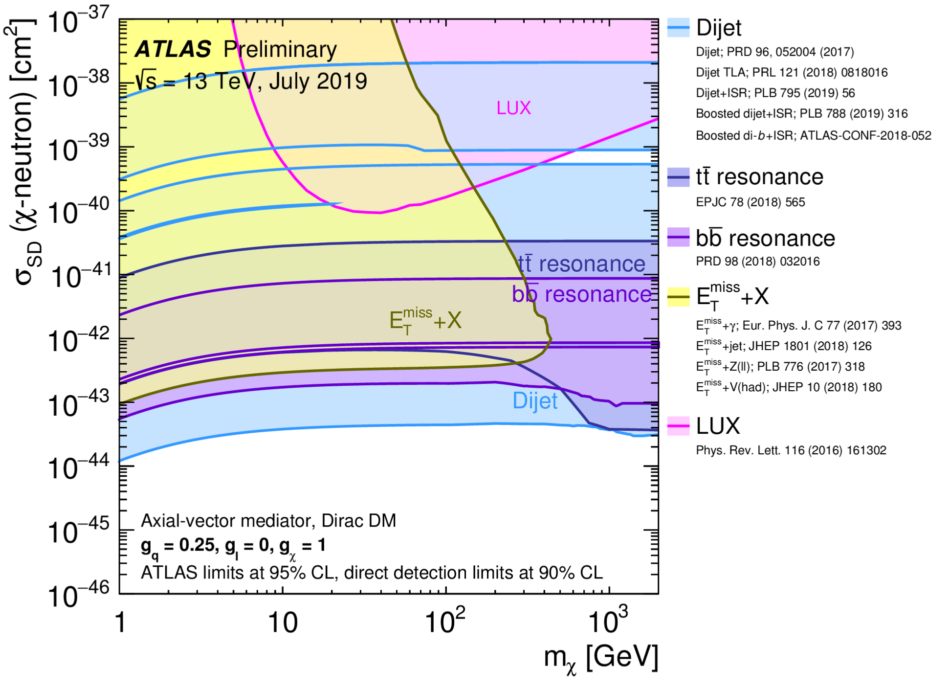


Fig. 19: The same as in Fig. 18 but for spin-dependent WIMP-nucleon elastic cross section.

antiprotons in cosmic rays (produced in WIMP annihilations in our Galaxy), see, e.g., Ref. [37]. These searches have produced interesting, albeit model-dependent limits on WIMP properties.

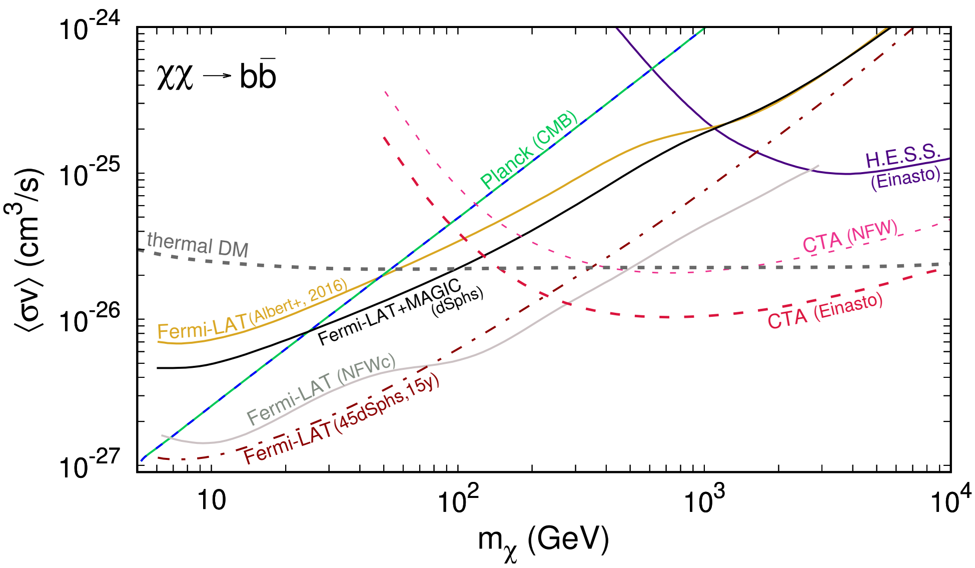


Fig. 20: Limits on WIMP annihilation cross section obtained by γ-ray telescopes (solid lines) and projected sensitivities of future γ-ray observatories (dashed lines). "NFW" and "Einasto" refer to different dark matter profiles in galaxies. Dashed line "thermal DM" is the prediction from cosmology in Eq. (41) under assumption of s-wave annihilation. Note the different units for $\langle\sigma v\rangle$ used in this figure and in Eq. (41).

6.4 WIMP summary

- While the WIMP hypothesis was very attractive for long time, and the SUSY neutralino was considered the best candidate, today the WIMP option is highly squeezed. On the one hand, the parameter space of most of the concrete models is strongly constrained by direct, LHC and indirect searches. On the other hand, SUSY searches at the LHC have moved colored superpartner masses into the TeV region, thus making SUSY less attractive from the viewpoint of solving the gauge hierarchy problem.
- This does not mean too much, however: we would like to discover *one theory* and *one point in its parameter space*.
- Hunt for WIMPs continues in numerous directions. Their potential is far from being exhausted. Concerning direct searches, we will soon face the neutrino floor problem—the situation where the cosmic neutrino background will show up. It is time to look into ways to go beneath the neutrino floor.
- With a null results of WIMP searches, it makes a lot of sense to strengthen also searches for other dark matter candidates.

7 Axions

An Axion is a consequence of the Peccei–Quinn solution to the strong CP problem. It is a pseudo-Nambu–Goldstone boson of an approximate Peccei–Quinn symmetry.

7.1 Strong CP problem

To understand the strong CP problem, we begin with considering QCD in the chiral limit $m_u = m_d = m_s = 0$. The Lagrangian is

$$L_{QCD,m=0} = -\frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu} + \sum_i \bar{q}_i i\gamma^\mu D_\mu q_i$$

$$= -\frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu} + \sum_i \left(\bar{q}_{L,i}i\gamma^\mu D_\mu q_{L,i} + \bar{q}_{R,i}i\gamma^\mu D_\mu q_{R,i}\right) \ ,$$

where $i = u, d, s$. As it stands, it is invariant under independent transformations of left and right quark fields $q_{L,i}$ and $q_{R,i}$, each with arbitrary unitary matrices. Naively, this means that the theory possesses a large symmetry

$$SU(3)_L \times U(1)_L \times SU(3)_R \times U(1)_R = SU(3)_L \times SU(3)_R \times U(1)_B \times U(1)_A \tag{43}$$

where vector $U(1)_B$ is baryon number symmetry, $q_i \to \mathrm{e}^{i\alpha}q_i$, while axial $U(1)_A$ act as $q_i \to \mathrm{e}^{i\beta\gamma^5}q_i$.

The symmetry in Eq. (43) is spontaneously broken: there exist quark condensates in the QCD vacuum:

$$\langle \bar{u}_L u_R\rangle = \langle \bar{d}_L d_R\rangle = \langle \bar{s}_L s_R\rangle = \frac{1}{2}\langle \bar{q}q\rangle \sim \Lambda^3_{\text{QCD}} \tag{44}$$

The unbroken symmetry $SU(3)_V$ rotates left and right quarks together (this is the well known flavor $SU(3)$); $U(1)_B$ also remains unbroken.

Spontaneous breaking of a global symmetry always leads to the presence of a Nambu–Goldstone bosons. Naively, one expects that there are 9 Nambu–Goldstone bosons: 8 of them come from symmetry breaking $SU(3)_L \times SU(3)_R \to SU(3)_V$, and one from $U(1)_B \times U(1)_A \to U(1)_B$ (since the original symmetry is explicitly broken by quark masses, these should be pseudo-Nambu–Goldstone bosons with non-zero mass). However, there are only 8 light pseudoscalar particles whose properties are well described by Nambu–Goldstone theory: these are $\pi^\pm$, π^0, $K^\pm$, K^0, $\bar{K}^0$, η. Indeed, their masses squared are proportional to quark masses, e.g., $m_\pi^2 = (m_u + m_d)\langle\bar{q}q\rangle/f_\pi^2$. Importantly, yet another pseudoscalar η' is heavy and does not behave like pseudo-Nambu–Goldstone boson.

The reason for this mismatch (absence of the 9th pseudo-Nambu–Goldstone boson) is that $U(1)_A$ is not, in fact, a symmetry of QCD even in the chiral limit. The corresponding axial current suffers, at the quantum level, an Adler–Bell–Jackiw (triangle, or axial) anomaly,

$$\partial_\mu J^\mu_A \neq 0 \ .$$

This means that the axial charge is not conserved, and thus the $U(1)_A$ is explicitly broken. We discuss this phenomenon in little more detail in Section 10.2 in the context of electroweak baryon number non-conservation.

The strong CP problem [38–40] emerges in the following way. One considers quark mass terms in the Standard Model Lagrangian, which are obtained from the Yukawa interaction terms with a non-zero Higgs expectation value. A common believe is that one can perform unitary rotations of quark fields to make quark mass terms real (and in this way generate the Cabibbo–Kobayashi–Maskawa (CKM) matrix in quark interactions with W-bosons). This is not quite true, precisely because one cannot freely use the $U(1)_A$-rotation. In fact, by performing a $SU(3)_L \times SU(3)_R \times U(1)_B$-rotation, one casts the mass term of light quarks into the form

$$L_m = \mathrm{e}^{i\theta} \cdot m_{ij}^{\text{CKM}} \bar{q}_{L,i}q_{R,j} + h.c. \ ,$$

where $m_{ij}^{\text{CKM}} = \text{diag}(m_u, m_d, m_s)$ is a real diagonal matrix, and θ is some phase. Naively, this phase can be rotated away by an axial rotation of all three light quark fields, $q_i \to \mathrm{e}^{-i\theta\gamma_5/2}q_i$, but, as we discussed, this is not an innocent field redefinition. What happens instead is that this transformation generates an extra term in the QCD Lagrangian

$$\Delta L = \frac{\alpha_s}{8\pi} \cdot \theta \cdot G^a_{\mu\nu}\tilde{G}^{\mu\nu\ a} \ , \tag{45}$$

where α_s is the $SU(3)_c$ gauge coupling, $G^a_{\mu\nu}$ is the gluon field strength, $\tilde{G}^{\mu\nu\ a} = \frac{1}{2}\epsilon^{\mu\nu\lambda\rho}G^a_{\lambda\rho}$ is the dual tensor. The term (45) is invariant under gauge symmetries of the Standard Model, but it violates P and

CP symmetry. A similar term, but with another parameter θ_0 instead of θ, can already exist in the initial QCD Lagrangian. The combined parameter

$$\bar{\theta} = \theta + \theta_0$$

is a “coupling constant” that cannot be removed by field redefinition, and QCD with a non-zero $\bar{\theta}$ violates CP symmetry.

Let us show explicitly that the parameter $\bar{\theta}$ is physical, i.e., some physical quantities depend on $\bar{\theta}$. To this end, we perform a chiral rotation of light quark fields $q_i \to \mathrm{e}^{+i\bar{\theta}\gamma_5/2} q_i$ to get rid of the term (45) and generate the phase in the quark mass terms

$$L_m = \sum_i \mathrm{e}^{i\bar{\theta}} m_i \bar{q}_{L,i} q_{R,i} + h.c.$$

Let us consider for simplicity two light quark flavors u and d with equal masses $m_{\mathrm{u}} = m_{\mathrm{d}} \equiv m_{\mathrm{q}} \sim 4$ MeV and calculate the vacuum energy density in such a theory. We use perturbation theory in quark masses, and work to the leading order. Then the $\bar{\theta}$-dependent part of the vacuum energy density is $V(\bar{\theta}) = -\langle L_m \rangle$. We recall that $\langle \bar{q}q \rangle$ is non-zero in the chiral limit, see Eq. (44), and observe that it is real, provided that the term (45) is absent (no spontaneous CP violation in the chiral limit). Importantly, $\langle \bar{q}q \rangle$ does not have an arbitrary phase, since the arbitrariness of this phase would mean that $U(1)_A$ is a (spontaneously broken) symmetry, which is not the case, as we discussed above. Thus, we obtain

$$V(\bar{\theta}) = -\langle L_m \rangle = -2m_{\mathrm{q}} \langle \bar{q}q \rangle \cos\bar{\theta} = -\frac{m_\pi^2 f_\pi^2}{4} \cos\bar{\theta} \; . \tag{46}$$

This shows explicitly that $\bar{\theta}$ is a physically relevant parameter. We note in passing that the expression for $V(\bar{\theta})$ is, in fact, more complicated, especially for $m_u \neq m_d$ and also for three quark flavors, but the main property—minimum at $\bar{\theta} = 0$—is intact.

Thus, $\bar{\theta}$ is a new coupling constant that can take any value in the interval $(-\pi, \pi)$. There is no reason to think that $\bar{\theta} = 0$. The term (45) has a dramatic phenomenological consequence: it generates a electric dipole moment (EDM) of the neutron d_n, which is estimated as [41]

$$d_n \sim \bar{\theta} \cdot 10^{-16} \cdot e \cdot \mathrm{cm} \; . \tag{47}$$

The neutron EDM is strongly constrained experimentally,

$$d_n \lesssim 3 \cdot 10^{-26} \cdot e \cdot \mathrm{cm} \; . \tag{48}$$

This leads to the bound on the parameter $\bar{\theta}$,

$$|\bar{\theta}| < 0.3 \cdot 10^{-9} \; .$$

The problem to explain such a small value of $\bar{\theta}$ is precisely the strong CP problem.

A solution to this problem does not exist within the Standard Model. The solution is offered by models with an axion. The idea of these models is to promote the $\bar{\theta}$-parameter to a field, which is precisely the axion field. This can be done in various ways. Two well-known ones are Dine–Fischler–Srednicki–Zhitnitsky [45,46] (DFSZ) and Kim–Shifman–Vainshtein–Zakharov [47,48] (KSVZ) mechanisms[9]. In either case, one introduces a complex scalar field Φ and makes sure that without QCD effects, the theory is invariant under global Peccei–Quinn $U(1)_{PQ}$ symmetry. Under this symmetry, the field Φ transforms as $\Phi \to \mathrm{e}^{i\alpha}\Phi$. One also arranges that the QCD effects make this symmetry anomalous,

[9] Earlier and even simpler is Weinberg–Wilczek model [43,44], but it is ruled out experimentally.

very much like $U(1)_A$, so that under the $U(1)_{PQ}$-transformation, the Lagrangian obtains an additional contribution

$$\Delta L = C\frac{\alpha_s}{8\pi} \cdot \alpha \cdot G^a_{\mu\nu}\tilde{G}^{\mu\nu\, a} \ , \tag{49}$$

where C is a model-dependent constant of order 1. A simple example is the KSVZ model: one adds a new quark ψ which interacts with Φ as follows:

$$L_{int} = h\Phi\bar{\psi}_L\psi_R + h.c. \tag{50}$$

where h is Yukawa coupling. Then the Peccei–Quinn transformation is

$$\Phi \to \mathrm{e}^{i\alpha}\Phi \ , \qquad \psi \to \mathrm{e}^{i\alpha\gamma^5/2}\psi \ ,$$

while "our" quark fields are $U(1)_{PQ}$-singlets. In the same way as above, this transformation induces the term (49), as required.

Now, one arranges the scalar potential for Φ in such a way that the Peccei–Quinn symmetry is spontaneously broken at very high energy. If not for QCD effects, the phase of Φ would be a massless Nambu–Goldstone boson, the axion. At low energies one writes $\Phi = f_{PQ} \cdot \mathrm{e}^{i\theta(x)}$, where f_{PQ} is the Peccei–Quinn vacuum expectation value. In the absence of QCD, the field θ is rotated away from the non-derivative part of the action by the Peccei–Quinn rotation, while it reappears in the form of Eq. (49) when QCD is switched on. We see that the parameter $\bar{\theta}$ is indeed promoted to a field, and this parameter disappears upon shifting $\theta(x) \to \theta(x) - \bar{\theta}$; we are free to set $\bar{\theta} = 0$. Now, there is a potential for the field θ; it is given precisely by Eq. (46) with $\bar{\theta}$ replaced by θ. Hence, the low energy axion Lagrangian reads

$$L_a = \frac{f_{PQ}^2}{2}\partial_\mu\theta\partial^\mu\theta - V(\theta) \ .$$

As usual, the first term here comes from the kinetic term for the field Φ. We recall that the minimum of $V(\theta)$ is at $\theta = 0$; at this value CP symmetry is not violated, the strong CP problem is solved! We now make a field redefinition, $\theta = a/f_{PQ}$ and find from Eq. (46) that the quadratic axion Lagrangian is

$$L_a = \frac{1}{2}\partial_\mu a\partial^\mu a - \frac{m_a^2}{2}a^2 \ ,$$

where

$$m_a = \frac{m_\pi f_\pi}{2f_{PQ}} \ . \tag{51}$$

The axion is a *pseudo*-Nambu-Goldstone boson.

To summarize, for large Peccei–Quinn scale $f_{PQ} \gg M_W$, the axion is a light particle whose interactions with the Standard Model fields are very weak. Like for any Nambu–Goldstone field, the tree-level interactions of the axion with quarks and leptons are described by the generalized Goldberger–Treiman formula

$$L_{af} = \frac{1}{f_{PQ}} \cdot \partial_\mu a \cdot J^\mu_{PQ} \ . \tag{52}$$

Here

$$J^\mu_{PQ} = \sum_f e_f^{(PQ)} \cdot \bar{f}\gamma^\mu\gamma^5 f \ . \tag{53}$$

The contributions of fermions to the current J^μ_{PQ} are proportional to their PQ charges $e_f^{(PQ)}$; these charges are model-dependent. There is necessarily an interaction of axions with gluons, see Eq. (49),

$$L_{ag} = C_g\frac{\alpha_s}{8\pi} \cdot \frac{a}{f_{PQ}} \cdot G^a_{\mu\nu}\tilde{G}^{\mu\nu\, a} \tag{54}$$

Finally, there is an axion-photon coupling

$$L_{a\gamma} = g_{a\gamma\gamma} \cdot aF_{\mu\nu}\tilde{F}^{\mu\nu} \ , \qquad g_{a\gamma\gamma} = C_\gamma \frac{\alpha}{8\pi f_{PQ}} \ , \tag{55}$$

The dimensionless constants C_g and C_γ are model-dependent and, generally speaking, not very much different from 1. The main free parameter is f_{PQ}, while the axion mass is related to it via Eq. (51); numerically,

$$m_a = 6 \ \mu\text{eV} \ \cdot \left(\frac{10^{12} \ \text{GeV}}{f_{PQ}} \right) \ . \tag{56}$$

There are astrophysical bounds on the strength of axion interactions f_{PQ}^{-1} and hence on the axion mass. Axions in theories with $f_{PQ} \lesssim 10^9$ GeV, which are heavier than about 10^{-2} eV, would be intensely produced in stars and supernovae explosions. This would lead to contradictions with observations. So, we are left with very light axions, $m_a \lesssim 10^{-2}$ eV. These very light and very weakly interacting axions are interesting dark matter candidates[10].

7.2 Axions in cosmology

Axions can serve as dark matter if they do not decay within the lifetime of the Universe. The main decay channel of a light axion is the decay into two photons. The axion width is calculated as

$$\Gamma_{a\to\gamma\gamma} = \frac{m_a^3}{4\pi} \left(C_\gamma \frac{\theta}{8\pi f_{PQ}} \right)^2 \ ,$$

where the quantity in parenthesis is the axion-photon coupling, see Eq. (55). We recall the relation in Eq. (51) and obtain an axion lifetime

$$\tau_a = \frac{1}{\Gamma_{a\to\gamma\gamma}} = \frac{64\pi^3 m_\pi^2 f_\pi^2}{C_\gamma^2 \alpha^2 m_a^5} \sim 10^{24} \ \text{s} \ \cdot \left(\frac{\text{eV}}{m_a} \right)^5 \ .$$

By requiring that this lifetime exceeds the age of the Universe, $\tau_a > t_0 \approx 14$ billion years, we find a very weak bound on the mass of an axion as a dark matter candidate, $m_a < 25$ eV.

Thermal production of axions in the early Universe is not very relevant, since even if they were in thermal equilibrium at high temperatures, their thermally produced number density at present is substantially smaller than that of photons and neutrinos, and with their tiny mass they do not contribute much into the energy density[11]. This is a welcome property, since thermally produced axions, if they composed substantial part of dark matter, would be *hot* dark matter, which is ruled out.

There are at least two mechanisms of axion production in the early Universe that can provide not only the right axion abundance, but also small initial velocities of axions. The latter property makes the axion a *cold* dark matter candidate, despite its very small mass.

One mechanism [50–52] is called the misalignment scenario. It assumes that the Peccei–Quinn symmetry is spontaneously broken before the beginning of the hot epoch, $\langle\Phi\rangle \neq 0$. This is indeed the case in the inflationary framework, if f_{PQ} is higher than both the inflationary Hubble parameter (towards the inflation end) and the reheat temperature of the Universe. In this case the axion field (the phase of the field Φ) is homogeneous over the entire visible Universe, and initially it can take any value θ_0 between $-\pi$ and π. As we have seen in Eq. (46), the axion potential is proportional to the quark condensate $\langle\bar{q}q\rangle$. This condensate vanishes at high temperatures, $T \gg \Lambda_{\text{QCD}}$, and the axion potential is negligibly small. As the temperature decreases, the axion potential builds up. This is shown in Fig. 21. Accordingly, the

[10] We note in passing that axions may be heavy instead [49]. This case is irrelevant for dark matter.

[11] If axions were in thermal equilibrium, they contribute to the effective number of "neutrino" species N_{eff}. This contribution, however, is smaller than the current precision [3] of the determination of N_{eff}, which is equal to ±0.17.

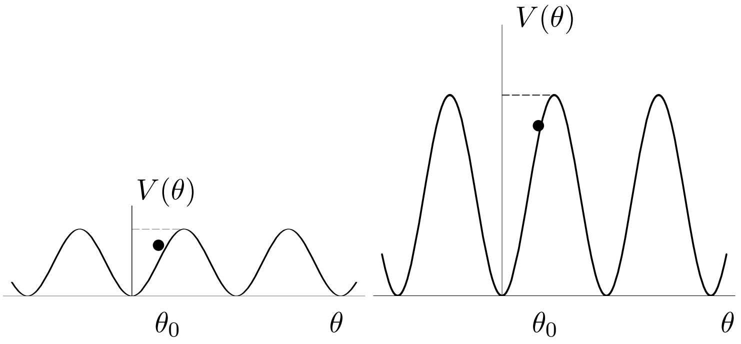


Fig. 21: The axion potential at higher temperature (left) and lower temperature (right). The bullet shows the initial value of the axion field. The field starts to roll down the potential at the time when $m(T) \sim H(T)$.

axion mass increases from zero to m_a; hereafter m_a denotes the zero-temperature axion mass. The axion field practically does not evolve when $m_a(T) \ll H(T)$ and at the time when $m_a(T) \sim H(T)$ it starts to roll down from the initial value θ_0 to the minimum $\theta = 0$ and then it oscillates. During all these stages of evolution, the axion field is homogeneous in space. The homogeneous oscillating field can be interpreted as a collection of scalar quanta with zero spatial momenta, the axion condensate. This is indeed cold dark matter.

Let us estimate the present energy density of the axion field in this picture. The oscillations start at the time t_{osc} when $m_a(t_{osc}) \sim H(t_{osc})$. At this time, the energy density of the axion field is estimated as

$$\rho_a(t_{osc}) \sim m_a^2(t_{osc}) a_0^2 = m_a^2(t_{osc}) f_{PQ}^2 \theta_0^2 \; .$$

The number density of axions at rest at the beginning of oscillations is estimated to be

$$n_a(t_{osc}) \sim \frac{\rho_a(t_{osc})}{m_a(t_{osc})} \sim m_a(t_{osc}) f_{PQ}^2 \theta_0^2 \sim H(t_{osc}) f_{PQ}^2 \theta_0^2 \; .$$

This number density, as any number density of non-relativistic particles, then decreases as a^{-3}. The axion-to-entropy ratio at time t_{osc} is

$$\frac{n_a}{s} \sim \frac{H(t_{osc}) f_{PQ}^2}{\frac{2\pi^2}{45} g_* T_{osc}^3} \cdot \theta_0^2 \simeq \frac{f_{PQ}^2}{\sqrt{g_*} T_{osc} M_{\rm Pl}} \cdot \theta_0^2 \; ,$$

where we use the usual relation $H = 1.66\sqrt{g_*} T^2 / M_{\rm Pl}$. The axion-to-entropy ratio remains constant after the beginning of the oscillations, such that the present mass density of axions is

$$\rho_{a,0} = \frac{n_a}{s} m_a s_0 \simeq \frac{m_a f_{PQ}^2}{\sqrt{g_*} T_{osc} M_{\rm Pl}} s_0 \cdot \theta_0^2 \; . \tag{57}$$

To obtain a simple estimate, let us set $T_{osc} \sim \Lambda_{\rm QCD} \simeq 200$ MeV and make use of Eq. (56). We find

$$\Omega_a \equiv \frac{\rho_{a,0}}{\rho_c} \simeq \left(\frac{10^{-6}\ \text{eV}}{m_a} \right) \theta_0^2 \; . \tag{58}$$

The natural assumption about the initial phase is $\theta_0 \sim \pi/2$. Hence, an axion of mass $m_a = (\text{a few}) \cdot 10^{-6}$ eV is a good dark matter candidate. Note that an axion of lower mass $m_a < 10^{-6}$ eV may also serve as a dark matter particle, if for some reason the initial phase θ_0 is much smaller than $\pi/2$.

A more precise estimate is obtained by taking into account the fact that that the axion mass smoothly depends on temperature:

$$\Omega_a \simeq 0.2 \cdot \theta_0^2 \cdot \left(\frac{4 \cdot 10^{-6}\ \text{eV}}{m_a} \right)^{1.2}$$

We see that our crude estimate in Eq. (58) is fairly accurate.

We note that in the misalignment scenario, and in the inflationary framework, the initial phase θ_0 is not quite homogeneous in space. At the inflationary stage, vacuum fluctuations of all massless or light scalar fields get enhanced. As a result, scalar fields become inhomogeneous on scales exceeding the inflationary Hubble scale H_{infl}^{-1}. The amplitudes of these inhomogeneities (for canonically normalized fields) are equal to $H_{infl}/(2\pi)$. Phase perturbations give rise to perturbations of the axion dark matter energy density, which are uncorrelated with perturbations of conventional matter. These uncorrelated dark matter perturbations are called isocurvature (or entropy) modes. Cosmological observations show that their contribution cannot exceed a few per cent of the dominant adiabatic mode. This leads to a constraint [53] on the inflationary Hubble parameter H_{infl} or, equivalently, on the energy scale of the inflation (energy density of the inflaton field)

$$V_{infl}^{1/4} \lesssim 10^{13}\ \text{GeV} .$$

This makes the misalignment mechanism somewhat contrived. Reversing the argument, detection of the dark matter entropy mode would be an interesting hint towards the nature of dark matter.

Another mechanism of axion production in the early Universe works under the assumption which is opposite to the main assumption of the misalignment scenario. Namely, one assumes that the Peccei–Quinn symmetry is restored at the beginning of the hot epoch, and gets spontaneously broken at a temperature of order $T \sim f_{PQ}$ at the hot stage. Then the phase of the field Φ is uncorrelated at distances exceeding the size of the horizon at that time. In principle, one should be able to predict the value of f_{PQ} and hence m_a in this scenario, since there is no uncertainty in the initial conditions. However, the dynamics in this case is quite complicated. Indeed, the uncorrelated phase gives rise to the production of global cosmic strings [54]—topological defects that exist in theories with a spontaneously broken global $U(1)$ symmetry ($U(1)_{PQ}$ in our case; for a discussion see, e.g., Ref. [55]). At the QCD transition epoch, defects of another type, axion domain walls, are created. Then all these defects get destructed, giving rise to the production of axions. The analysis of this dynamics has been made by various authors, see, e.g., Refs. [56, 57], but it is fair to say that there is no compelling prediction for m_a yet. A reasonable estimate of the axion mass is (Ref. [56] claims $m_a = 2.6 \cdot 10^{-5}$ eV)

$$m_a = (\text{a few}) \cdot 10^{-5}\ \text{eV} .$$

To end up with cosmological aspects of axion dark matter, we note that it has interesting phenomenology in the present Universe. Axions tend to form mini-clusters [58] which can be disrupted and form streams of dark matter [59]. Axions also form Bose-stars [60]. All this exotica is relevant to both astrophysics and the axion search.

7.3 Axion search

The search for dark matter axions with mass $m_a \sim 10^{-5} - 10^{-6}$ eV is difficult, but not impossible. One way is to search for an axion-photon conversion in a resonant cavity filled with a strong magnetic field. Indeed, in the background magnetic field, the axion-photon interaction in Eq. (55) leads to the conversion $a \to \gamma$, see Fig. 22. Axions of mass $10^{-5} - 10^{-6}$ eV are converted to photons of frequency $\nu = m/(2\pi) = 2 - 0.2$ GHz (radio waves; $m = 10^{-6}$ eV $\longleftrightarrow \nu = 240$ MHz). To collect a reasonable number of conversion photons, one needs cavities of a high quality factor Q, which have a small bandwidths. This means that one goes in small steps in m_a, and the whole search takes a long time. This is illustrated in Fig. 23.

The hunt for dark matter axions has been intensified recently. A new set of resonant cavity experiments, CAPP, is under preparation, see Fig. 24. A new approach to search for heavier dark matter axions with $m_a \gtrsim 4 \cdot 10^{-5}$ eV has been suggested by the MADMAX interest group [64]. Other axion search experiments are reviewed in Refs. [63, 65].

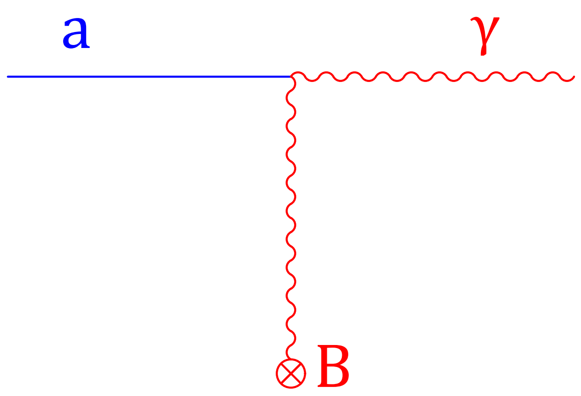


Fig. 22: Axion-photon conversion in magnetic field.

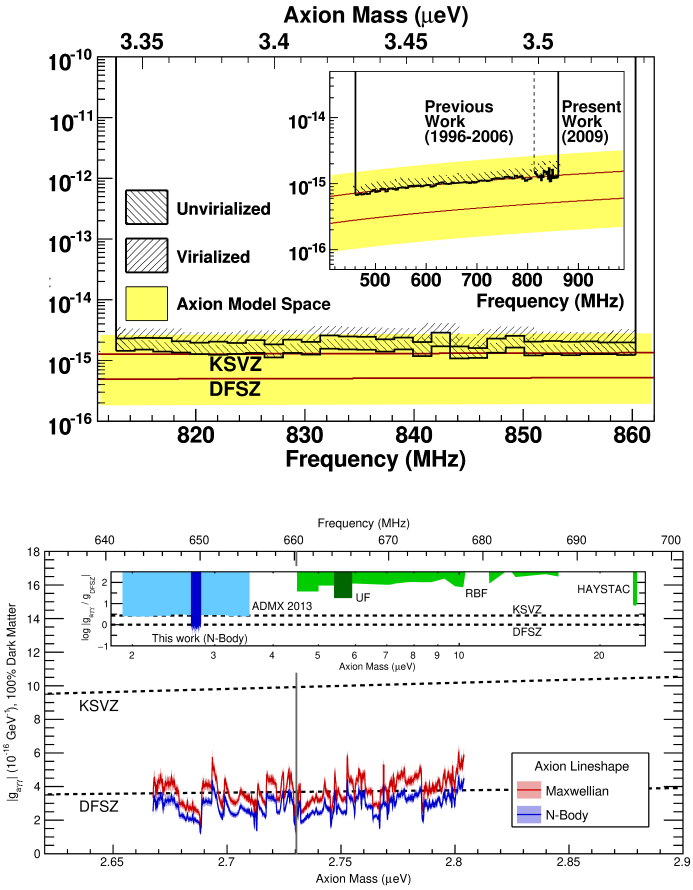


Fig. 23: Limits on the axion-photon coupling for various axion masses. Lines labeled KSVZ and DSVZ refer to predictions of the two axion models under the assumption that axions make the whole of dark matter. Shown are limits published by ADMX collaboration in 2010 [61] (upper panel) and in 2018 [62]. Note the limited ranges of masses spanned during the long period of time. Note also that the recent limits (lower panel) reach almost entire range of axion-photon couplings predicted by various axion models.

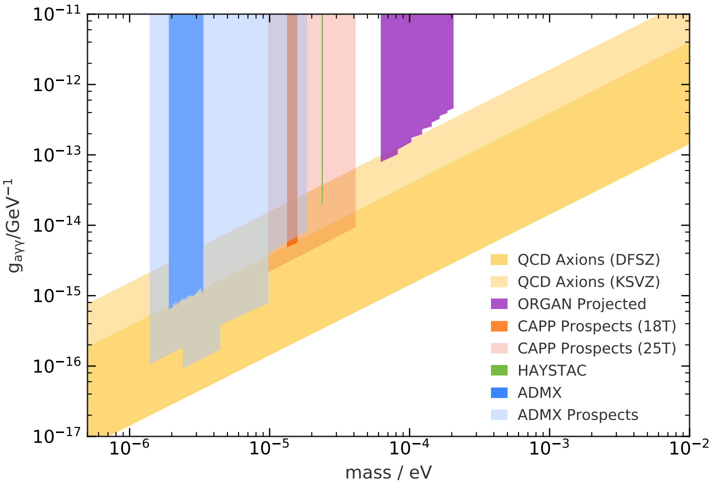


Fig. 24: Future prospects of dark matter axion searches with resonant cavities [63].

7.4 Axion-like particles (ALPs)

There may exist light, weakly interacting scalar or pseudoscalar particles other than axions. They are called axion-like particles, ALPs, and they may emerge as pseudo-Nambu–Goldstone bosons of some new approximate global symmetry. We have discussed one example, fuzzy dark matter, in Section 5.2. Unlike the axion case, where the axion-photon coupling is related, albeit in somewhat model-dependent way, to its mass via Eqs. (55) and (56), the ALP mass and coupling to photons are both arbitrary parameters. Also, ALPs may interact with the Standard Model fermions, and that coupling is again a free parameter. ALPs may or may not be dark matter candidates; searches for them is of interest independently of the dark matter problem.

If the ALP is a dark matter candidate, instruments described in previous subsection—“haloscopes”—are capable for searching for dark matter ALPs, and it makes sense to extend the search to an as wide mass range as possible. In this regard, it is worth mentioning that the CASPEr experiment [66] is going to be sensitive to very light ALPs, $m \lesssim 10^{-9}$ eV, and very small ALP-fermion couplings.

Bounds and prospects for search for light ALPs are summarized Fig. 26.

ALPs may be produced in the Sun, and their flux may be detectable by “helioscopes”, instruments searching for the axion-photon conversion in the magnetic field of a magnet looking at the Sun. One such instrument, CAST, has been operating for a long time, whereas other experiments, IAXO and TASTE,

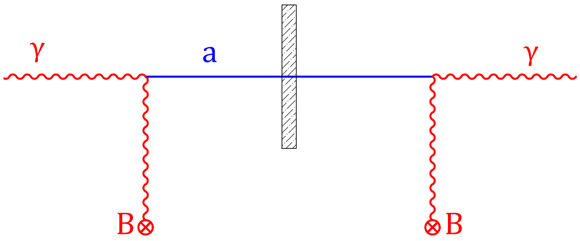


Fig. 25: “Light shining through a wall”: laser light shining from the left is converted into axions in the magnetic field of a magnet placed before the wall, axions pass through the wall and are converted into photons by a magnet behind the wall; the latter photons are detected by a highly sensitive photon detector.

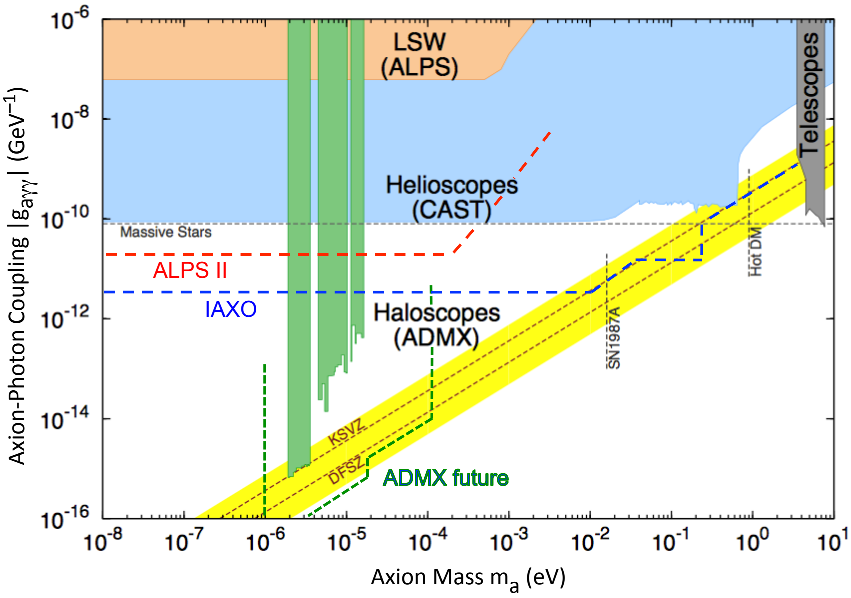


Fig. 26: Bounds on ALPs: ALP-photon coupling vs ALP mass [65]. The inclined straight strip with lines labeled “KSVZ” and “DFSZ” is the range of predictions of axion models. Shaded regions are limits from existing experiments, dashed lines show sensitivities of future searches.

are planned. Another way to search for ALPs makes use of the idea of “light shining through a wall”, see Fig. 25; this idea is implemented in the ALPS-I, ALPS-II experiments. For a review of these approaches see, e.g., Ref. [65]. Finally, ALPs can be searched in beam-dump experiments and in decays of K- and B-mesons. Interesting limits are obtained by the CHARM and BaBar experiments, and a promising planned experiment is SHiP at CERN [67].

8 Warm dark matter: sterile neutrinos

As we discussed in Section 5.1, there are arguments, albeit not yet conclusive, which favor warm, rather than cold, dark matter. If WDM particles were in kinetic equilibrium at some epoch in the early Universe, then their mass should be in the range of $3-10$ keV. Reasonably well motivated particles of this mass are sterile neutrinos.

Sterile neutrinos—massive leptons N which do not participate in the Standard Model gauge interactions—are most probably required for giving masses to ordinary, “active” neutrinos. The masses of sterile neutrinos cannot be predicted theoretically. Although sterile neutrinos of a mass of $m_N = 3-10$ keV are not particularly plausible from the particle physics prospective, they are not pathological either. In the simplest case the creation of sterile neutrino states $|N\rangle$ in the early Universe occurs due to their mixing with active neutrinos $|\nu_\alpha\rangle$, $\alpha = e, \mu, \tau$. In the approximation of mixing between two states only, we have

$$|\nu_\alpha\rangle = \cos\theta|\nu_1\rangle + \sin\theta|\nu_2\rangle \;, \qquad |N\rangle = -\sin\theta|\nu_1\rangle + \cos\theta|\nu_2\rangle \;,$$

where $|\nu_\alpha\rangle$ and $|N\rangle$ are active and sterile neutrino states, $|\nu_1\rangle$ and $|\nu_2\rangle$ are mass eigenstates of masses m_1 and m_2, where we order $m_1 < m_2$, and θ is the vacuum mixing angle between sterile and active neutrino. This mixing should be weak, $\theta \ll 1$, otherwise sterile neutrinos would decay too rapidly, see below. The heavy state is mostly the sterile neutrino $|\nu_2\rangle \approx |N\rangle$, and $m_2 \equiv m_N$ is the sterile neutrino mass.

The calculation of the sterile neutrino abundance is fairly complicated, and we do not reproduce it here. If there is no sizeable lepton asymmetry in the Universe, the estimate is

$$\Omega_N \simeq 0.2 \cdot \left(\frac{\sin\theta}{10^{-4}}\right)^2 \cdot \left(\frac{m_N}{1\ \text{keV}}\right)^2 . \tag{59}$$

The energy spectrum of sterile neutrinos is nearly thermal. Thus, sterile neutrino of mass $m_\nu \gtrsim 1$ keV and small mixing angle $\theta_\alpha \lesssim 10^{-4}$ would serve as a dark matter candidate. However, this range of masses and mixing angles is ruled out. The point is that due to its mixing with an active neutrino, the sterile neutrino can decay into an active neutrino and a photon, see Fig. 27,

$$N \to \nu_\alpha + \gamma .$$

The sterile neutrino decay width is proportional to $\sin^2\theta$. If sterile neutrinos are dark matter particles,

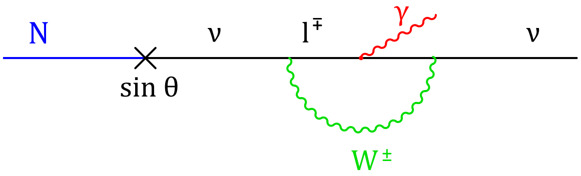


Fig. 27: Sterile neutrino decay $N \to \nu_\alpha + \gamma$.

their decays would produce a narrow line in X-ray flux from the cosmos (the orbiting velocity of dark matter particles in galaxies is small, $v \lesssim 10^{-3}$, hence the photons produced in their two-body decays are nearly monochromatic). Leaving aside a hint towards a 3.5 keV line advocated in Refs. [68, 69] (see the discussion of its status in Ref. [70]), one makes use of strong limits on such a line and translates them into limits on $\sin^2\theta$. These limits as function of the sterile neutrino mass are shown in Fig. 28; they rule out the range of masses that are giving the right mass density of dark matter, Eq. (59).

A (rather baroque) way out [71] is to assume that there is a fairly large lepton asymmetry in the Universe. Then the oscillations of an active neutrino into a sterile neutrino may be enhanced due to the MSW effect, as at some temperature they occur in the Mikheev–Smirnov resonance regime. In that case the right abundance of sterile neutrinos is obtained at smaller θ, and may be consistent with the X-ray bounds. This is also shown in Fig. 28.

Direct laboratory searches for a sterile neutrino are currently sensitive to substantially larger sterile-active mixing angles. This is shown in Fig. 29 and also in Fig. 28, projected KATRIN limit, dashed line.

9 Dark matter summary

In the first place, the mechanisms discussed here are by no means the only ones capable of producing dark matter, and the particles we discussed are by no means the only dark matter candidates. Other dark matter candidates include gravitinos, Q-balls, very heavy relics produced towards the end of inflation (wimpzillas), primordial black holes, etc. Hence, even though there are grounds to hope that the dark matter problem will be solved soon, there is no guarantee at all. Indeed, some of the candidates, like a gravitino or a sterile wimpzilla, interact with Standard Model particles so weakly that their direct discovery is hopeless. Concerning the candidates we have presented, we make a few comments:

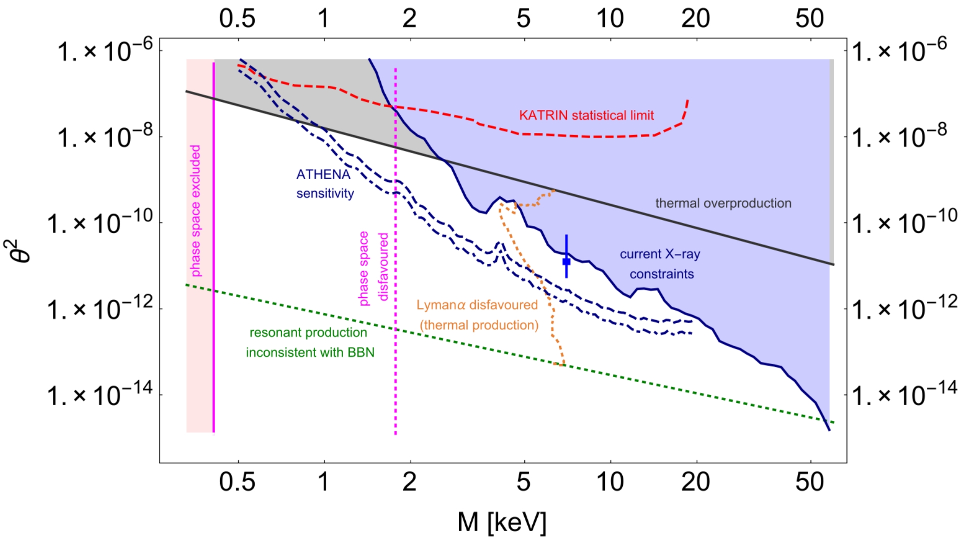


Fig. 28: Limits on sterile neutrino parameters (mass M, mixing angle squared θ^2) obtained from X-ray telescopes [70]. The straight solid line refers to sterile neutrino dark matter produced in non-resonant oscillations, Eq. (59). The region between this line and the dotted line corresponds to the resonant mechanism that works in a Universe with a fairly large lepton asymmetry. Vertical lines show very conservative limits coming from phase space and Lyman-α considerations, see Section 5.1. Regions left of these lines are disfavored. In fact, for a non-resonant mechanism, the phase space constraint is $M \gtrsim 6$ keV. The bullet with vertical interval shows the point corresponding to a putative 3.5 keV line.

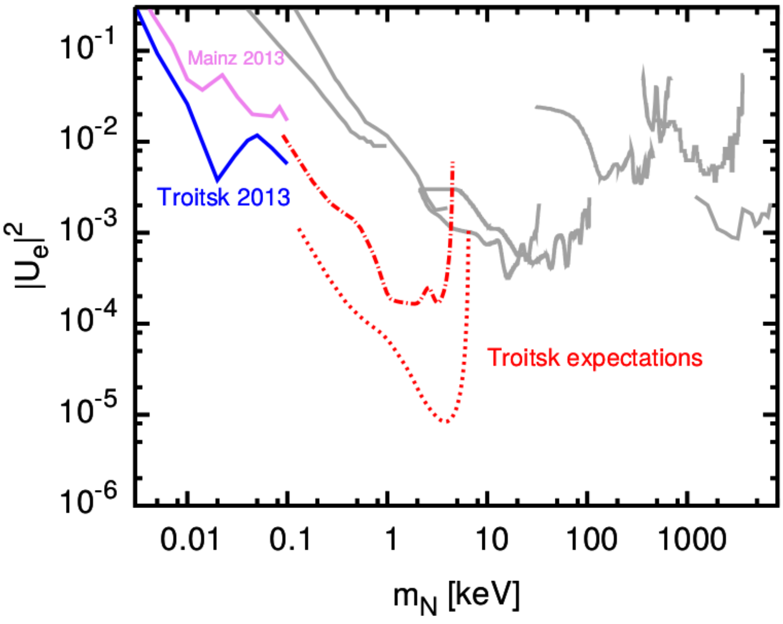


Fig. 29: Existing laboratory limits on a sterile neutrino mixing with an electron neutrino, $|U_e|^2 = \theta^2_{N\,\nu_e}$, and projected sensitivity of the Troitsk nu-mass experiment [72].

– With the exception of axions/ALPs, the plausible candidates are strongly constrained already. However, as we pointed out, this does not mean much, since the actual values of model parameters may still be in the unexplored region of the parameter space.

- The null results obtained so far suggest that it makes sense to look for less motivated candidates, and employ diverse search strategies. This happens already: we note in this regard existing and proposed experiments like NA64, SHiP, Troitsk nu-mass, Katrin, etc.
- Astrophysics and cosmology may well provide hints towards the nature of dark matter (CDM vs WDM vs SIMP vs fuzzy DM, etc.)
- WIMPs are attacked from different directions. If dark matter particles are indeed WIMPs, and the relevant energy scale is of order 1 TeV, then the Hot Big Bang theory will be probed experimentally up to a temperature of $(\text{a few}) \cdot (10-100)$ GeV and down to an age $10^{-9} - 10^{-11}$ s (compare to 1 MeV and 1 s accessible today through Big Bang nucleosynthesis). With microscopic physics to be known from collider experiments, the WIMP abundance will be reliably calculated and checked against the data from observational cosmology. Thus, the WIMP scenario offers a window to a very early stage of the evolution of the Universe.
- Searches for dark matter axions, ALPs or a signal from a light sterile neutrino make use of completely different methods. Yet there is a good chance for discovery, if either of these particles make dark matter.

All this shows that the situation with dark matter is controversial but extremely interesting.

10 Baryon asymmetry of the Universe

As we discussed in Section 2.6, the baryon asymmetry of the Universe is characterized by the baryon-to-entropy ratio, which at high temperatures is defined as follows,

$$\Delta_{\mathrm{B}} = \frac{n_{\mathrm{B}} - n_{\bar{\mathrm{B}}}}{s} = \frac{1}{3}\frac{n_{\mathrm{q}} - n_{\bar{\mathrm{q}}}}{s} \; ,$$

where n_{q} and $n_{\bar{\mathrm{q}}}$ are the number densities of quarks and antiquarks, respectively (baryon number of a quark equals $1/3$), and s is the entropy density. If the baryon number is conserved and the Universe expands adiabatically (which is the case at least after the electroweak epoch, $T \lesssim 100$ GeV), Δ_B is time-independent and equal to its present value $\Delta_B \approx 0.86 \cdot 10^{-10}$, see Eq. (24). At early times, at temperatures well above 100 MeV, the cosmic plasma contained many quark-antiquark pairs, whose number density was of the order of the entropy density, $n_{\mathrm{q}} + n_{\bar{\mathrm{q}}} \sim s$. Hence, in terms of quantities characterizing the very early epoch, the baryon asymmetry may be expressed as

$$\Delta_B \sim \frac{n_{\mathrm{q}} - n_{\bar{\mathrm{q}}}}{n_{\mathrm{q}} + n_{\bar{\mathrm{q}}}} \; .$$

We see that there was one extra quark per about 10 billion quark-antiquark pairs! It is this tiny excess that is responsible for the entire baryonic matter in the present Universe: as the Universe expanded and cooled down, antiquarks annihilated with quarks, and only the excessive quarks remained and formed baryons.

There is no logical contradiction to suppose that the tiny excess of quarks over antiquarks was built in as an initial condition. This would be very contrived, however. Furthermore, the inflationary scenario predicts that the Universe was baryon-symmetric at inflation (no quarks, no antiquarks). Hence, the baryon asymmetry must be explained dynamically [73, 74], by some mechanism of its generation in the early Universe.

10.1 Sakharov conditions

There are three necessary conditions for the generation of a baryon asymmetry from an initially baryon-symmetric state. These are the Sakharov conditions:

(i) baryon number non-conservation;

(ii) C- and CP violation;

(iii) deviation from a thermal equilibrium.

All three conditions are easily understood. (i) If baryon number were conserved, and the initial net baryon number in the Universe vanishes, the Universe today would still be baryon-symmetric. (ii) If C or CP were conserved, then the rate of reactions with particles would be the same as the rate of reactions with antiparticles, and no asymmetry would be generated. (iii) Thermal equilibrium means that the system is stationary (no time-dependence at all). Hence, if the initial baryon number is zero, it is zero forever, unless there are deviations from the thermal equilibrium. Furthermore, if there are processes that violate baryon number, and the system approaches thermal equilibrium, then the baryon number tends to be washed out rather than generated (with a qualification, see below).

At the epoch of the baryon asymmetry generation, all three Sakharov conditions have to be met simultaneously. There is a qualification, however. These conditions would be literally correct if there were no other relevant quantum numbers that characterize the cosmic medium. In reality, however, lepton numbers also play a role. As we will see shortly, baryon and lepton numbers are rapidly violated by anomalous electroweak processes at temperatures above, roughly, 100 GeV. What is conserved in the Standard Model is the combination $(B - L)$, where L is the total lepton number[12]. So, there are two options. One is to generate the baryon asymmetry at or below the electroweak epoch, $T \lesssim 100$ GeV, and make sure that the electroweak processes do not wash out the baryon asymmetry after its generation. This leads to the idea of electroweak baryogenesis (another possibility is Affleck–Dine baryogenesis [75]). Another is to generate $(B - L)$-asymmetry before the electroweak epoch, i.e., at $T \gg 100$ GeV: if the Universe is $(B - L)$-asymmetric above 100 GeV, the electroweak physics reprocesses $(B - L)$ partially into baryon number and partially into lepton number, so that in the thermal equilibrium with conserved $(B - L)$ one has

$$B = C \cdot (B - L) \ , \qquad L = (C - 1) \cdot (B - L) \ , \tag{60}$$

where C is a constant of order 1 ($C = 28/79$ in the Standard Model at $T \gtrsim 100$ GeV). In the second scenario, the first Sakharov condition applies to $(B - L)$ rather than baryon number itself.

There are two most commonly discussed mechanisms of baryon number non-conservation. One emerges in Grand Unified Theories and is due to the exchange of super-massive particles. The scale of these new, baryon number violating interactions is the Grand Unification scale, presumably of the order of $M_{GUT} \simeq 10^{16}$ GeV. It is not very likely, however, that the baryon asymmetry was generated due to this mechanism: the relevant temperature would have to be of order M_{GUT}, and so a high reheat temperature after inflation is difficult to obtain.

Another mechanism is non-perturbative [38] and is related to the triangle anomaly in the baryonic current (a keyword here is "sphaleron" [76, 77]). It exists already in the Standard Model, and, possibly with mild modifications, operates in all its extensions. The two main features of this mechanism, as applied to the early Universe, is that it is effective over a wide range of temperatures, 100 GeV $< T <$ 10^{11} GeV, and, as we pointed out above, that it conserves $(B - L)$. A detailed analysis can be found in the book [78] and in references therein, as well as in the lecture notes of a similar School [31]. Here we only sketch its main ingredients.

10.2 Electroweak baryon number non-conservation

Let us consider the baryonic current,

$$B^\mu = \frac{1}{3} \cdot \sum_i \bar{q}_i \gamma^\mu q_i \ ,$$

where the sum runs over all quark flavors. Naively, it is conserved, but at the quantum level its divergence is non-zero because of the triangle anomaly (we discussed similar effect in the QCD context in

[12] Masses of neutrinos, if Majorana, violate lepton number. This effect, however, is by itself negligible.

Section 7.1; there, the axial current J_A^μ is not conserved even in the chiral limit),

$$\partial_\mu B^\mu = \frac{1}{3} \cdot 3_{colors} \cdot 3_{generations} \cdot \frac{g^2}{16\pi^2} F^a_{\mu\nu} \tilde{F}^{a\,\mu\nu} \ ,$$

where $F^a_{\mu\nu}$ and g are the field strength of the $SU(2)_W$ gauge field and the $SU(2)_W$ gauge coupling, respectively, and $\tilde{F}^{a\,\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\lambda\rho}F^a_{\lambda\rho}$ is the dual tensor, cf. Eq. (45). Likewise, each leptonic current ($\alpha = e, \mu, \tau$) is anomalous in the Standard Model (we disregard here neutrino masses and mixings, which violate lepton numbers too),

$$\partial_\mu L^\mu_\alpha = \frac{g^2}{16\pi^2} F^a_{\mu\nu} \tilde{F}^{a\,\mu\nu} . \tag{61}$$

A non-trivial fact is that there exist large field fluctuations, $F^a_{\mu\nu}(\mathbf{x}, t) \propto g^{-1}$, such that

$$Q \equiv \int d^3x dt \, \frac{g^2}{16\pi^2} \cdot F^a_{\mu\nu} \tilde{F}^{a\,\mu\nu} \neq 0 \ . \tag{62}$$

Furthermore, for any physically relevant fluctuation, the value of Q is integer ("physically relevant" means that the gauge field strength vanishes at infinity in space-time). In four space-time dimensions such fluctuations exist only in *non-Abelian* gauge theories.

Suppose now that a fluctuation with non-vanishing Q has occurred. Then the baryon numbers in the end and beginning of the process are different,

$$B_{fin} - B_{in} = \int d^3x dt \, \partial_\mu B^\mu = 3Q \ . \tag{63}$$

Likewise

$$L_{\alpha,\,fin} - L_{\alpha,\,in} = Q \ . \tag{64}$$

This explains the selection rule mentioned above: B is violated, $(B - L) \equiv (B - \sum_\alpha L_\alpha)$ is not.

At zero temperature, the field fluctuations that induce baryon and lepton number violation are vacuum fluctuations are called instantons [79]. Since these are *large* field fluctuations, their probability is exponentially suppressed. The suppression factor in the Standard Model is[13]

$$\mathrm{e}^{-\frac{16\pi^2}{g^2}} \sim 10^{-165} \ .$$

Therefore, the rate of baryon number violating processes at zero temperature is totally negligible. On the other hand, at high temperatures there are large *thermal* fluctuations ("sphalerons") whose rate is not necessarily small. And, indeed, B-violation in the early Universe is rapid as compared to the cosmological expansion at sufficiently high temperatures, provided that (see Ref. [80] for details)

$$\langle\phi\rangle_T < T \ , \tag{65}$$

where $\langle\phi\rangle_T$ is the Englert–Brout–Higgs expectation value at temperature T.

10.3 Electroweak baryogenesis: what can make it work

Rapid electroweak baryon number non-conservation at high temperatures appears to open up an intriguing possibility that the baryon asymmetry was generated just by these electroweak processes. This should occur at electroweak temperatures, $T_{EW} \sim 100$ GeV, since whatever baryon asymmetry is generated by electroweak processes at higher temperatures, it would be washed out by the same processes as the Universe cools down to T_{EW}. There are two obstacles, however:

[13] Similar fluctuations of the gluon field in QCD are not suppressed, since QCD is strongly coupled at low energies. This explains why the axial current J_A^μ is not conserved, even approximately.

– CP violation (2nd Sakharov condition) is too weak in the Standard Model: the CKM mechanism alone is insufficient to generate an realistic value of the baryon asymmetry.
– Departure from thermal equilibrium (3d Sakharov condition) is problematic as well. At temperatures of order $T_{EW} \sim 100$ GeV, the Universe expands very slowly: the cosmological time scale at these temperatures,
$$H^{-1}(T_{EW}) = \frac{M^*_{\rm Pl}}{T^2_{EW}} \sim 10^{-10}\ {\rm s}\ , \qquad (66)$$
is very large by the electroweak physics standards.

Let us discuss what can make the electroweak mechanism work. We begin with the second obstacle. It appears that the only way to have strong departure from a thermal equilibrium at $T_{EW} \sim 100$ GeV is a first order phase transition. Indeed, at temperatures well above 100 GeV electroweak symmetry is restored, and the expectation value of the Englert–Brout–Higgs field ϕ is zero, while it is non-zero in vacuum.

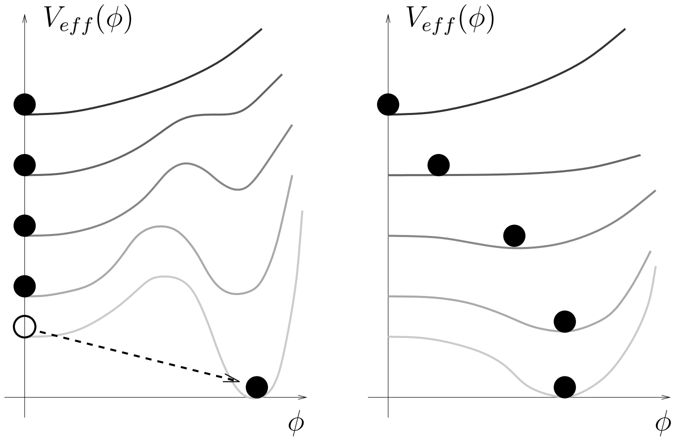


Fig. 30: Effective potential as function of ϕ at different temperatures. Left: first order phase transition. Right: second order phase transition. Upper curves correspond to higher temperatures. Black blobs show the expectation value of ϕ in thermal equilibrium. The arrow in the left panel illustrates the transition from the metastable, supercooled state to the ground state.

This suggests that there may be a phase transition from the phase with $\langle\phi\rangle = 0$ to the phase with $\langle\phi\rangle \neq 0$. In fact, the situation is subtle here, as ϕ is not gauge invariant, and hence cannot serve as an order parameter, so the notion of phases with $\langle\phi\rangle = 0$ and $\langle\phi\rangle \neq 0$ is vague. This is similar to a liquid-vapor system, which does not have an order parameter and, depending on the pressure, may or may not undergo a vapor-liquid phase transition as the temperature decreases.

Continuing to use somewhat sloppy terminology, we recall that in thermal equilibrium any system is at the global minimum of its *free energy*. To figure out the expectation value of ϕ at a given temperature, one introduces the temperature-dependent effective potential $V_{\rm eff}(\phi; T)$, which is equal to the free energy density in the system under the constraint that the average field is equal to a prescribed value ϕ, but otherwise there is thermal equilibrium. Then the global minimum of $V_{\rm eff}$ at a given temperature is at the equilibrium value of ϕ, while local minima correspond to metastable states.

The interesting case for us is a first order phase transition. In this case, the system evolves as follows. At high temperatures, there exists one minimum of $V_{\rm eff}$ at $\phi = 0$, and the expectation value of the Englert–Brout–Higgs field is zero. As the temperature decreases, another minimum appears at finite ϕ, and then becomes lower than the minimum at $\phi = 0$, see left panel of Fig. 30. However, the minima

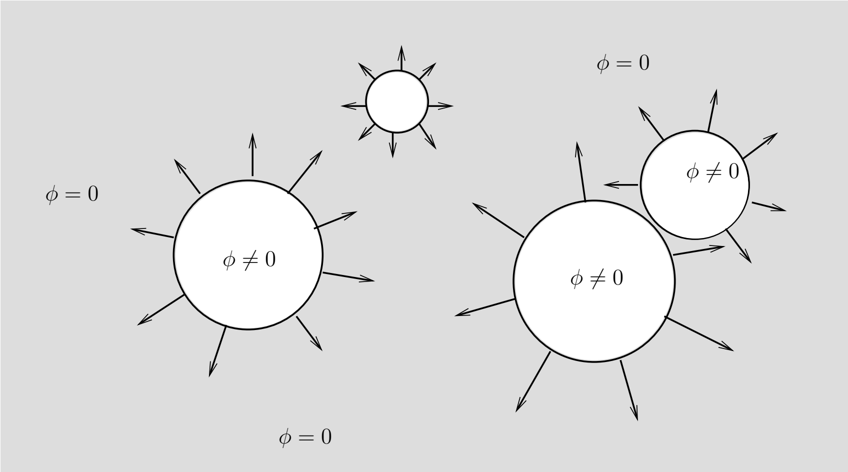


Fig. 31: First order phase transition: a boiling Universe.

with $\phi = 0$ and $\phi \neq 0$ are separated by a barrier of V_{eff}, the probability of the transition from the phase $\phi = 0$ to the phase $\phi \neq 0$ is very small for some time, and the system gets overcooled. The transition occurs when the temperature becomes sufficiently low, and the transition probability sufficiently high. This is to be contrasted to the case, e.g., of the second order phase transition, right panel of Fig. 30. In the latter case, the field slowly evolves, as the temperature decreases, from zero to non-zero vacuum value, and the system remains very close to thermal equilibrium at all times.

The dynamics of a first order phase transition is highly inequilibrium. Thermal fluctuations spontaneously create bubbles of the new phase inside the old phase. These bubbles then grow, their walls eventually collide, and the new phase finally occupies the entire space. The Universe boils, Fig. 31. In the cosmological context, this process happens when the bubble nucleation rate per Hubble time per Hubble volume is roughly of order 1, i.e., when a few bubbles are created in a Hubble volume in Hubble time. The velocity of the bubble wall in the relativistic cosmic plasma is roughly of the order of the speed of light (in fact, it is somewhat smaller, a factor from 0.1 to 0.01). Hence, the bubbles grow large before their walls collide: their size at collision is roughly of order of the Hubble size (in fact, one or two orders of magnitude smaller). In other words, the bubbles are born microscopic, their initial sizes are determined by the electroweak scale and are roughly of order

$$R_{init} \sim (100 \text{ GeV})^{-1} \sim 10^{-16} \text{ cm} .$$

Their final sizes at the time the bubble walls collide are of order

$$R_{fin} \sim 10^{-2} - 10^{-3} \text{ cm} ,$$

as follows from (66). One may hope that the baryon asymmetry may be generated during this inequilibrium process.

Does this really happen in the Standard Model? Unfortunately, no: with the Higgs boson mass $m_H = 125$ GeV, there is no phase transition in the Standard Model at all; there is smooth crossover instead [81].

Nevertheless, the first order phase transition may be characteristic of some extensions of the Standard Model. Generally speaking, one needs the existence of new bosonic fields that have large enough couplings to the Englert–Brout–Higgs field(s). To have an effect on the dynamics of the transition, the new bosons must be present in the cosmic plasma at the transition temperature, $T_{EW} \sim 100$ GeV, so their masses should not be very much higher than T_{EW}.

Let us turn to the first obstacle, CP violation. In the course of the first order phase transition, the baryon asymmetry is generated in the interactions of quarks and leptons with the bubble walls. Therefore, CP violation must occur at the walls. Now, the walls are made of the scalar field(s), and this points towards the necessity of CP violation in the scalar sector, which may only be the case in a theory containing scalar fields other than the Standard Model Englert–Brout–Higgs field.

In concrete models with successful electroweak baryogenesis, CP violation responsible for the baryon asymmetry often leads to sizeable electric dipole moments (EDMs) of neutron and electron. The limits on EDMs are so strong that many such models are actually ruled out. An example is the non-minimal split supersymmetric Standard Model, for which only a few years ago successful electroweak baryogenesis [82] was demonstrated. The predictions of this model for electron EDM are shown in Fig. 32. In 2016, when Ref. [82] was written, part of the parameter space was still allowed, but the recent ACME limit [83]

$$d_e < 1.1 \cdot 10^{-29} e \cdot \text{cm}$$

rules out the entire parameter space with efficient electroweak baryogenesis.

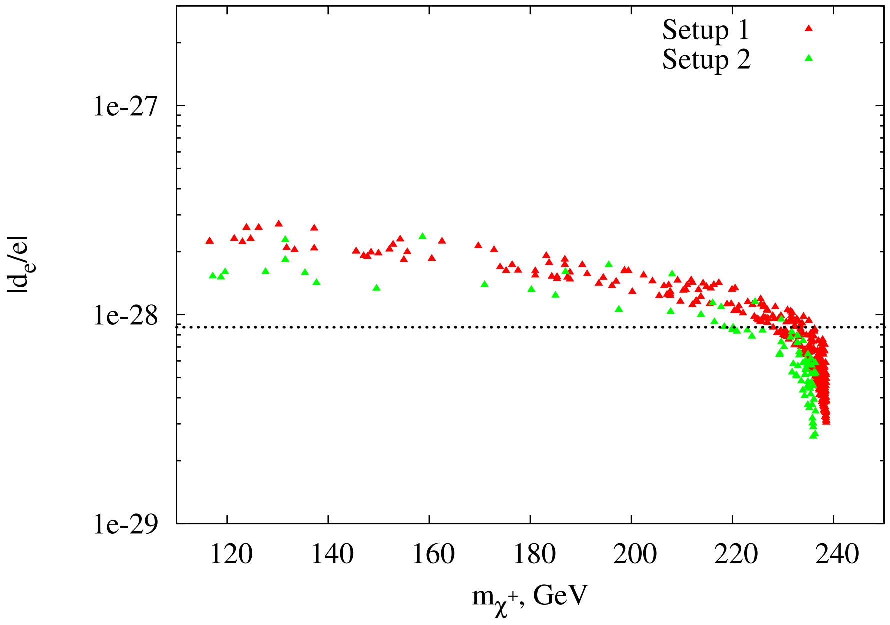


Fig. 32: Electron EDM predicted by the Non-Minimal Split Supersymmetric Standard Model with parameters suitable for electroweak baryogenesis. The current limit $d_e < 1.1 \cdot 10^{-29} e \cdot \text{cm}$ rules out all these models.

To summarize, electroweak baryogenesis requires a considerable extension of the Standard Model, often with masses of new particles in the TeV range or lower. Hence, this mechanism will most likely be ruled out or confirmed by the LHC or its successors. Moreover, limits on electron and neutron EDMs make the design of such an extension very difficult. Still, the issue is not decided yet, and the effort to construct the models with successful electroweak baryogenesis continues [84].

10.4 Baryogenesis in sterile neutrino oscillations

Let us mention another baryogenesis mechanism interesting from the viewpoint of terrestrial experiments, namely, leptogenesis in oscillations of sterile neutrinos [86, 87]. The general idea of leptogenesis [85] is that one or another mechanism generates a *lepton* asymmetry in the Universe before the electroweak transition, and electroweak sphalerons automatically reprocess part of the lepton asymmetry into a baryon asymmetry, see Eq. (60). The particular version of leptogenesis that we briefly discuss here assumes that there are at least two heavy Majorana neutrinos in the mass range $1 - 10$ GeV, and that there is strong enough CP violation in the sterile neutrino sector. Then asymmetries in the sterile neutrino sector may be generated and transmitted to active neutrino sector via Yukawa interactions responsible for see-saw masses of the active neutrinos. In the case when there are effectively two sterile neutrino species participating in leptogenesis, the correct value of the baryon asymmetry is obtained when the two sterile neutrinos are nearly degenerate,

$$\frac{|M_1^2 - M_2^2|}{M_{1,2}^2} \lesssim 10^{-6} ,$$

which makes the model rather contrived. However, with three sterile neutrino species, the degeneracy is no longer required [88]. The sterile neutrinos of masses in the GeV range and parameters suitable for leptogenesis in their oscillations are typically accessible through rare decays of B-mesons, Z-bosons, as well as in future beam dump experiments such as SHiP.

An important point concerning this and virtually all other leptogenesis mechanisms is that CP violation in the sector of active neutrinos, which will hopefully be discovered in oscillation experiments, does not have direct relevance to the leptogenesis: the value of lepton and hence baryon asymmetry is determined by the CP-violating parameters *in the sterile neutrino sector.*

10.5 Baryogenesis summary

We briefly considered here two mechanisms of baryogenesis which may be directly tested, at least in principle, in particle physics experiments. These are certainly not the only mechanisms proposed, and, arguably, not the most plausible mechanisms. One particularly strong competitor is thermal leptogenesis [85], for reviews see, e.g., Ref. [89]. Its idea is that the lepton asymmetry is generated in decays of heavy Majorana sterile neutrinos. The masses of these new particles are well above the experimentally accessible energies. On the one hand, this is in line with the see-saw idea; on the other, direct proof of this mechanism does not appear possible. Interestingly, thermal leptogenesis works only with light active neutrinos: the neutrino masses inferred from cosmology and oscillation experiments are just in the right ballpark.

There are numerous alternative mechanisms of baryogenesis. To name a few, we have already mentioned the Affleck–Dine baryogenesis [75]; early discussions concentrated mostly on GUT baryogenesis [90]; there is even a possibility to generate the baryon asymmetry at the inflationary epoch [91]. Unfortunately, most of these proposals will be very difficult, if at all possible, to test. So, there is no guarantee at all that we will understand in foreseeable future the origin of matter in the Universe.

11 Before the hot epoch

With the Big Bang nucleosynthesis theory and observations, and due to evidence, albeit indirect, for relic neutrinos, we are confident of the theory of the early Universe at temperatures up to $T \simeq 1$ MeV, which correspond to age of $t \simeq 1$ s. With the LHC, we are learning the Universe up to temperatures $T \sim 100$ GeV and down to an age of $t \sim 10^{-10}$ s. Are we going to have a handle on even earlier epoch?

Let us summarize the current status of this issue.

- On the one hand, we are confident that the hot cosmological epoch was not the first one; it was preceded by some other, entirely different stage.
- On the other hand, we do not know for sure what was that earlier epoch; an excellent guess is inflation, but alternative scenarios are not ruled out.
- It is conceivable (although not guaranteed) that future cosmological observations will enable us to understand the nature of the pre-hot epoch.

All this makes the situation very interesting. It is fascinating that by studying the Universe at large we may be able to learn about the earliest cosmological epoch which happened at an extremely high energy density and expansion rate of our Universe.

11.1 Cosmological perturbations

The key players in this Section are cosmological perturbations. These are inhomogeneities in the energy density and associated gravitational potentials, in the first place. It is these inhomogeneities that, among other things, serve as seeds for structures—galaxies, clusters of galaxies, etc. This type of inhomogeneities is called scalar perturbations, as they are described by 3-scalars. There may exist perturbations of another type, called tensor; these are primordial gravity waves. Tensor modes have not been observed (yet), so we mostly concentrate on scalar perturbations. While perturbations of a present size of the order of 10 Mpc and smaller have large amplitudes today and are non-linear, amplitudes of all known perturbations were small in the past, and a linearized theory is applicable. Indeed, CMB temperature anisotropy tells us that the perturbations at the recombination epoch were roughly at the level of

$$\delta \equiv \frac{\delta\rho}{\rho} = 10^{-4} - 10^{-5} \ . \tag{67}$$

We are sloppy here in characterizing the scalar perturbations by the density contrast $\delta\rho/\rho$; we are going to skip technicalities and use this notation in what follows.

Linearized perturbations are most easily studied in momentum space, since the background FLRW metric in Eq. (1) does not explicitly depend on $\mathbf{x}$. The spatial Fourier transformation reads

$$\delta(\mathbf{x}, t) = \int e^{i\mathbf{kx}}\delta(\mathbf{k}, t)\; d^3k \ .$$

Each Fourier mode $\delta(\mathbf{k}, t)$ obeys its own linearized equation and hence can be treated separately. Note that the physical distance between neighboring points is $a(t)d\mathbf{x}$. Thus, $\mathbf{k}$ is *not* the physical momentum (wavenumber); the physical momentum is $\mathbf{k}/a(t)$. While for a given mode the comoving (or coordinate) momentum $\mathbf{k}$ remains constant in time, the physical momentum gets redshifted as the Universe expands, see also Section 2.1. In what follows we set the present value of the scale factor equal to 1, $a_0 \equiv a(t_0) = 1$; then $\mathbf{k}$ is the *present* physical momentum and $2\pi/k$ is the present physical wavelength, which is also called comoving wavelength.

Properties of scalar perturbations are measured in various ways. Perturbations of fairly large spatial scales (fairly low $\mathbf{k}$) give rise to a CMB temperature anisotropy and polarization, so we have very detailed knowledge of them. Somewhat shorter wavelengths are studied by analysing distributions of galaxies and quasars at present and in the relatively near past. There are several other methods, some of which can probe even shorter wavelengths. There is good overall consistency of the results obtained by different methods, so we have a reasonably good understanding of many aspects of the scalar perturbations.

The cosmic medium in our Universe has several components that interact only gravitationally: baryons, photons, neutrinos, dark matter. Hence, there may be and, in fact, there are perturbations in each of these components. As we pointed out in Section 4, electromagnetic interactions between baryons, electrons and photons were strong before recombination, so to a reasonable approximation these species made a single fluid, and it is appropriate to talk about perturbations in this fluid. After recombination, baryons and photons evolved independently.

11.2 Subhorizon and superhorizon regimes

It is instructive to compare the wavelength of a perturbation with the horizon size. To this end, recall (see Section 2.6) that the horizon size $l_{\mathrm{H}}(t)$ is the size of the largest region which is causally connected by the time t, and that

$$l_{\mathrm{H}}(t) \sim H^{-1}(t) \sim t$$

at the radiation domination epoch and later, see Eq. (18). The latter relation, however, holds *under the assumption that the hot epoch was the first one in cosmology*, i.e., that the radiation domination started right after the Big Bang. This assumption is at the heart of what can be called the hot Big Bang theory. We will find that this assumption in fact is *not valid* for our Universe; we are going to see this ad absurdum, so let us stick to the hot Big Bang theory for the time being.

The physical wavelength of a perturbation grows slower than the horizon size. As an example, at the radiation domination epoch

$$\lambda(t) = \frac{2\pi a(t)}{k} \propto \sqrt{t} ,$$

while at the matter domination epoch $\lambda(t) \propto t^{2/3}$. For an obvious reason, the modes with $\lambda(t) \ll H^{-1}(t)$ and $\lambda(t) \gg H^{-1}(t)$ are called subhorizon and superhorizon at the time t, respectively. We are interested in the modes which are subhorizon *today*; longer modes are homogeneous throughout the visible Universe and are not observed. However, *the wavelengths which are subhorizon today were superhorizon at some earlier epoch.* In other words, the physical momentum $k/a(t)$ was smaller than $H(t)$ at early times; at the time $t_\times$ such that

$$q(t_\times) \equiv \frac{k}{a(t_\times)} = H(t_\times) ,$$

the mode entered the horizon, and after that evolved in the subhorizon regime $k/a(t) \gg H(t)$. It is straightforward to see that for all cosmologically interesting wavelengths, the horizon crossing occurs at temperatures below 1 MeV, i.e., at the time we are confident about (repeating the calculation of Section 5.1 we find that the present wavelength of order 100 kpc entered the horizon at $T \sim 4$ keV). So, there is no guesswork at this point.

Another way to look at the superhorizon–subhorizon behaviour of perturbations is to introduce a new time coordinate (cf. Eq. (16)),

$$\eta = \int_0^t \frac{dt'}{a(t')} . \tag{68}$$

Note that this integral converges at the lower limit in the hot Big Bang theory. In terms of this time coordinate, the FLRW metric in Eq. (1) reads

$$ds^2 = a^2(\eta)(d\eta^2 - d\mathbf{x}^2) .$$

In coordinates $(\eta, \mathbf{x})$, the light cones $ds = 0$ are the same as in Minkowski space, and η is the coordinate size of the horizon, see Fig. 33. Every mode of perturbation has the time-independent coordinate wavelength $2\pi/k$, and at small η it is in the superhorizon regime, $2\pi/k \gg \eta$.

11.3 Hot epoch was not the first

This picture falsifies the hot Big Bang theory. Indeed, within this theory, we see the horizon at recombination $l_{\rm H}(t_{\rm rec})$ at an angle $\Delta\theta \approx 2^\circ$, as schematically shown in Fig. 33. By causality, at recombination there should be no perturbations of larger wavelengths, as any perturbation can be generated within the causal light cone only. In other words, the CMB temperature must be isotropic when averaged over angular scales exceeding 2°; there should be no cold or warm regions of an angular size larger than 2°.

We now take a look at the CMB photographic picture shown in Fig. 2. It is seen by naked eye that there are cold and warm regions whose angular size much exceeds 2°; in fact, there are perturbations of all angular sizes up to those comparable to the entire sky. We come to an important conclusion: the scalar perturbations were built in at the very beginning of the hot epoch, i.e., the cosmological perturbations were generated before the hot epoch.

Another manifestation of the fact that the scalar perturbations were there already at the beginning of the hot epoch is the existence of peaks in the angular spectrum of the CMB temperature, as seen in

Fig. 3. In general, perturbations in the baryon-photon medium before recombination are acoustic waves (cf. Section 4.3),

$$\delta_B(\mathbf{k},t) = A(\mathbf{k})e^{i\mathbf{kx}}\cos\left[\int_0^t v_s \frac{k}{a(t')}dt' + \psi_{\mathbf{k}}\right] , \qquad (69)$$

where v_s is the sound speed, $A(\mathbf{k})$ is the time-independent amplitude and $\psi_{\mathbf{k}}$ is a time-independent phase. This expression is valid, however, in the subhorizon regime only, i.e., at late times. The two solutions in the superhorizon regime at the radiation domination epoch are

$$\delta_B(t) = \text{const} , \qquad (70a)$$

$$\delta(t)_B = \frac{\text{const}}{t^{3/2}} . \qquad (70b)$$

If the perturbations existed at the very beginning of the hot epoch, they were in the superhorizon regime at sufficiently early times, and were described by the solutions in Eq. (70). The consistency of the whole cosmology requires that the amplitude of the perturbations was small at the beginning of the hot stage. The solution in Eq. (70b) rapidly decays away, and towards the horizon entry the perturbation is in constant mode in Eq. (70a). So, the initial condition for the further evolution is unique modulo the amplitude $A(\mathbf{k})$, and hence the phase $\psi(\mathbf{k})$ is uniquely determined: we have $\psi(\mathbf{k}) = 0$ for modes entering horizon at the radiation domination epoch. As discussed in Section 4.3, this leads to oscillatory behavior of baryon-photon perturbations at the recombination epoch *as function of* k, and translates into oscillations of the CMB temperature multipole C_l as function of multipole number l.

Were the perturbations generated in a causal way at the radiation domination epoch, they would be always in the subhorizon regime. In that case the solutions in Eq. (70) would be irrelevant, and there would be no reason for a particular choice of phase $\psi_{\mathbf{k}}$ in Eq. (69). One would rather expect that $\psi_{\mathbf{k}}$ is a random function of $\mathbf{k}$, so $\delta_B(\mathbf{k}, t_r)$ would not oscillate as function of $\mathbf{k}$, and oscillations of C_l would not exist. This is indeed the case for specific mechanisms of the generation of density perturbations at hot epoch [92].

We conclude that the facts that the CMB angular spectrum has oscillatory behavior and that there are sizeable temperature fluctuations at $l < 50$ (angular scale greater than the angular size 2° of the

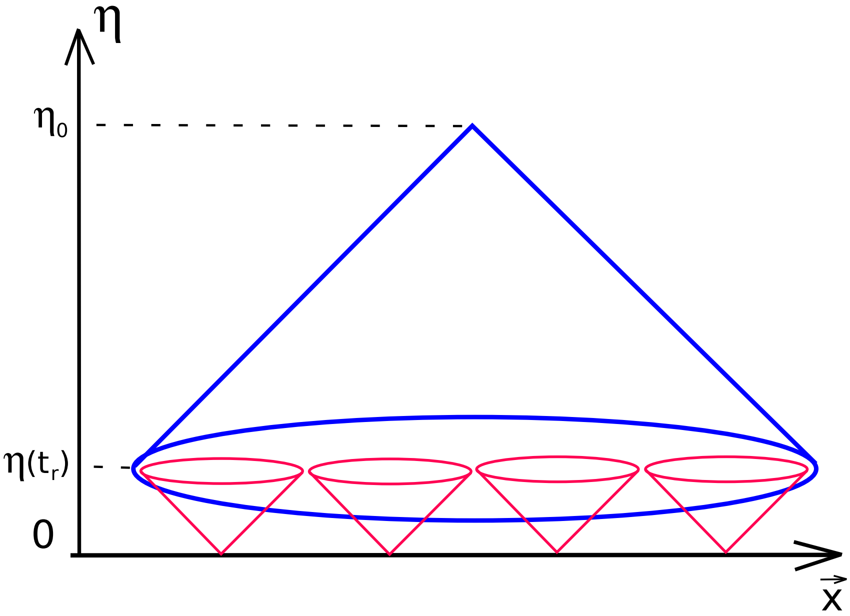


Fig. 33: Causal structure of space-time in the hot Big Bang theory. η_r and η_0 are conformal times at recombination and today, respectively.

horizon at the recombination epoch) unambiguously tell us that the density perturbations were indeed in the superhorizon regime at the hot cosmological stage. The hot epoch was preceded by some other epoch—the epoch of the generation of perturbations.

11.4 Inflation or not?

The pre-hot epoch must be long in terms of the time variable η introduced in Eq. (68). What we would like to have is that the large part of the Universe be causally connected towards the end of that epoch, see Fig. 34. Long duration in η does not necessarily mean long duration in physical time t; in fact, the pre-hot epoch may be very short in physical time.

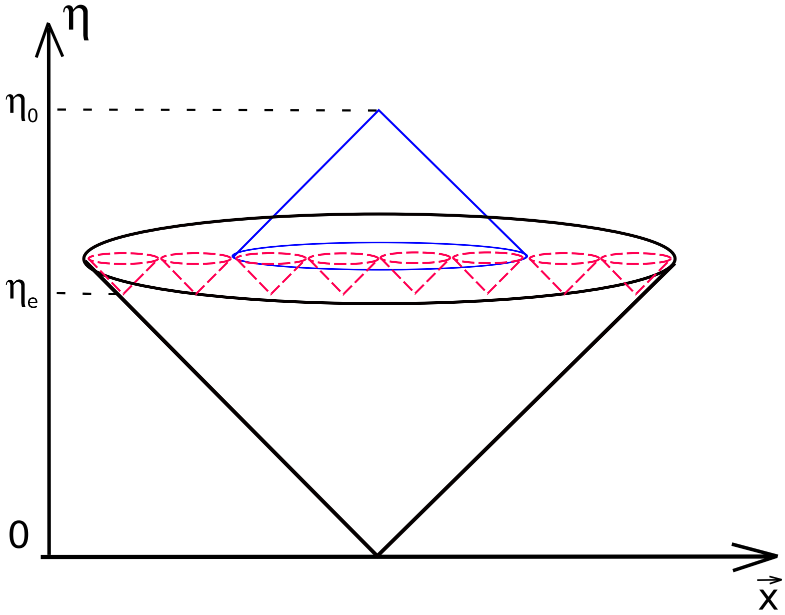


Fig. 34: Causal structure of space-time in the real Universe

An excellent hypothesis on the pre-hot stage is inflation, the epoch of nearly exponential expansion [93],

$$a(t) = \mathrm{e}^{\int H dt} , \qquad H \approx \text{const} .$$

If this epoch lasts many Hubble times, the whole visible Universe, and likely much a greater region of space, is causally connected already at very early times.

From the viewpoint of perturbations, the physical momentum $q(t) = k/a(t)$ decreases (gets redshifted) at the time of inflation, while the Hubble parameter stays almost constant. So, every mode is first in the subhorizon regime ($q(t) \gg H(t)$), and later in the superhorizon regime ($q(t) \ll H(t)$). This situation is opposite to what happens at the radiation and matter domination epoch; this is precisely the pre-requisite for generating the density perturbations. Indeed, inflation does generate primordial density perturbations [94], whose properties are consistent with everything we know about them.

Inflation is not the only hypothesis proposed so far. One alternative option is the bouncing Universe scenario, which assumes that the cosmological evolution begins from contraction, then the contracting stage terminates at some moment of time (bounce) and is followed by expansion. A version is the cycling Universe scenario with many cycles of contraction–bounce–expansion, see Ref. [95] for reviews. Another scenario is that the Universe starts out from nearly flat and static state with nearly vanishing energy density. Then the energy density increases (!), and according to the Friedmann equation, the expansion speeds up. This goes under the name of the Genesis scenario [96]. Theoretical realizations of these scenarios are surprisingly difficult, but not impossible, as became clear recently.

12 Towards understanding the earliest epoch

Since cosmological perturbations originate from the earliest epoch that occurred before the hot stage, properties of these perturbations will hopefully give us a clue on that epoch. Presently, we know only very basic things about the cosmological perturbations. Let us discuss this point, and at the same time consider promising directions where further study may lead to breakthrough.

Of course, since the properties we know of are established by observations, they are valid within certain error bars. Conversely, deviations from the results listed below, if observed, would be extremely interesting.

12.1 Adiabaticity of scalar perturbations

Primordial scalar perturbations are **adiabatic**. This means that there are perturbations in the energy density, but *not in composition*. More precisely, the baryon to entropy ratio and the dark matter to entropy ratio are constant in space,

$$\delta\left(\frac{n_B}{s}\right) = \text{const} \, , \qquad \delta\left(\frac{n_{\rm DM}}{s}\right) = \text{const} \, . \tag{71}$$

This is consistent with the generation of the baryon asymmetry and of dark matter at the hot cosmological epoch: in that case, all particles were in thermal equilibrium early at the hot epoch, and as long as physics behind the baryon asymmetry and dark matter generation is the same everywhere in the Universe, the baryon and dark matter abundances (relative to the entropy density) are necessarily the same everywhere. In principle, there may exist *entropy* (or isocurvature) perturbations that violate (one of) the relations in Eq. (71). No admixture of the entropy perturbations have been detected so far, but it is worth emphasizing that even small admixture will show that many popular mechanisms for generating dark matter and/or baryon asymmetry have nothing to do with reality. One will have to think, instead, that the baryon asymmetry and/or dark matter were generated before the beginning of the hot stage. A notable example is the axion misalignment mechanism discussed in Section 7.

12.2 Gaussianity

The primordial scalar perturbations are a **Gaussian random field**. Gaussianity means that the three-point and all odd correlation functions vanish, while the four-point and higher order even correlation functions are expressed through the two-point function via Wick's theorem:

$$\begin{aligned} \langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle &= 0 \\ \langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\delta(\mathbf{k}_4)\rangle &= \langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\rangle\cdot\langle\delta(\mathbf{k}_3)\delta(\mathbf{k}_4)\rangle \\ &\quad + \text{permutations of momenta} \, . \end{aligned}$$

We note that this property is characteristic of *vacuum fluctuations of non-interacting (linear) quantum fields.* A free quantum field has the general form

$$\phi(\mathbf{x},t) = \int d^3k e^{-i\mathbf{k}\mathbf{x}} \left(f_{\mathbf{k}}^{(+)}(t) a_{\mathbf{k}}^\dagger + e^{i\mathbf{k}\mathbf{x}} f_{\mathbf{k}}^{(-)}(t) a_{\mathbf{k}} \right) \, ,$$

where $a_{\mathbf{k}}^\dagger$ and $a_{\mathbf{k}}$ are creation and annihilation operators. For the field in Minkowski space-time one has $f_{\mathbf{k}}^{(\pm)}(t) = \mathrm{e}^{\pm i\omega_k t}$, while enhancement, e.g. due to the evolution in the time-dependent background, means that $f_{\mathbf{k}}^{(\pm)}$ are large. But in any case, Wick's theorem is valid, provided that the state of the system is the vacuum, $a_{\mathbf{k}}|0\rangle = 0$. Hence, it is quite likely that the density perturbations originate from the enhanced vacuum fluctuations of non-interacting or weakly interacting quantum field(s).

Search for *non-Gaussianity* is an important topic of current research. It would show up as a deviation from Wick's theorem. As an example, the three-point function (bispectrum) may be non-vanishing,

$$\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle = \delta(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\, G(k_i^2;\ \mathbf{k}_1\mathbf{k}_2;\ \mathbf{k}_1\mathbf{k}_3) \neq 0\ .$$

The functional dependence of $G(k_i^2;\ \mathbf{k}_1\mathbf{k}_2;\ \mathbf{k}_1\mathbf{k}_3)$ on its arguments is different in different models of generation of primordial perturbations, so this shape is a potential discriminator. In some models the bispectrum vanishes, e.g., due to symmetries. In that case the trispectrum (connected 4-point function) may be measurable instead. For the time being, non-Gaussianity has not been detected.

Inflation does the job of producing Gaussian primordial perturbations very well. At the inflationary epoch, fluctuations of all light fields get enhanced greatly due to the fast expansion of the Universe. This is true, in particular, for the inflaton, the field that dominates the energy density at the time of inflation. Enhanced vacuum fluctuations by inflaton are reprocessed into adiabatic perturbations in the hot medium after the end of inflation. The inflaton field is very weakly coupled, so the non-Gaussianity in the primordial scalar perturbations is very small [97]. In fact, it is so small that its detection is problematic even in the distant future. It is worth noting that this refers to the simplest, single field inflationary models. In models with more than one relevant field the situation may be different, and sizeable non-Gaussianity may be generated.

The generation of the density perturbations is less automatic in scenarios alternative to inflation. Most models proposed so far can be adjusted in such a way that non-Gaussianity is not particularly strong, but potentially observable. In many cases the bispectrum $G(k_i^2;\ \mathbf{k}_1\mathbf{k}_2;\ \mathbf{k}_1\mathbf{k}_3)$ and/or trispectrum are different from inflationary theories.

12.3 Nearly flat power spectrum

Another important property is that the primordial power spectrum of density perturbations **is nearly, but not exactly flat**. For homogeneous and anisotropic Gaussian random field, the power spectrum completely determines its only characteristic, the two-point function. A convenient definition is

$$\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle = \frac{1}{4\pi k^3}\mathcal{P}(k)\delta(\mathbf{k}+\mathbf{k}')\ . \tag{72}$$

The power spectrum $\mathcal{P}(k)$ defined in this way determines the fluctuation in a logarithmic interval of momenta,

$$\langle\delta^2(\mathbf{x})\rangle = \int_0^\infty \frac{dk}{k}\,\mathcal{P}(k)\ .$$

By definition, the flat, scale-invariant spectrum is such that $\mathcal{P}$ is independent of k. The flat spectrum was conjectured by Harrison [98], Zeldovich [99] and Peebles and Yu [100] in the beginning of 1970's, long before realistic mechanisms of the generation of density perturbations have been proposed.

In view of the approximate flatness, a natural parametrization is

$$\mathcal{P}(k) = A_s\left(\frac{k}{k_*}\right)^{n_s-1}\ , \tag{73}$$

where A_s is the amplitude, (n_s-1) is the tilt and k_* is a fiducial momentum, chosen at one's convenience. The flat spectrum in this parametrization has $n_s = 1$. The cosmological data give [3]

$$n_s = 0.965 \pm 0.004\ . \tag{74}$$

This quantifies what we mean by a nearly, but not exactly flat power spectrum.

The approximate flatness of the primordial power spectrum in an inflationary theory is explained by the symmetry of the de Sitter space-time, which is the space-time of constant Hubble rate,

$$ds^2 = dt^2 - \mathrm{e}^{2Ht}d\mathbf{x}^2\ , \qquad H = \text{const}\ .$$

This metric is invariant under spatial dilatations supplemented by time translations,

$$\mathbf{x} \to \lambda \mathbf{x} , \quad t \to t - \frac{1}{2H} \log \lambda .$$

Therefore, all spatial scales are alike, as required for the flat power spectrum. At inflation, H and the inflaton field are almost constant in time, and the de Sitter symmetry is an approximate symmetry. For this reason inflation automatically generates nearly a flat power spectrum. However, neither H nor inflaton are exactly time-independent. This naturally leads to the slight tilt in the spectrum. Overall, this picture is qualitatively consistent with the result in Eq. (74), though the quantitative prediction depends on the concrete inflationary model.

The situation is not so straightforward in alternatives to inflation: the approximate flatness of the scalar power spectrum is not at all automatic. So, one has to work hard to obtain this property. Similarly to an inflationary theory, the flatness of the scalar power spectrum may be due to some symmetry. One candidate symmetry is conformal invariance [101, 102]. The point is that the conformal group includes dilatations, $x^\mu \to \lambda x^\mu$. This property indicates that the theory possesses no scale, and has good chance for producing the flat spectrum. This idea is indeed realized at least at the toy model level.

12.4 Statistical isotropy

In principle, the power spectrum of scalar perturbations may depend on the direction of momentum, e.g.,

$$\mathcal{P}(\mathbf{k}) = \mathcal{P}_0(k) \left(1 + w_{ij}(k) \frac{k_i k_j}{k^2} + \dots \right) ,$$

where w_{ij} is a fundamental tensor in our part of the Universe (odd powers of k_i would contradict commutativity of the Gaussian random field $\delta(\mathbf{k})$). Such a dependence would imply that the Universe was anisotropic at the pre-hot stage, when the primordial perturbations were generated. This statistical anisotropy is rather hard to obtain in inflationary models, though it is possible in inflation with strong vector fields [103]. On the other hand, statistical anisotropy is natural in some other scenarios, including conformal models [104]. The statistical anisotropy would show up in correlations [105]

$$\langle a_{lm} a_{l'm'} \rangle \qquad \text{with } \; l' \neq l \; \text{ and/or } \; m' \neq m \; .$$

At the moment, the constraints [106] on statistical anisotropy obtained by analysing the CMB data are getting into the region, which is interesting from the viewpoint of some (though not many) models of the pre-hot epoch.

12.5 Tensor modes

The distinguishing property of inflation is *the generation of tensor modes (primordial gravity waves)* of sizeable amplitude and a nearly flat power spectrum. The gravity waves are thus smoking guns for inflation (although there is some debate on this point). Indeed, there seems to be no way of generating a nearly flat tensor power spectrum in alternatives to inflation; in fact, most, if not all, alternative scenarios predict unobservably small tensor modes. The reason for their generation at the time of inflation is that the exponential expansion of the Universe enhances vacuum fluctuations of all fields, including the gravitational field itself. Particularly interesting are gravity waves whose present wavelengths are huge, 100 Mpc and larger, and periods are of the order of a billion years and larger. Many inflationary models predict their amplitudes to be very large, of order 10^{-6} or so. Shorter gravity waves are generated too, but their amplitudes decay after horizon entry at the radiation domination epoch, and today they have much smaller amplitudes making them inaccessible to gravity wave detectors like LIGO/VIRGO, eLISA, etc. A conventional characteristic of the amplitude of primordial gravity waves is the tensor-to-scalar ratio

$$r = \frac{\mathcal{P}_T}{\mathcal{P}} ,$$

where $\mathcal{P}$ is the scalar power spectrum defined in Eq. (72) and $\mathcal{P}_T$ is the tensor power spectrum defined in a similar way, but for transverse traceless metric perturbations h_{ij}.

Until recently, the most sensitive probe of the tensor perturbations has been the CMB temperature anisotropy [107]. Nowadays, the best tool is the CMB polarization. The point is that a certain class of polarization patterns (called B-mode) is generated by tensor perturbations, while scalar perturbations are unable to create it [108]. Hence, dedicated experiments aiming at measuring the CMB polarization may well discover the tensor perturbations, i.e., relic gravity waves. Needless to say, this would be a profound discovery. To avoid confusion, let us note that the CMB polarization has been already observed, but it belongs to another class of patterns (so called E-mode) and is consistent with the existence of the scalar perturbations only.

The result of the search for effects of the tensor modes on the CMB temperature anisotropy is shown in Fig. 35. This search has already ruled out some of the popular inflationary models.

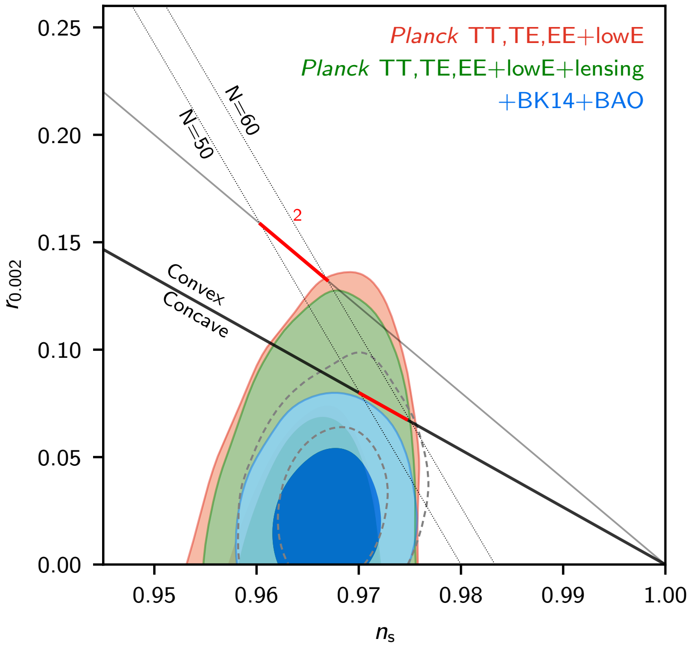


Fig. 35: Allowed regions (at 68% and 95% CL) in the plane (n_s, r), where n_s is the scalar spectral index and r is the tensor-to-scalar ratio [3], obtained by the Planck collaboration alone and by combining the Planck data with the BAO data and CMB polarization data from the BICEP2/KEK experiments. The right corner (the point $(1.0, 0.0)$) is the Harrison–Zeldovich point (flat scalar spectrum, no tensor modes). Intervals show predictions of inflationary models with quadratic and linear inflaton potentials.

13 Conclusion

The present situations in particle physics, on one side, and cosmology, on the other, have much in common. The Standard Model of particle physics and the Standard Model of cosmology, ΛCDM, have been shaped. Both fields enjoyed fairly unexpected discoveries: neutrino oscillations and accelerated expansion of the Universe.

There is strong evidence that the two Standard Models are both incomplete. Therefore, in both fields one hopes for new, revolutionary discoveries. In the context of these lectures, we hope to learn what is the dark matter particle; we may learn the origin of the matter-antimatter asymmetry in the Universe;

the discoveries of new properties of cosmological perturbations will hopefully reveal the nature of the pre-hot epoch.

However, there is no guarantee of new discoveries in particle physics or cosmology. Nature may hide its secrets. Whether or not we will be able to reveal these secrets is the biggest open question in fundamental physics.

References

[1] S. Dodelson, *Modern cosmology,* Academic Press, Amsterdam, 2003, doi:10.1016/B978-0-12-219141-1.X5019-0;
V. Mukhanov, *Physical foundations of cosmology,* Cambridge University Press, Cambridge, 2005, doi:10.1017/CBO9780511790553;
S. Weinberg, *Cosmology*, Oxford University Press, Oxford, 2008, WorldCat;
A.R. Liddle and D.H. Lyth, *The primordial density perturbation: Cosmology, inflation and the origin of structure,* Cambridge University Press, 2009, doi:10.1017/CBO9780511819209;
D.S. Gorbunov and V.A. Rubakov, *Introduction to the theory of the early universe: Hot big bang theory,* 2nd ed., World Scientific, Singapore, 2018, doi:10.1142/10447;
D.S. Gorbunov and V.A. Rubakov, *Introduction to the theory of the early universe: Cosmological perturbations and inflationary theory,* World Scientific, Singapore, 2011, doi:10.1142/7873.

[2] E. Di Valentino, A. Melchiorri and J. Silk, “Planck evidence for a closed universe and a possible crisis for cosmology”, *Nature Astron.* **4** (2019) 196–203, doi:10.1038/s41550-019-0906-9, arXiv:1911.02087 [astro-ph.CO].

[3] N. Aghanim *et al.* [Planck Collaboration], “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, arXiv:1807.06209 [astro-ph.CO].

[4] E. Gawiser and J. Silk “The cosmic microwave background radiation”, *Phys. Rept.* **333** (2000) 245, doi:10.1016/S0370-1573(00)00025-9, arXiv:astro-ph/0002044.

[5] N.G. Busca *et al.*, “Baryon acoustic oscillations in the Lyα forest of BOSS quasars”, *Astron. Astrophys.* **552** (2013) A96, doi:10.1051/0004-6361/201220724, arXiv:1211.2616 [astro-ph.CO].

[6] S. Weinberg, “Anthropic bound on the cosmological constant”, *Phys. Rev. Lett.* **59** (1987) 2607, doi:10.1103/PhysRevLett.59.2607.

[7] A.D. Linde, “Inflation and quantum cosmology”, in: *Three hundred years of gravitation*, Eds. S.W. Hawking and W. Israel (Cambridge Univ. Press, Cambridge, 1987) pp. 604–630.

[8] C. Amsler *et al.* [Particle Data Group], *Phys. Lett.* **B667** (2008) 1, doi:10.1016/j.physletb.2008.07.018 and 2009 partial update for the 2010 edition https://pdg.lbl.gov/2009/.

[9] K.A. Olive, “Dark matter”, arXiv:astro-ph/0301505;
G. Bertone, D. Hooper and J. Silk, “Particle dark matter: Evidence, candidates and constraints”, *Phys. Rept.* **405** (2005) 279, doi:10.1016/j.physrep.2004.08.031, arXiv:hep-ph/0404175;
A. Boyarsky, O. Ruchayskiy and M. Shaposhnikov, “The role of sterile neutrinos in cosmology and astrophysics”, *Ann. Rev. Nucl. Part. Sci.* **59** (2009) 191, doi:10.1146/annurev.nucl.010909.083654, arXiv:0901.0011 [hep-ph];
M. Kawasaki and K. Nakayama, “Axions: Theory and cosmological role”, *Ann. Rev. Nucl. Part. Sci.* **63** (2013) 69, doi:10.1146/annurev-nucl-102212-170536, arXiv:1301.1123 [hep-ph];

H. Baer *et al.*, "Dark matter production in the early universe: beyond the thermal WIMP paradigm", *Phys. Rept.* **555** (2015) 1, doi:10.1016/j.physrep.2014.10.002, arXiv:1407.0017 [hep-ph].

[10] L. Roszkowski, E.M. Sessolo and S. Trojanowski, "WIMP dark matter candidates and searches—current status and future prospects", *Rept. Prog. Phys.* **81** (2018) 066201, doi:10.1088/1361-6633/aab913, arXiv:1707.06277 [hep-ph].

[11] G. Arcadi *et al.*, "The waning of the WIMP? A review of models, searches, and constraints", *Eur. Phys. J.*.**C78** (2018) 203, doi:10.1140/epjc/s10052-018-5662-y, arXiv:1703.07364 [hep-ph].

[12] J.S. Bullock and M. Boylan-Kolchin, "Small-scale challenges to the ΛCDM paradigm", *Ann. Rev. Astron. Astrophys.* **55** (2017) 343, doi:10.1146/annurev-astro-091916-055313, arXiv:1707.04256 [astro-ph.CO].

[13] K.G. Begeman, A.H. Broeils and R.H. Sanders "Extended rotation curves of spiral galaxies: Dark haloes and modified dynamics", *Mon. Not. Roy. Astron. Soc.* **249** (1991) 523, doi:10.1093/mnras/249.3.523.

[14] J.P. Kneib *et al.*, "A wide field Hubble Space Telescope study of the cluster Cl0024+1654 at Z=0.4. II. The cluster mass distribution, *Astrophys. J.* **598** (2003) 804, doi:10.1086/378633, astro-ph/0307299.

[15] D. Clowe *et al.*, "A direct empirical proof of the existence of dark matter", *Astrophys. J.* **648** (2006) L109, doi:10.1086/508162, arXiv:astro-ph/0608407.

[16] A.D. Sakharov, "The initial stage of an expanding universe and the appearance of a nonuniform distribution of matter", *Zh. Eksp. Teor. Fiz.* **49** 345, English transl. in *Sov. Phys. JETP* **22** (1966) 241, http://jetp.ras.ru/cgi-bin/e/index/e/22/1/p241?a=list.

[17] D.J. Eisenstein *et al.* [SDSS Collaboration], "Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies", *Astrophys. J.* **633** (2005) 560, doi:10.1086/466512, arXiv:astro-ph/0501171.

[18] S.Y. Kim, A.H.G. Peter and J.R. Hargis, "Missing satellites problem: Completeness corrections to the number of satellite galaxies in the Milky Way are consistent with cold dark matter predictions", *Phys. Rev. Lett.* **121** (2018) 211302, doi:10.1103/PhysRevLett.121.211302, arXiv:1711.06267 [astro-ph.CO].

[19] S. Shen *et al.* "The baryon cycle of dwarf galaxies: Dark, bursty, gas-rich polluters", *Astrophys. J.* **792** (2014) 99, doi:10.1088/0004-637X/792/2/99, arXiv:1308.4131 [astro-ph.CO].

[20] Y. Revaz and P. Jablonka, "Pushing back the limits: detailed properties of dwarf galaxies in a ΛCDM universe", *Astron. Astrophys.* **616** (2018) A96, doi:10.1051/0004-6361/201832669, arXiv:1801.06222 [astro-ph.GA].

[21] S. Garrison-Kimmel *et al.*, "Not so lumpy after all: modelling the depletion of dark matter subhaloes by Milky Way-like galaxies", *Mon. Not. Roy. Astron. Soc.* **471** (2017) 1709, doi:10.1093/mnras/stx1710, arXiv:1701.03792 [astro-ph.GA].

[22] A. Drlica-Wagner *et al.* [LSST Dark Matter Group], "Probing the fundamental nature of dark matter with the Large Synoptic Survey Telescope", arXiv:1902.01055 [astro-ph.CO];
K. Bechtol *et al.*, "Dark matter science in the era of LSST", arXiv:1903.04425 [astro-ph.CO].

[23] A. Boyarsky, O. Ruchayskiy and D. Iakubovskyi, "A lower bound on the mass of dark matter particles", *JCAP* **03** (2009) 005, doi:10.1088/1475-7516/2009/03/005, arXiv:0808.3902 [hep-ph].

[24] D. Gorbunov, A. Khmelnitsky and V. Rubakov, "Constraining sterile neutrino dark matter by phase-space density observations", *JCAP* **10** (2008) 041, doi:10.1088/1475-7516/2008/10/041, arXiv:0808.3910 [hep-ph].

[25] V. Iršič *et al.*, “New constraints on the free-streaming of warm dark matter from intermediate and small scale Lyman-α forest data”, *Phys. Rev.* **D96** (2017) 023522, doi:10.1103/PhysRevD.96.023522, arXiv:1702.01764 [astro-ph.CO].

[26] M.R. Lovell *et al.*, “Addressing the too big to fail problem with baryon physics and sterile neutrino dark matter”, *Mon. Not. Roy. Astron. Soc.* **468** (2017) 2836, doi:10.1093/mnras/stx621, arXiv:1611.00005 [astro-ph.GA].

[27] D.N. Spergel and P.J. Steinhardt, “Observational evidence for selfinteracting cold dark matter”, *Phys. Rev. Lett.* **84** (2000) 3760, doi:10.1103/PhysRevLett.84.3760, arXiv:astro-ph/9909386.

[28] L. Hui *et al.*, “Ultralight scalars as cosmological dark matter”, *Phys. Rev.* **D95** (2017) 043541, doi:10.1103/PhysRevD.95.043541, arXiv:1610.08297 [astro-ph.CO].

[29] V. Iršič *et al.*, “First constraints on fuzzy dark matter from Lyman-α forest data and hydrodynamical simulations”, *Phys. Rev. Lett.* **119** (2017) 031302, doi:10.1103/PhysRevLett.119.031302, arXiv:1703.04683 [astro-ph.CO];
M. Nori *et al.*, “Lyman α forest and non-linear structure characterization in fuzzy dark matter cosmologies”, *Mon. Not. Roy. Astron. Soc.* **482** (2019) 3227, doi:10.1093/mnras/sty2888, arXiv:1809.09619 [astro-ph.CO].

[30] A. Khmelnitsky and V. Rubakov, “Pulsar timing signal from ultralight scalar dark matter”, *JCAP* **02** (2014) 019, doi:10.1088/1475-7516/2014/02/019, arXiv:1309.5888 [astro-ph.CO];
N.K. Porayko *et al.*, “Parkes pulsar timing array constraints on ultralight scalar-field dark matter”, *Phys. Rev.* **D98** (2018) 102002, doi:10.1103/PhysRevD.98.102002, arXiv:1810.03227 [astro-ph.CO].

[31] V.A. Rubakov, “Cosmology”, *CERN Yellow Rep. School Proc.* **2** (2017) 239, doi:10.23730/CYRSP-2017-002.239, arXiv:1804.11230 [gr-qc].

[32] M. Cirelli, N. Fornengo and A. Strumia, “Minimal dark matter”, *Nucl. Phys.* **B753** (2006) 178, doi:10.1016/j.nuclphysb.2006.07.012, arXiv:hep-ph/0512090;
M. Cirelli and A. Strumia, “Minimal dark matter: Model and results”, *New J. Phys.* **11** (2009) 105005, doi:10.1088/1367-2630/11/10/105005, arXiv:0903.3381 [hep-ph].

[33] A. Bottino and N. Fornengo, “Dark matter and its particle candidates”, pres. at the 5th ICTP School on Nonaccelerator Particle Astrophysics, Trieste, Italy, 29 Jun.–10 Jul. 1998, hep-ph/9904469.

[34] E. Aprile *et al.* [XENON Collaboration], “Dark matter search results from a one ton-year exposure of XENON1T”, *Phys. Rev. Lett.* **121** (2018) 111302,doi:10.1103/PhysRevLett.121.111302, arXiv:1805.12562 [astro-ph.CO].

[35] M. Lisanti, “Lectures on dark matter physics”, Proc. Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings (TASI 2015), Boulder, CO, USA, 1–26 Jun. 2015, Eds. J. Polchinski, P. Vieira and O. DeWolfe (World Scientific, Singapore, 2017), pp. 399-446, doi:10.1142/9789813149441_0007, arXiv:1603.03797 [hep-ph].

[36] A.D. Avrorin *et al.* [Baikal Collaboration], “Search for neutrino emission from relic dark matter in the Sun with the Baikal NT200 detector”, *Astropart. Phys.* **62** (2014) 12, doi:10.1016/j.astropartphys.2014.07.006, arXiv:1405.3551 [astro-ph.HE].

[37] L. Bergström *et al.*, “New limits on dark matter annihilation from AMS cosmic ray positron data”, *Phys. Rev. Lett.* **111** (2013) 171101, doi:10.1103/PhysRevLett.111.171101, arXiv:1306.3983 [astro-ph.HE].

[38] G. ’t Hooft, “Symmetry breaking through Bell-Jackiw anomalies”, *Phys. Rev. Lett.* **37** (1976) 8, doi:10.1103/PhysRevLett.37.8.

[39] C.G. Callan, R.F. Dashen and D.J. Gross, “The structure of the gauge theory vacuum”, *Phys. Lett.***B63** (1976) 334, doi:10.1016/0370-2693(76)90277-X.

[40] R. Jackiw and C. Rebbi, "Vacuum periodicity in a Yang-Mills quantum theory", *Phys. Rev. Lett.* **37** (1976) 172, doi:10.1103/PhysRevLett.37.172.

[41] J.E. Kim and G. Carosi, "Axions and the strong CP problem", *Rev. Mod. Phys.* **82** (2010) 557, doi:10.1103/RevModPhys.82.557, arXiv:0807.3125 [hep-ph], Erratum: *Rev. Mod. Phys.* **91** (2019) 049902, doi:10.1103/RevModPhys.91.049902.

[42] R.D. Peccei and H.R. Quinn, "CP conservation in the presence of instantons", *Phys. Rev. Lett.* **38** (1977) 1440, doi:10.1103/PhysRevLett.38.1440.

[43] S. Weinberg, "A new light boson?", *Phys. Rev. Lett.* **40** (1978) 223, doi:10.1103/PhysRevLett.40.223.

[44] F. Wilczek, "Problem of strong P and T invariance in the presence of instantons", *Phys. Rev. Lett.* **40** (1978) 279, doi:10.1103/PhysRevLett.40.279.

[45] M. Dine, W. Fischler and M. Srednicki, "A simple solution to the strong CP problem with a harmless axion", *Phys. Lett.* **B104** (1981) 199, doi:10.1016/0370-2693(81)90590-6.

[46] A.R. Zhitnitsky, "On possible suppression of the axion hadron interactions", *Yad. Fiz.* **31** (1980) 497, English transl. publ. in *Sov. J. Nucl. Phys.* **31** (1980) 260.

[47] J.E. Kim, "Weak interaction singlet and strong CP invariance", *Phys. Rev. Lett.* **43** (1979) 103, doi:10.1103/PhysRevLett.43.103.

[48] M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, "Can confinement ensure natural CP invariance of strong interactions?", *Nucl. Phys.* **B166** (1980) 493, doi:10.1016/0550-3213(80)90209-6.

[49] V.A. Rubakov, *JETP Lett.* **65** (1997) 621, doi:10.1134/1.567390, arXiv:hep-ph/9703409;
P. Agrawal and K. Howe, "Factoring the strong CP problem", *JHEP* **12** (2018) 029, doi:10.1007/JHEP12(2018)029, arXiv:1710.04213 [hep-ph];
M.K. Gaillard *et al.*, "Color unified dynamical axion", *Eur. Phys. J.* **C78** (2018) 972, doi:10.1140/epjc/s10052-018-6396-6, arXiv:1805.06465 [hep-ph].

[50] J. Preskill, M.B. Wise and F. Wilczek, "Cosmology of the invisible axion", *Phys. Lett.* **B120** (1983) 127, doi:10.1016/0370-2693(83)90637-8.

[51] L.F. Abbott and P. Sikivie, "A cosmological bound on the invisible axion", *Phys. Lett.* **B120** (1983) 133, doi:10.1016/0370-2693(83)90638-X.

[52] M. Dine and W. Fischler, "The not-so-harmless axion", *Phys. Lett.* **B120** (1983) 137, doi:10.1016/0370-2693(83)90639-1.

[53] L. Visinelli and P. Gondolo, "Dark matter axions revisited", *Phys. Rev.* **D80** (2009) 035024, doi:10.1103/PhysRevD.80.035024, arXiv:0903.4377 [astro-ph.CO].

[54] A. Vilenkin and A.E. Everett, "Cosmic strings and domain walls in models with goldstone and pseudogoldstone bosons", *Phys. Rev. Lett.* **48** (1982) 1867, doi:10.1103/PhysRevLett.48.1867.

[55] R.A. Battye and E.P.S. Shellard, "Axion string cosmology and its controversies", Proc. 2nd Int. Conf. Physics beyond the Standard Model, Ed. H.V. Klapdor-Kleingrothaus (Springer, Berlin, 1999), pp. 565–572, arXiv:astro-ph/9909231.

[56] V.B. Klaer and G.D. Moore, "The dark-matter axion mass", *JCAP* **11** (2017) 049, doi:10.1088/1475-7516/2017/11/049, arXiv:1708.07521 [hep-ph].

[57] M. Kawasaki, K. Saikawa and T. Sekiguchi, "Axion dark matter from topological defects", *Phys. Rev.* **D91** (2015) 065014, doi:10.1103/PhysRevD.91.065014, arXiv:1412.0789 [hep-ph];
M. Kawasaki *et al.*, "Long-term dynamics of cosmological axion strings", PTEP **2018** (2018) 091E01, doi:10.1093/ptep/pty098, arXiv:1806.05566 [hep-ph].

[58] E.W. Kolb and I.I. Tkachev, "Axion miniclusters and Bose stars", *Phys. Rev. Lett.* **71** (1993) 3051, doi:10.1103/PhysRevLett.71.3051 [hep-ph/9303313].

[59] P. Tinyakov, I. Tkachev and K. Zioutas, “Tidal streams from axion miniclusters and direct axion searches”, *JCAP* **01** (2016) 035, doi:10.1088/1475-7516/2016/01/035, arXiv:1512.02884 [astro-ph.CO].

[60] D.G. Levkov, A.G. Panin and I.I. Tkachev, “Gravitational Bose-Einstein condensation in the kinetic regime”, *Phys. Rev. Lett.* **121** (2018) 151301, doi:10.1103/PhysRevLett.121.151301, arXiv:1804.05857 [astro-ph.CO].

[61] S.J. Asztalos *et al.* [ADMX Collaboration], “A SQUID-based microwave cavity search for dark-matter axions”, *Phys. Rev. Lett.* **104** (2010) 041301, doi:10.1103/PhysRevLett.104.041301, arXiv:0910.5914 [astro-ph.CO].

[62] N. Du *et al.* [ADMX Collaboration], “A search for invisible axion dark matter with the Axion Dark Matter Experiment”, *Phys. Rev. Lett.* **120** (2018) 151301, doi:10.1103/PhysRevLett.120.151301, arXiv:1804.05750 [hep-ex].

[63] R. Battesti *et al.*, “High magnetic fields for fundamental physics”, *Phys. Rept.* **765-766** (2018) 1, doi:10.1016/j.physrep.2018.07.005, arXiv:1803.07547 [physics.ins-det].

[64] B. Majorovits *et al.* [MADMAX interest Group], “MADMAX: A new road to axion dark matter detection”, *J. Phys. Conf. Ser.* 1342 (2020) 012098, doi:10.1088/1742-6596/1342/1/012098, arXiv:1712.01062 [physics.ins-det].

[65] P.W. Graham *et al.*, “Experimental searches for the axion and axion-like particles”, *Ann. Rev. Nucl. Part. Sci.* **65** (2015) 485, doi:10.1146/annurev-nucl-102014-022120, arXiv:1602.00039 [hep-ex].

[66] D. Budker *et al.*, “Proposal for a Cosmic Axion Spin Precession Experiment (CASPEr)”, *Phys. Rev.* **X4** (2014) 021030, doi:10.1103/PhysRevX.4.021030, arXiv:1306.6089 [hep-ph].

[67] S. Alekhin *et al.*, “A facility to search for hidden particles at the CERN SPS: the SHiP physics case”, *Rept. Prog. Phys.* **79** (2016) 124201, doi:10.1088/0034-4885/79/12/124201, arXiv:1504.04855 [hep-ph].

[68] 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, arXiv:1402.2301 [astro-ph.CO].

[69] 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, arXiv:1402.4119 [astro-ph.CO].

[70] A. Boyarsky *et al.*, “Sterile neutrino dark matter”, *Prog. Part. Nucl. Phys.* **104** (2019) 1, doi:10.1016/j.ppnp.2018.07.004, arXiv:1807.07938 [hep-ph].

[71] X.-D. Shi and G.M. Fuller, “A new dark matter candidate: Nonthermal sterile neutrinos”, *Phys. Rev. Lett.* **82** (1999) 2832 , doi:10.1103/PhysRevLett.82.2832, arXiv:astro-ph/9810076.

[72] D.N. Abdurashitov *et al.*, “The current status of "Troitsk nu-mass" experiment in search for sterile neutrino”, *JINST* **10** (2015) T10005, doi:10.1088/1748-0221/10/10/T10005, arXiv:1504.00544 [physics.ins-det].

[73] A.D. Sakharov, “Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe”, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32, English transl. reprinted in *Sov.Phys.Usp.*, **34** (1991) 392, doi:10.1070/PU1991v034n05ABEH002497.

[74] V.A. Kuzmin, “CP-noninvairiance and baryon asymmetry of the universe”, *Pisma Zh. Eksp. Teor. Fiz.* **12** (1970) 335, English transl. in *JETP Letters* **12** (1970) 228, http://jetpletters.ru/ps/1730/article_26297.shtml.

[75] I. Affleck and M. Dine, “A new mechanism for baryogenesis”, *Nucl. Phys.* **B249** (1985) 361, doi:10.1016/0550-3213(85)90021-5.

[76] F.R. Klinkhamer and N.S. Manton, “A saddle point solution in the Weinberg-Salam theory”, *Phys. Rev.* **D30** (1984) 2212, doi:10.1103/PhysRevD.30.2212.

[77] 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.

[78] V.A. Rubakov, *Classical theory of gauge fields* (Princeton Univ. Press, Princeton, 2002).

[79] A.A. Belavin *et al.*, “Pseudoparticle solutions of the Yang-Mills equations”, *Phys. Lett.* **B59** (1975) 85, doi:10.1016/0370-2693(75)90163-X.

[80] V.A. Rubakov and M.E. Shaposhnikov, “Electroweak baryon number nonconservation in the early universe and in high-energy collisions”, *Usp. Fiz. Nauk* **166** (1996) 493, doi:10.3367/UFNr.0166.199605d.0493, English vers. publ. in *Phys. Usp.* **39** (1996) 461, doi:10.1070/PU1996v039n05ABEH000145, arXiv:hep-ph/9603208.

[81] K. Kajantie *et al.*, ‘The electroweak phase transition: A nonperturbative analysis”, *Nucl. Phys.* **B466** (1996) 189, doi:10.1016/0550-3213(96)00052-1, arXiv:hep-lat/9510020.

[82] S.V. Demidov, D.S. Gorbunov and D.V. Kirpichnikov, “Split NMSSM with electroweak baryogenesis”, *JHEP* **11** (2016) 148, doi:10.1007/JHEP11(2016)148, arXiv:1608.01985 [hep-ph], Erratum: *JHEP* **08** (2017) 080, doi:10.1007/JHEP08(2017)080.

[83] V. Andreev *et al.* [ACME Collaboration], “Improved limit on the electric dipole moment of the electron”, *Nature* **562** (2018) 355, doi:10.1038/s41586-018-0599-8.

[84] T. Konstandin, “Quantum transport and electroweak baryogenesis”, *Usp. Fiz. Nauk* **183** (2013) 785, doi:10.3367/UFNr.0183.201308a.0785, English vers. publ. in *Phys. Usp.* **56** (2013) 747, doi:10.3367/UFNe.0183.201308a.0785, arXiv:1302.6713 [hep-ph];
M. Chala, G. Nardini and I. Sobolev, “Unified explanation for dark matter and electroweak baryogenesis with direct detection and gravitational wave signatures”, *Phys. Rev.* **D94** (2016) 055006, doi:10.1103/PhysRevD.94.055006, arXiv:1605.08663 [hep-ph];
J.M. Cline, K. Kainulainen and D. Tucker-Smith, “Electroweak baryogenesis from a dark sector”, *Phys. Rev.* **D95** (2017) 115006, doi:10.1103/PhysRevD.95.115006, arXiv:1702.08909 [hep-ph];
S. Bruggisser *et al.*, “Electroweak phase transition and baryogenesis in composite Higgs models”, *JHEP* **12** (2018) 099, doi:10.1007/JHEP12(2018)099, arXiv:1804.07314 [hep-ph];
I. Baldes and G. Servant, “High scale electroweak phase transition: baryogenesis & symmetry non-restoration”, *JHEP* **10** (2018) 053, doi:10.1007/JHEP10(2018)053, arXiv:1807.08770 [hep-ph];
M. Carena, M. Quirós and Y. Zhang, “Electroweak baryogenesis from dark-sector CP Violation”, *Phys. Rev. Lett.* **122** (2019) 201802, doi:10.1103/PhysRevLett.122.201802, arXiv:1811.09719 [hep-ph];
A. Glioti, R. Rattazzi and L. Vecchi, “Electroweak baryogenesis above the electroweak scale”, *JHEP* **04** (2019) 027, doi:10.1007/JHEP04(2019)027, arXiv:1811.11740 [hep-ph].

[85] M. Fukugita and T. Yanagida, “Baryogenesis without grand unification”, *Phys. Lett.* **B174** (1986) 45, doi:10.1016/0370-2693(86)91126-3.

[86] E.K. Akhmedov, V.A. Rubakov and A.Y. Smirnov, “Baryogenesis via neutrino oscillations”, *Phys. Rev. Lett.* **81** (1998) 1359, doi:10.1103/PhysRevLett.81.1359, arXiv:hep-ph/9803255.

[87] T. Asaka and M. Shaposhnikov, “The νMSM, dark matter and baryon asymmetry of the universe”, *Phys. Lett.* **B620** (2005) 17, doi:10.1016/j.physletb.2005.06.020, arXiv:hep-ph/0505013.

[88] M. Drewes and B. Garbrecht, “Leptogenesis from a GeV seesaw without mass degeneracy”, *JHEP* **03** (2013) 096, doi:10.1007/JHEP03(2013)096, arXiv:1206.5537 [hep-ph].

[89] W. Buchmuller, R.D. Peccei and T. Yanagida, “Leptogenesis as the origin of matter”, *Ann. Rev. Nucl. Part. Sci.* **55** (2005) 311, 10.1146/annurev.nucl.55.090704.15558, arXiv:hep-ph/0502169;
S. Davidson, E. Nardi and Y. Nir, “Leptogenesis”, *Phys. Rept.* **466** (2008) 105, doi:10.1016/j.physrep.2008.06.002, arXiv:0802.2962 [hep-ph];
C.S. Fong, E. Nardi and A. Riotto, “Leptogenesis in the universe”, *Adv. High Energy Phys.* **2012** (2012) 158303, doi:10.1155/2012/158303, arXiv:1301.3062 [hep-ph].

[90] E.W. Kolb and M.S. Turner, “Grand Unified Theories and the origin of the baryon asymmetry”, *Ann. Rev. Nucl. Part. Sci.* **33** (1983) 645, doi:10.1146/annurev.ns.33.120183.003241;
A. Riotto and M. Trodden, “Recent progress in baryogenesis”, *Ann. Rev. Nucl. Part. Sci.* **49** (1999) 35, doi:10.1146/annurev.nucl.49.1.35, arXiv:hep-ph/9901362.

[91] E. Babichev, D. Gorbunov and S. Ramazanov, “Dark matter and baryon asymmetry from the very dawn of the Universe”, *Phys. Rev.* **D97** (2018) 123543, doi:10.1103/PhysRevD.97.123543, arXiv:1805.05904 [astro-ph.CO].

[92] J. Urrestilla *et al.*, “Cosmic microwave anisotropies from BPS semilocal strings”, *JCAP* **07** (2008) 010 , doi:10.1088/1475-7516/2008/07/010, arXiv:0711.1842 [astro-ph].

[93] A.A. Starobinsky, “Spectrum of relict gravitational radiation and the early state of the universe”, *Pisma Zh. Eksp. Teor. Fiz.* **30** (1979) 719, English transl. publ. in *JETP Lett.* **30** (1979) 682, http://jetpletters.ru/ps/1370/article_20738.shtml;
A.A. Starobinsky, “A new type of isotropic cosmological models without singularity”, *Phys. Lett.* **B91** (1980) 99, doi:10.1016/0370-2693(80)90670-X;
A.H. Guth, “The inflationary universe: A possible solution to the horizon and flatness problems”, *Phys. Rev.* **D23** (1981) 347, doi:10.1103/PhysRevD.23.347;
A.D. Linde, “A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems”, *Phys. Lett.* **B108** (1982) 389, doi:10.1016/0370-2693(82)91219-9;
A. Albrecht and P.J. Steinhardt, “Cosmology for Grand Unified Theories with radiatively induced symmetry breaking”, *Phys. Rev. Lett.* **48** (1982) 1220, doi:10.1103/PhysRevLett.48.1220;
A.D. Linde, “Chaotic inflation”, *Phys. Lett.* **B129** (1983) 177, doi:10.1016/0370-2693(83)90837-7.

[94] V.F. Mukhanov and G.V. Chibisov, "Quantum fluctuation and nonsingular universe”, *Pisma Zh. Eksp. Teor. Fiz.* **33** (1981) 549, English vers. publ. in *JETP Lett.* **33** (1981) 532, http://jetpletters.ru/ps/1510/article_23079.shtml;
S.W. Hawking, “The development of irregularities in a single bubble inflationary universe”, *Phys. Lett.* **B115** (1982) 295, doi:10.1016/0370-2693(82)90373-2;
A.A. Starobinsky, “Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations”, *Phys. Lett.* **B117** (1982) 175, doi:10.1016/0370-2693(82)90541-X;
A.H. Guth and S.Y. Pi, “Fluctuations in the new inflationary universe”, *Phys. Rev. Lett.* **49** (1982) 1110, doi:10.1103/PhysRevLett.49.1110;
J.M. Bardeen, P.J. Steinhardt and M.S. Turner, “Spontaneous creation of almost scale-free density perturbations in an inflationary universe”, *Phys. Rev.* **D28** (1983) 679, doi:10.1103/PhysRevD.28.679.

[95] J.L. Lehners, “Ekpyrotic and cyclic cosmology”, *Phys. Rept.* **465** (2008) 223, doi:10.1016/j.physrep.2008.06.001, arXiv:0806.1245 [astro-ph];
R.H. Brandenberger, “Unconventional cosmology”, *Lect. Notes Phys.* **863** (2013) 333, doi:10.1007/978-3-642-33036-0_12, arXiv:1203.6698 [astro-ph.CO].

[96] P. Creminelli, A. Nicolis and E. Trincherini, “Galilean genesis: An alternative to inflation”, *JCAP* **11** (2010) 021, doi:10.1088/1475-7516/2010/11/021, arXiv:1007.0027 [hep-th].

[97] J.M. Maldacena, “Non-Gaussian features of primordial fluctuations in single field inflationary models”, *JHEP* **05** (2003) 013, doi:10.1088/1126-6708/2003/05/013, arXiv:astro-ph/0210603.

[98] E.R. Harrison, “Fluctuations at the threshold of classical cosmology”, *Phys. Rev.* **D1** (1970) 2726, doi:10.1103/PhysRevD.1.2726.

[99] Y.B. Zeldovich, “A hypothesis, unifying the structure and the entropy of the universe”, *Mon. Not. Roy. Astron. Soc.* **160** (1972) 1P, doi:10.1093/mnras/160.1.1P.

[100] P.J.E. Peebles and J.T. Yu, ‘Primeval adiabatic perturbation in an expanding universe”, *Astrophys. J.* **162** (1970) 815, doi:10.1086/150713.

[101] I. Antoniadis, P.O. Mazur and E. Mottola, “Conformal invariance and cosmic background radiation”, *Phys. Rev. Lett.* **79** (1997) 14, doi:10.1103/PhysRevLett.79.14, arXiv:astro-ph/9611208.

[102] V.A. Rubakov, “Harrison–Zeldovich spectrum from conformal invariance”, *JCAP* **09** (2009) 030, doi:10.1088/1475-7516/2009/09/030, arXiv:0906.3693 [hep-th];
K. Hinterbichler and J. Khoury, “The pseudo-conformal universe: Scale invariance from spontaneous breaking of conformal symmetry”, *JCAP* **04** (2012) 023, doi:10.1088/1475-7516/2012/04/023, arXiv:1106.1428 [hep-th].

[103] M.A. Watanabe, S. Kanno and J. Soda, “Inflationary universe with anisotropic hair”, *Phys. Rev. Lett.* **102** (2009) 191302, doi:10.1103/PhysRevLett.102.191302, arXiv:0902.2833 [hep-th];
T.R. Dulaney and M.I. Gresham, “Primordial power spectra from anisotropic inflation”, *Phys. Rev.* **D81** (2010) 103532, doi:10.1103/PhysRevD.81.103532, arXiv:1001.2301 [astro-ph.CO];
A.E. Gumrukcuoglu, B. Himmetoglu and M. Peloso, “Scalar-scalar, scalar-tensor, and tensor-tensor correlators from anisotropic inflation”, *Phys. Rev.* **D81** (2010) 063528, doi:10.1103/PhysRevD.81.063528, arXiv:1001.4088 [astro-ph.CO].

[104] M. Libanov and V. Rubakov, “Cosmological density perturbations from conformal scalar field: infrared properties and statistical anisotropy”, *JCAP* **11** (2010) 045, doi:10.1088/1475-7516/2010/11/045, arXiv:1007.4949 [hep-th];
M. Libanov, S. Ramazanov and V. Rubakov, “Scalar perturbations in conformal rolling scenario with intermediate stage”, *JCAP* **06** (2011) 010, doi:10.1088/1475-7516/2011/06/010, arXiv:1102.1390 [hep-th].

[105] L. Ackerman, S.M. Carroll and M.B. Wise, “Imprints of a primordial preferred direction on the microwave background”, *Phys. Rev.* **D75** (2007) 083502, doi:10.1103/PhysRevD.75.083502, Erratum: *Phys. Rev.* **D80** (2009) 069901, doi:10.1103/PhysRevD.80.069901, arXiv:astro-ph/0701357;
A.R. Pullen and M. Kamionkowski, “Cosmic microwave background statistics for a direction-dependent primordial power spectrum”, *Phys. Rev.* **D76** (2007) 103529, doi:10.1103/PhysRevD.76.103529, arXiv:0709.1144 [astro-ph].

[106] J. Kim and E. Komatsu, “Limits on anisotropic inflation from the Planck data”, *Phys. Rev.* **D88** (2013) 101301, doi:https:10.1103/PhysRevD.88.101301, arXiv:1310.1605 [astro-ph.CO];
G.I. Rubtsov and S.R. Ramazanov, “Revisiting constraints on the (pseudo)conformal universe with Planck data”, *Phys. Rev.* **D91** (2015) 043514, doi:10.1103/PhysRevD.91.043514, arXiv:1406.7722 [astro-ph.CO];
S. Ramazanov *et al.*, “General quadrupolar statistical anisotropy: Planck limits”, *JCAP* **03** (2017) 039, doi:10.1088/1475-7516/2017/03/039, arXiv:1612.02347 [astro-ph.CO].

[107] V.A. Rubakov, M.V. Sazhin and A.V. Veryaskin, “Graviton creation in the inflationary universe and the grand unification scale”, *Phys. Lett.* **B115** (1982) 189, doi:10.1016/0370-2693(82)90641-4;

R. Fabbri and M.D. Pollock, “The effect of primordially produced gravitons upon the anisotropy of the cosmological microwave background radiation”, *Phys. Lett.* **B125** (1983) 445, doi:10.1016/0370-2693(83)91322-9;
L.F. Abbott and M.B. Wise, “Constraints on generalized inflationary cosmologies”, *Nucl. Phys.* **B244** (1984) 541, doi:10.1016/0550-3213(84)90329-8;
A.A. Starobinsky, “Cosmic background anisotropy induced by isotropic flat-spectrum gravitational-wave perturbations”, *Pis'ma Astron. Zh.* **11** (1985) 323, English transl. publ. in *Sov. Astron. Lett.* **11** (1985) 133, https://ui.adsabs.harvard.edu/abs/1985SvAL...11..133S/abstract.

[108] M. Kamionkowski, A. Kosowsky and A. Stebbins, “A probe of primordial gravity waves and vorticity”, *Phys. Rev. Lett.* **78** (1997) 2058, doi:https://doi.org/10.1103/PhysRevLett.78.2058, arXiv:astro-ph/9609132;
U. Seljak and M. Zaldarriaga, “Signature of gravity waves in polarization of the microwave background”, *Phys. Rev. Lett.* **78** (1997) 2054, doi:10.1103/PhysRevLett.78.2054, arXiv:astro-ph/9609169.

Proceedings of the 2019 European School of High-Energy Physics, St Petersburg, Russia, 14–27 September 2019, edited by C. Duhr and M. Mulders, CERN-2021-005 (CERN, Geneva, 2021)

Practical statistics for particle physics

R. J. Barlow
The University of Huddersfield, Huddersfield, United Kingdom

Abstract

This is the write-up of a set of lectures given at the CERN European School of High Energy Physics in St Petersburg, Russia in September 2019, to an audience of PhD students in all branches of particle physics. They cover the different meanings of 'probability', particularly Frequentist and Bayesian, the binomial, the Poisson and the Gaussian distributions, hypothesis testing, estimation, errors (including asymmetric and systematic errors) and goodness of fit. Several different methods used in setting upper limits are explained, followed by a discussion on why 5 sigma are conventionally required for a 'discovery'.

Keywords

Lectures; statistics; particle physics, probability, estimation, confidence limits.

1 Introduction

To interpret the results of your particle physics experiment and see what it implies for the relevant theoretical model and parameters, you need to use statistical techniques. These are a part of your experimental toolkit, and to extract the maximum information from your data you need to use the correct and most powerful statistical tools.

Particle physics (like, probably, any field of science) has is own special set of statistical processes and language. Our use is in some ways more complicated (we often fit multi-parameter functions, not just straight lines) and in some ways more simple (we do not have to worry about ethics, or law suits). So the generic textbooks and courses you will meet on 'Statistics' are not really appropriate. That's why HEP schools like this one include lectures on statistics as well as the fundamental real physics, like field theory and physics beyond the Standard Model (BSM).

There are several textbooks [1–6] available which are designed for an audience of particle physicists. You will find these helpful—more helpful than general statistical textbooks. You should find one whose language suits you and keep a copy on your bookshelf—preferably purchased—but at least on long term library loan. You will also find useful conference proceedings [7–9], journal papers (particularly in Nuclear Instruments and Methods) and web material: often your own experiment will have a set of pages devoted to the topic.

2 Probability

We begin by looking at the concept of probability. Although this is familiar (we use it all the time, both inside and outside the laboratory), its use is not as obvious as you would think.

2.1 What is probability?

A typical exam for Statistics101 (or equivalent) might well contain the question:

Q1 Explain what is meant by the *probability* P_A of an event A [1]

---

http://doi.org/10.23730/CYRSP-2021-005.197

The '1' in square brackets signifies that the answer carries one mark. That's an indication that just a sentence or two are required, not a long essay.

Asking a group of physicists this question produces answers falling into four different categories

1. P_A is number obeying certain mathematical rules,
2. P_A is a property of A that determines how often A happens,
3. For N trials in which A occurs N_A times, P_A is the limit of N_A/N for large N,
4. P_A is my belief that A will happen, measurable by seeing what odds I will accept in a bet.

Although all these are generally present, number 3 is the most common, perhaps because it is often explicitly taught as the definition. All are, in some way, correct! We consider each in turn.

2.2 Mathematical probability

The Kolmogorov axioms are: For all $A \subset S$

$$\begin{aligned} P_A &\geq 0 \\ P_S &= 1 \\ P_{A\cup B} = P_A + P_B \text{ if } A \cap B = \phi \text{ and } A, B &\subset S \quad . \end{aligned} \tag{1}$$

From these simple axioms a complete and complicated structure of theorems can be erected. This is what pure mathematicians do. For example, the 2nd and 3rd axiom show that the probability of not-A $P_{\overline{A}}$, is $1 - P_A$, and then the 1st axiom shows that $P_A \leq 1$: probabilities must be less than 1.

But the axioms and the ensuing theorems says nothing about what P_A actually means. Kolmogorov had Frequentist probability in mind, but these axioms apply to any definition: he explicitly avoids tying P_A down in this way. So although this apparatus enables us to compute numbers, it does not tell us what we can use them for.

2.3 Real probability

Also known as Classical probability, this was developed during the 18th–19th centuries by Pascal, Laplace and others to serve the gambling industry.

If there are several possible outcomes and there is a symmetry between them so they are all, in a sense, identical, then their individual probabilities must be equal. For example, there are two sides to a coin, so if you toss it there must be a probability $\frac{1}{2}$ for each face to land uppermost. Likewise there are 52 cards in a pack, so the probability of a particular card being chosen is $\frac{1}{52}$. In the same way there are 6 sides to a dice, and 33 slots in a roulette wheel.

This enables you to answer questions like 'What is the probability of rolling more than 10 with 2 dice?'. There are 3 such combinations (5-6, 6-5 and 6-6) out of the $6 \times 6 = 36$ total possibilities, so the probability is $\frac{1}{12}$. Compound instances of A are broken down into smaller instances to which the symmetry argument can be applied. This is satisfactory and clearly applicable—you know that if someone offers you a 10 to 1 bet on this dice throw, you should refuse; in the long run knowledge of the correct probabilities will pay off.

The problem with this 'equally likely' approach is that it cannot be applied to continuous variables. This is brought out in many ways, for example in one of Bertrand's paradoxes:

In a circle of radius R an equilateral triangle is drawn. A chord is drawn at random. What is the probability that the length of the chord is greater than the side of the triangle?

Considering Fig. 1 one can give three answers:

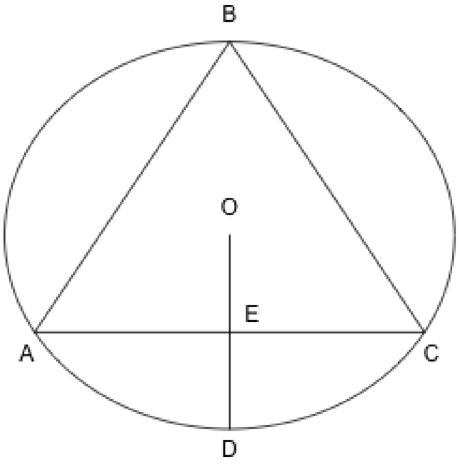


Fig. 1: Bertrand's paradox

1. If the chord, without loss of generality, starts at A, then it will be longer than the side if the end point is anywhere between B and C. So the answer is obviously $\frac{1}{3}$.
2. If the centre of the chord, without loss of generality, is chosen at random along the line OD, then it will be longer than the side of the triangle if it is in OE rather than ED. E is the midpoint of OD so the answer is obviously $\frac{1}{2}$.
3. If the centre of the chord, without loss of generality, is chosen at random within the circle, then it will be longer than the side of the triangle if it lies within the circle of radius $\frac{R}{2}$. So the answer is obviously $\frac{1}{4}$.

So we have three obvious but contradictory answers. The whole question is built on a false premise: drawing a chord 'at random' is, unlike tossing a coin or throwing a dice, not defined. Another way of seeing this is that a distribution which is uniform in one variable, say θ, is not uniform in any non-trivial transformation of that variable, say $\cos\theta$ or $\tan\theta$. Real probability has therefore to be discarded.

2.4 Frequentist probability

Because of such difficulties, Real Probability was replaced by Frequentist Probability in the early 20th century. This is the usual definition taught in schools and undergraduate classes. A very readable account is given by von Mises [10]:

$$P_A = \lim_{N\to\infty} \frac{N_A}{N} \quad .$$

N is the total number of events in the ensemble (or collective). It can be visualised as a Venn diagram, as in Fig. 2.

The probability of a coin landing heads up is $\frac{1}{2}$ because if you toss a coin 1000 times, one side will come down ~ 500 times. That is an empirical definition (Frequentist probability has roots in the Vienna school and logical positivism). Similarly, the lifetime of a muon is 2.2μs because if you take 1000 muons and wait 2.2μs, then ~ 368 (that's a fraction e^{-1}) will remain.

With this definition P_A is not just a property of A but a joint property of A and the ensemble. The same coin will have a different probability for showing head depending on whether it is in a purse or

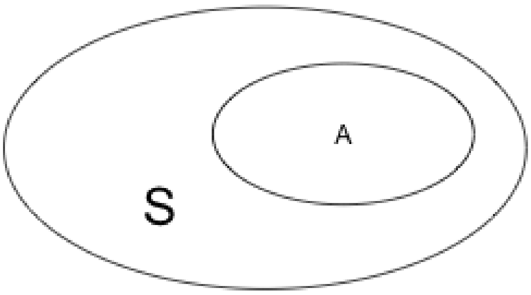


Fig. 2: Frequentist probability

in a numismatic collection. This leads to two distinctive properties (or, some would say, problems) for Frequentist probability.

Firstly, there may be more than one ensemble. To take an everyday example from von Mises, German life insurance companies pay out on 1.1% of 40 year old male clients. Your friend Hans is 40 today. What is the probability that he will survive to see his 41st birthday? 98.9% is an answer (if he's insured). But he is also a non-smoker and non-drinker—so perhaps the figure is higher (maybe 99.8%)? But if he drives a Harley-Davidson it should be lower (maybe 98.0%)? All these numbers are acceptable. The individual Hans belongs to several different ensembles, and the probability will be different for each of them.

To take an example from physics, suppose your experiment has a Particle Identification (PID) system using Cherenkov, time-of-flight and/or $\frac{dE}{dx}$ measurements. You want to talk about the probability that a K^+ will be correctly recognised by your PID. You determine this by considering many K^+ mesons and counting the number accepted to get $P = N_{acc}/N_{tot}$. But these will depend on the kaon sample you work with. It could be all kaons, or kaons above a certain energy threshold, or that actually enter the detector. The ensemble can be defined in various ways, each giving a valid but different value for the probability.

On the other hand, there may be no ensemble. To take an everyday example we might want to calculate the probability that it will rain tomorrow. This is impossible. There is only one tomorrow. It will either rain or not rain. P_{rain} is either 0 or 1, and we won't know which until tomorrow gets here. Von Mises insists that statements like 'It will probably rain tomorrow' are loose and unscientific.

To take an example from physics, consider the probability that there is a supersymmetric particle with mass below 2 TeV. Again, either there is or there isn't.

But, despite von Mises' objections, it does seem sensible, as the pressure falls and the gathering clouds turn grey, to say 'It will probably rain'. So this is a drawback to the frequentist definition. We will return to this and show how frequentists can talk meaningfully and quantitatively about unique events in the discussion of confidence intervals in Section 8.1.

2.5 Bayes' theorem

Before presenting Bayesian statistics we need to discuss Bayes' theorem, though we point out that Bayes' theorem applies (and is useful) in any probability model: it goes right back to the Kolmogorov axioms.

First we need to define the conditional probability: $P(A|B)$: this is the probability for A, given that B is true. For example: if a playing card is drawn at random from a pack of 52, then $P(\spadesuit A) = \frac{1}{52}$, but if you are told that the card is black, then $P(\spadesuit A|Black) = \frac{1}{26}$ (and obviously $P(\spadesuit A|Red) = 0$).

Bayes' theorem is just

$$P(A|B) = \frac{P(B|A)}{P(B)} \times P(A) \quad . \tag{2}$$

The proof is gratifyingly simple: the probability that A and B are both true can be written in two ways

$$P(A|B) \times P(B) = P(A\&B) = P(B|A) \times P(A) \quad .$$

Throw away middle term and divide by $P(B)$ to get the result.

As a first example, we go back to the ace of spades above. A card is drawn at random, and you are told that it is black. Bayes' theorem says

$$P(\spadesuit A|Black) = \frac{P(Black|\spadesuit A)}{P(Black)} P(\spadesuit A) = \frac{1}{\frac{1}{2}} \times \frac{1}{52} = \frac{1}{26} \quad ;$$

i.e. the original probability of drawing $\spadesuit A$, $\frac{1}{52}$, is multiplied by the probability that the ace of spades is black (just 1) and divided by the overall probability of drawing a black card ($\frac{1}{2}$) to give the obvious result.

For a less trivial example, suppose you have a momentum-selected beam which is 90% π and 10% K. This goes through a Cherenkov counter for which pions exceed the threshold velocity but kaons do not. In principle pions will give a signal, but suppose there is a 5% chance, due to inefficiencies, that they will not. Again in principle kaons always give no Cherenkov signal, but suppose that probability is only 95% due to background noise. What is the probability that a particle identified as a kaon, as it gave no signal, is truly one?

Bayes' theorem runs

$$P(K|no\ signal) = \frac{P(no\ signal|K)}{P(no\ signal)} \times P(K) = \frac{0.95}{0.95 \times 0.1 + 0.05 \times 0.9} \times 0.1 = 0.68 \quad ,$$

showing that the probability is only $\frac{2}{3}$. The positive identification is not enough to overwhelm the 9:1 $\pi : K$ ratio. Incidentally this uses the (often handy) expression for the denominator: $P(B) = P(B|A) \times P(A) + P(B|\overline{A}) \times \overline{P(A)}$.

2.6 Bayesian probability

The Bayesian definition of probability is that P_A represents your belief in A. 1 represents certainty, 0 represents total disbelief. Intermediate values can be calibrated by asking whether you would prefer to bet on A, or on a white ball being drawn from an urn containing a mix of white and black balls.

This avoids the limitations of Frequentist probability—coins, dice, kaons, rain tomorrow, existence of supersymmetry (SUSY) can all have probabilities assigned to them.

The drawback is that your value for P_A may be different from mine, or anyone else's. It is also called subjective probability.

Bayesian probability makes great use of Bayes' theorem, in the form

$$P(Theory|Data) = \frac{P(Data|Theory)}{P(Data)} \times P(Theory) \quad . \tag{3}$$

$P(Theory)$ is called the *prior*: your initial belief in $Theory$. $P(Data|Theory)$ is the *Likelihood*: the probability of getting $Data$ if $Theory$ is true. $P(Theory|Data)$ is the *Posterior*: your belief in $Theory$ in the light of a particular $Data$ being observed.

So this all works very sensibly. If the data observed is predicted by the theory, your belief in that theory is boosted, though this is moderated by the probability that the data could have arisen anyway. Conversely, if data is observed which is disfavoured by the theory, your belief in that theory is weakened.

The process can be chained. The posterior from a first experiment can be taken as the prior for a second experiment, and so on. When you write out the factors you find that the order doesn't matter.

2.6.1 Prior distributions

Often, though, the theory being considered is not totally defined: it may contain a parameter (or several parameters) such as a mass, coupling constant, or decay rate. Generically we will call this a, with the proviso that it may be multidimensional.

The prior is now not a single number $P(Theory)$ but a probability distribution $P_0(a)$. $\int_{a_1}^{a_2} P_0(a)\, da$ is your prior belief that a lies between a_1 and a_2. $\int_{-\infty}^{\infty} P_0(a)\, da$ is your original $P(Theory)$. This is generally taken as 1, which is valid provided the possibility that the theory that is false is matched by some value of a—for example if the coupling constant for a hypothetical particle is zero, that accommodates any belief that it might not exist. Bayes' theorem then runs:

$$P_1(a;x) \propto L(a;x) P_0(a) \quad . \tag{4}$$

If the range of a is infinite, $P_0(a)$ may be vanishingly small (this is called an 'improper prior'). However this is not a problem. Suppose, for example, that all we know about a is that it is non-negative, and we are genuinely equally open to its having any value. We write $P_0(a)$ as C, so $\int_{a_1}^{a_2} P_0(a)\, da = C(a_2 - a_1)$. This probability is vanishingly small: if you were offered the choice of a bet on a lying within the range $[a_1, a_2]$ or of drawing a white ball from an urn containing 1 white ball and N black balls, you would choose the latter, however large N was. However it is not zero: if the urn contained N black balls, but no white ball, your betting choice would change. After a measurement you have $P_1(a;x) = \frac{L(a;x)}{\int L(a';x) C da'} C$, and the factors of C can be cancelled (which, and this is the point, you could *not* do if C were exactly zero) giving $P_1(a;x) = \frac{L(a;x)}{\int L(a';x) da'}$ or, $P_1(a;x) \propto L(a;x)$, and you can then just normalize $P_1(a)$ to 1.

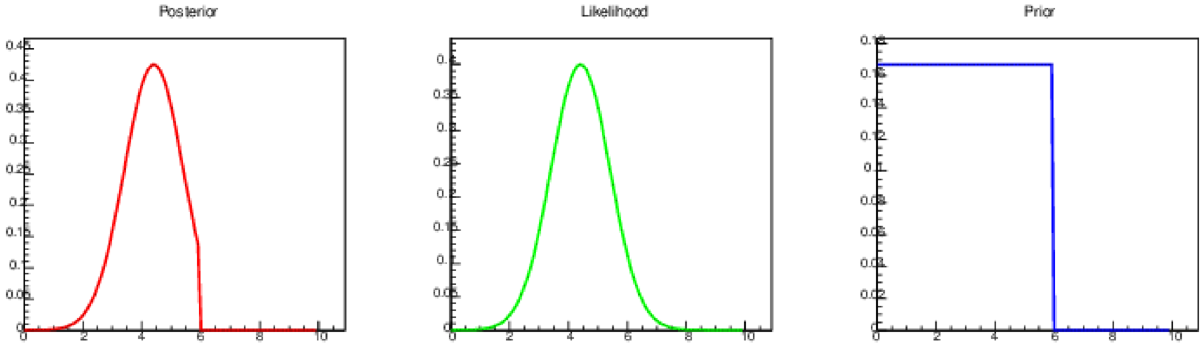


Fig. 3: Bayes at work

Figure 3 shows Eq. 4 at work. Suppose a is known to lie between 0 and 6, and the prior distribution is taken as flat, as shown in the right hand plot. A measurement of a gives a result 4.4 ± 1.0 , as shown in the central plot. The product of the two gives (after normalization) the posterior, as shown in the left hand plot.

2.6.2 Likelihood

The likelihood—the number $P(Data|Theory)$—is now generalised to the function $L(a, x)$, where x is the observed value of the data. Again, x may be multidimensional, but in what follows it is not misleading to ignore that.

This can be confusing. For example, anticipating Section 3.2.2, the probability of getting x counts from a Poisson process with mean a is

$$P(x,a) = e^{-a}\frac{a^x}{x!} \quad . \tag{5}$$

We also write

$$L(a,x) = e^{-a}\frac{a^x}{x!} \quad . \tag{6}$$

What's the difference? Technically there is none. These are identical joint functions of two variables (x and a) to which we have just happened to have given different names. Pragmatically we regard Eq. 5 as describing the probability of getting various different x from some fixed a, whereas Eq. 6 describes the likelihood for various different a from some given x. But be careful with the term 'likelihood'. If $P(x_1, a) > P(x_2, a)$ then x_1 is more probable (whatever you mean by that) than x_2. If $L(a_1, x) > L(a_2, x)$ it does not mean that a_1 is more likely (however you define that) than a_2.

2.6.3 Shortcomings of Bayesian probability

The big problem with Bayesian probability is that it is subjective. Your $P_0(a)$ and my $P_0(a)$ may be different—so how can we compare results? Science does, after all, take pride in being objective: it handles real facts, not opinions. If you present a Bayesian result from your search for the X particle this embodies the actual experiment and your irrational prior prejudices. I am interested in your experiment but not in your irrational prior prejudices—I have my own—and it is unhelpful if you combine the two.

Bayesians sometimes ask about the right prior they should use. This is the wrong question. The prior is what you believe, and only you know that.

There is an argument made for taking the prior as uniform. This is sometimes called the 'Principle of ignorance' and justified as being impartial. But this is misleading, even dishonest. If $P_0(a)$ is taken as constant, favouring no particular value, then it is not constant for a^2 or $\sqrt{a}$ or $\ln a$, which are equally valid parameters.

It is true that with lots of data, $P_1(a)$ decouples from $P_0(a)$ (this is the Bernstein-von Mises theorem). The final result depends only on the measurements. But this is not the case with little data—and that's the situation we're usually in—when doing statistics properly matters.

As an example, suppose you make a Gaussian measurement (anticipating slightly Section 3.2.3). You consider a prior flat in a and a prior flat in $\ln a$. This latter is quite sensible—it says you expect a result between 0.1 and 0.2 as being equally likely as a result between 1 and 2, or 10 and 20. The posteriors are shown in Fig. 4. For an 'accurate' result of 3 ± 0.5 the posteriors are very close. For an 'intermediate' result of 4.0 ± 1.0 there is an appreciable difference in the peak value and the shape. For a 'poor' measurement of 5.0 ± 2.0 the posteriors are *very* different.

So you should never just quote results from a single prior. Try several forms of prior and examine the spread of results. If they are pretty much the same you are vindicated. This is called 'robustness under choice of prior' and it is standard practice for statisticians. If they are different then the data are telling you about the limitations of your results.

2.6.4 Jeffreys' prior

Jeffreys [11] suggested a technique now known as the Jeffreys' or *objective prior*: that you should choose a prior flat in a transformed variable a' for which the Fisher information, $\mathcal{I} = -\left\langle \frac{\partial^2 L(x;a)}{\partial a^2} \right\rangle$ is constant. The Fisher information (which is important in maximum likelihood estimation, as described in Section 5.2) is a measure of how much a measurement tells you about the parameter: a large $\mathcal{I}$ has a likelihood function with a sharp peak and will tell you (by some measure) a lot about a; a small $\mathcal{I}$ has a featureless likelihood function which will not be useful. Jeffrey's principle is that the prior should not favour or disfavour particular values of the parameter. It is equivalently—and more conveniently—used as taking a prior in the original a which is proportional to $\sqrt{\mathcal{I}}$.

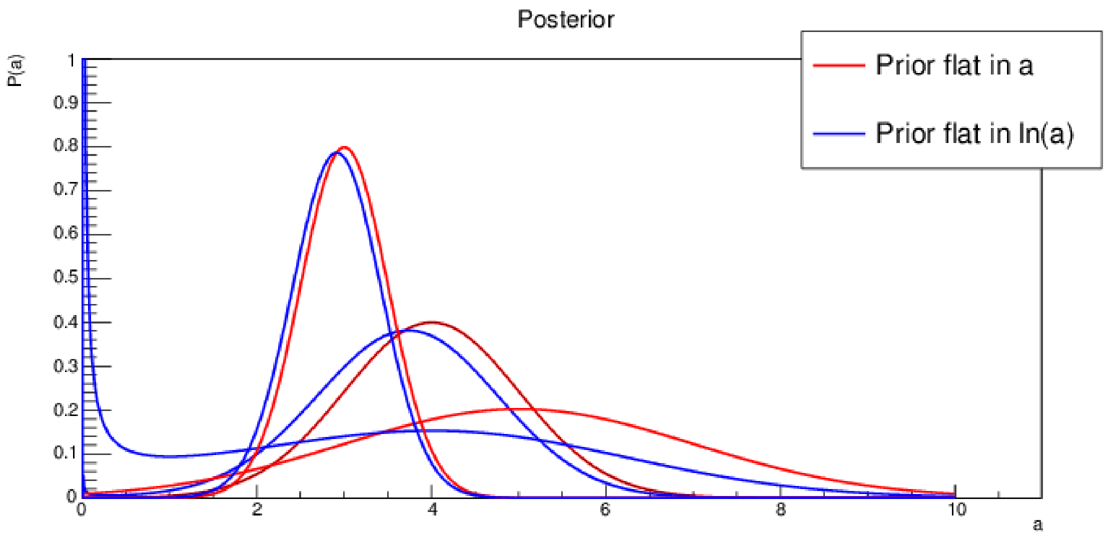


Fig. 4: Posteriors for two different priors for the results 3.0 ± 0.5, 4.0 ± 1.0 and 5.0 ± 2.0

It has not been universally adopted for various reasons. Some practitioners like to be able to include their own prior belief into the analysis. It also makes the prior dependent on the experiment (in the form of the likelihood function). Thus if ATLAS and CMS searched for the same new X particle they would use different priors for $P_0(M_X)$, which is (to some people) absurd.

So it is not universal—but when you are selecting a bunch of priors to test robustness—the Jeferys' prior is a strong contender for inclusion.

2.7 Summary

So mathematical probability has no meaning, and real probability is discredited. That leaves the Frequentist and Bayesian definitions. Both are very much in use.

They are sometimes presented as rivals, with adherents on either side ('Frequentists versus Bayesians'). This is needless drama. They are both tools that help us understand our results. Both have drawbacks. Sometimes it is clear which is the best tool for a particular job, sometimes it is not and one is free to choose either. It is said—probably accurately—that particle physicists feel happier with Frequentist probability as they are used to large ensembles of similar but different events, whereas astrophysicists and cosmologists are more at home with Bayesian probability as they only have one universe to consider.

What is important is not which version you prefer—these are not football teams—but that you know the limitations of each, that you use the best definition when there is a reason to do so, and, above all, that you are aware of which form you are using.

As a possibly heretical afterthought, perhaps classical probability still has a place? Quantum Mechanics, after all, gives probabilities. If P_A is not 'real'—either because it depends on an arbitrary ensemble, or because is a subjective belief—then it looks like there is nothing 'real' in the universe.

The state of a coin—or an electron spin—having probability $\frac{1}{2}$ makes sense. There is a symmetry that dictates it. The lifetime of a muon—i.e. probability per unit time that it will decay—seems to be a well-defined quantity, a property of the muon and independent of any ensemble, or any Bayesian belief.

The probability a muon will produce a signal in your muon detector seems like a 'real well-defined quantity', if you specify the 4 momentum and the state of the detector. Of course the inverse probability

'What is the probability that a muon signal in my detector comes from a real muon, not background' is not intrinsically defined, So perhaps classical probability has a place in physics—but not in interpreting results. However you should not mention this to a statistician or they will think you're crazy.

3 Probability distributions and their properties

We have to make a simple distinction between two sorts of data: *integer* data and *real-number* data[1].

The first covers results which are of their nature whole numbers: the numbers of kaons produced in a collision, or the number of entries falling into some bin of a histogram. Generically let's call such numbers r. They have probabilities $P(r)$ which are dimensionless.

The second covers results whose values are real (or floating-point) numbers. There are lots of these: energies, angles, invariant masses . . . Generically let's call such numbers x, and they have probability density functions $P(x)$ which have dimensions of $[x]^{-1}$, so $\int_{x_1}^{x_2} P(x)dx$ or $P(x)\,dx$ are probabilities.

You will also sometimes meet the cumulative distribution $C(x) = \int_{-\infty}^{x} P(x')\,dx'$.

3.1 Expectation values

From $P(r)$ or $P(x)$ one can form the expectation value

$$\langle f \rangle = \sum_r f(r)P(r) \qquad \text{or} \qquad \langle f \rangle = \int f(x)P(x)\,dx \quad , \tag{7}$$

where the sum or integral is taken as appropriate. Some authors write this as $E(f)$, but I personally prefer the angle-bracket notation. You may think it looks too much like quantum mechanics, but in fact it's quantum mechanics which looks like statistics: an expression like $\langle\psi|\hat{Q}|\psi\rangle$ is the average value of an operator $\hat{Q}$ in some state ψ, where 'average value' has exactly the same meaning and significance.

3.1.1 Mean and standard deviation

In particular the *mean*, often written μ, is given by

$$\langle r \rangle = \textstyle\sum_r rP(r) \qquad \text{or} \qquad \langle x \rangle = \int xP(x)\,dx \quad .$$

Similarly one can write higher *moments*

$$\mu_k = \langle r^k \rangle = \textstyle\sum_r r^k P(r) \qquad \text{or} \qquad \langle x^k \rangle = \int x^k P(x)\,dx \quad ,$$

and *central moments*

$$\mu'_k = \langle (r-\mu)^k \rangle = \textstyle\sum_r (r-\mu)^k P(r) \qquad \text{or} \qquad \langle (x-\mu)^k \rangle = \int (x-\mu)^k P(x)\,dx \quad .$$

The second central moment is known as the *variance*

$$\mu'_2 = V = \textstyle\sum_r (r-\mu)^2 P(r) = \langle r^2 \rangle - \langle r \rangle^2 \qquad \text{or} \qquad \int (x-\mu)^2 P(x)\,dx = \langle x^2 \rangle - \langle x \rangle^2$$

It is easy to show that $\langle (x-\mu)^2 \rangle = \langle x^2 \rangle - \mu^2$. The *standard deviation* is just the square root of the variance $\sigma = \sqrt{V}$.

Statisticians usually use variance, perhaps because formulae come out simpler. Physicists usually use standard deviation, perhaps because it has the same dimensions as the variable being studied, and can be drawn as an error bar on a plot.

You may also meet *skew*, which is $\gamma = \langle (x-\mu)^3 \rangle / \sigma^3$ and *kurtosis*, $h = \langle (x-\mu)^4 \rangle / \sigma^4 - 3$. Definitions vary, so be careful. Skew is a dimensionless measure of the asymmetry of a distribution. Kurtosis is (thanks to that rather arbitrary looking 3 in the definition) zero for a Gaussian distribution (see Section 3.2.3): positive kurtosis indicates a narrow core with a wide tail, negative kurtosis indicates the tails are reduced.

[1] Other branches of science have to include a third, *categorical* data, but we will ignore that.

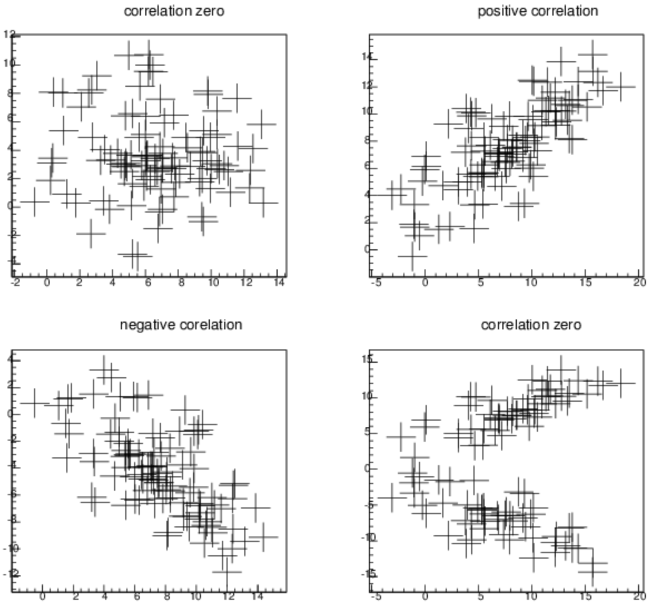


Fig. 5: Examples of two dimensional distributions. The top right has positive covariance (and correlation), the bottom left negative. In the top left the covariance is zero and x and y are independent; in the bottom right the covariance is also zero, but they are not independent.

3.1.2 Covariance and correlation

If your data are 2-dimensional pairs (x, y), then besides forming $\langle x \rangle, \langle y \rangle, \sigma_x$ etc., you can also form the *Covariance*

$$\mathrm{Cov}(x, y) = \langle (x - \mu_x)(y - \mu_y) \rangle = \langle xy \rangle - \langle x \rangle \langle y \rangle \quad .$$

Examples are shown in Fig. 5. If there is a tendency for positive fluctuations in x to be associated with positive fluctuations in y (and therefore negative with negative) then the product $(x_i - \overline{x})(y_i - \overline{y})$ tends to be positive and the covariance is greater than 0. A negative covariance, as in the 3rd plot, happens if a positive fluctuation in one variable is associated with a negative fluctuation in the other. If the variables are independent then a positive variation in x is equally likely to be associated with a positive or a negative variation in y and the covariance is zero, as in the first plot. However the converse is not always the case, there can be two-dimensional distributions where the covariance is zero, but the two variables are not independent, as is shown in the fourth plot.

Covariance is useful, but it has dimensions. Often one uses the *correlation*, which is just

$$\rho = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y} \quad . \tag{8}$$

It is easy to show that ρ lies between 1 (complete correlation) and -1 (complete anticorrelation). $\rho = 0$ if x and y are independent.

If there are more than two variables—the alphabet runs out so let's call them $(x_1, x_2, x_3 \dots x_n)$—then these generalise to the *covariance matrix*

$$\mathbf{V}_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$$

and the *correlation matrix*

$$\rho_{ij} = \frac{\mathbf{V}_{ij}}{\sigma_i \sigma_j} \quad .$$

The diagonal of $\mathbf{V}$ is σ_i^2. The diagonal of ρ is 1.

3.2 Binomial, Poisson and Gaussian

We now move from considering the general properties of distributions to considering three specific ones. These are the ones you will most commonly meet for the distribution of the original data (as opposed to quantities constructed from it). Actually the first, the binomial, is not nearly as common as the second, the Poisson; and the third, the Gaussian, is overwhelmingly more common. However it is useful to consider all three as concepts are built up from the simplest to the more sophisticated.

3.2.1 The binomial distribution

The binomial distribution is easy to understand as it basically describes the familiar tossing of coins. It describes the number r of successes in N trials, each with probability p of success. r is discrete so the process is described by a probability distribution

$$P(r;p,N) = \frac{N!}{r!(N-r)!}p^r q^{N-r} \quad , \tag{9}$$

where $q \equiv 1 - p$.

Some examples are shown in Fig. 6.

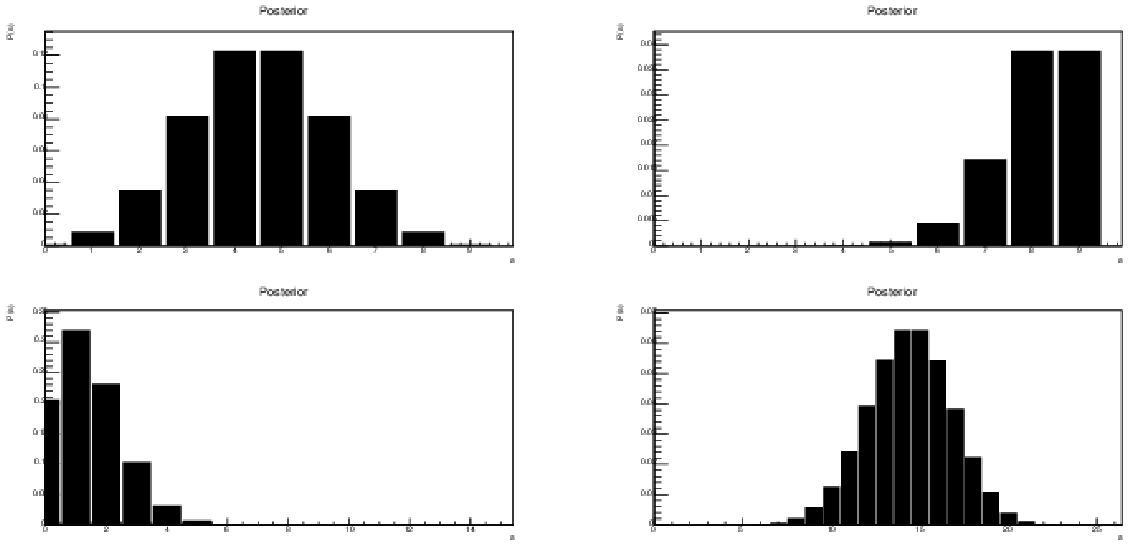


Fig. 6: Some examples of the binomial distribution, for (1) $N = 10, p = 0.6$, (2) $N = 10, p = 0.9$, (3) $N = 15, p = 0.1$, and (4) $N = 25, p = 0.6$.

The distribution has mean $\mu = Np$, variance $V = Npq$, and standard deviation $\sigma = \sqrt{Npq}$.

3.2.2 The Poisson distribution

The Poisson distribution also describes the probability of some discrete number r, but rather than a fixed number of 'trials' it considers a random rate λ:

$$P(r;\lambda) = e^{-\lambda}\frac{\lambda^r}{r!} \quad . \tag{10}$$

It is linked to the binomial—the Poisson is the limit of the binomial—as $N \to \infty$, $p \to 0$ with

$np = \lambda = constant$. Figure 7 shows various examples. It has mean $\mu = \lambda$, variance $V = \lambda$, and standard deviation $\sigma = \sqrt{\lambda} = \sqrt{\mu}$.

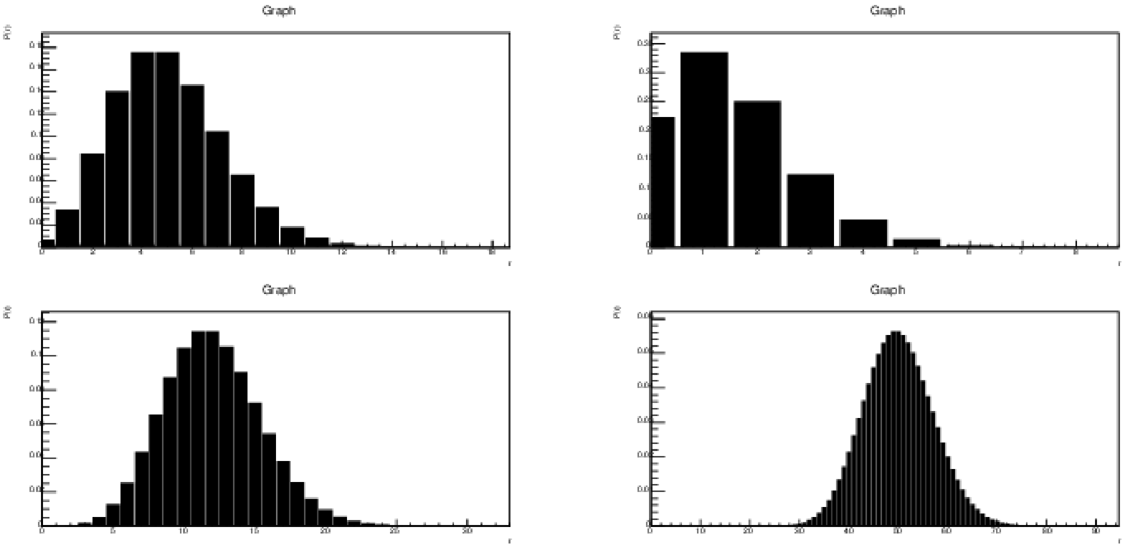


Fig. 7: Poisson distributions for (1) $\lambda = 5$, (2) $\lambda = 1.5$, (3) $\lambda = 12$ and (4) $\lambda = 50$

The clicks of a Geiger counter are the standard illustration of a Poisson process. You will meet it a lot as it applies to event counts—on their own or in histogram bins.

To help you think about the Poisson, here is a simple question (which describes a situation I have seen in practice, more than once, from people who ought to know better).

You need to know the efficiency of your PID system for positrons.

You find 1000 data events where 2 tracks have a combined mass of 3.1 GeV (J/ψ) and the negative track is identified as an e^- ('Tag-and-probe' technique).

In 900 events the e^+ is also identified. In 100 events it is not. The efficiency is 90%.

What about the error?

Colleague A says $\sqrt{900} = 30$ so efficiency is $90.0 \pm 3.0\%$.

Colleague B says $\sqrt{100} = 10$ so efficiency is $90.0 \pm 1.0\%$.

Which is right?

Please think about this before turning the page...

Neither—both are wrong. This is binomial not Poisson: $p = 0.9, N = 1000$.
The error is $\sqrt{Npq} = \sqrt{1000 \times 0.9 \times 0.1}$ (or $\sqrt{1000 \times 0.1 \times 0.9}$) $=\sqrt{90} = 9.49$ so the efficiency is 90.0 ± 0.9 %.

3.2.3 *The Gaussian distribution*

This is by far the most important statistical distribution. The probability density function (PDF) for a variable x is given by the formula

$$P(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad . \tag{11}$$

Pictorially this is shown in Fig. 8.

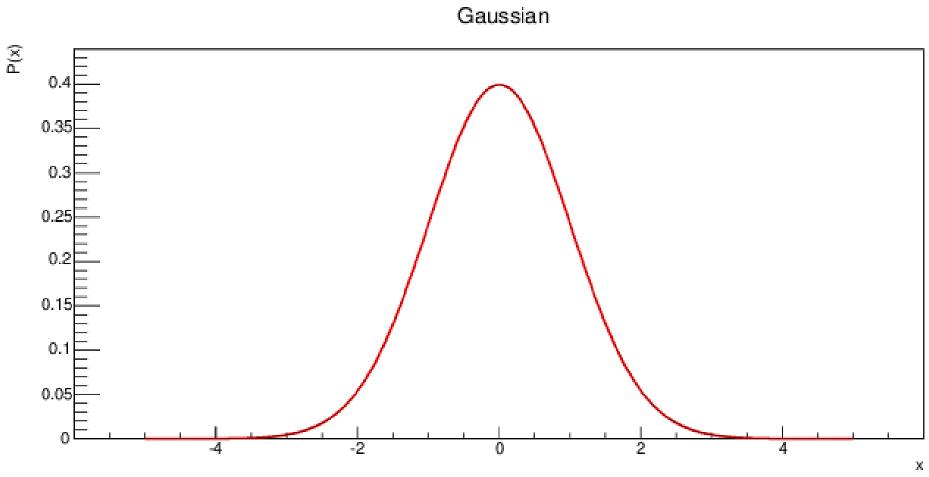


Fig. 8: The Gaussian distribution

This is sometimes called the 'bell curve', though in fact a real bell does not have flared edges like that. There is (in contrast to the Poisson and binomial) only one Gaussian curve, as μ and σ are just location and scale parameters.

The mean is μ and the standard deviation is σ. The Skew is zero, as it is symmetric, and the kurtosis is zero by construction.

In statistics, and most disciplines, this is known as the *normal distribution*. Only in physics is it known as 'The Gaussian'—perhaps because the word 'normal' already has so many meanings.

The reason for the importance of the Gaussian is the *central limit theorem* (CLT) that states: if the variable X is the sum of N independent variables $x_1, x_2 \dots x_N$ then:

1. Means add: $\langle X \rangle = \langle x_1 \rangle + \langle x_2 \rangle + \dots \langle x_N \rangle$,
2. Variances add: $V_X = V_1 + V_2 + \dots V_N$,
3. If the variables x_i are identically distributed then $P(X)$ tends to a Gaussian for large N.

(1) is obvious, (2) is pretty obvious, and means that standard deviations add in quadrature, and that the standard deviation of an average falls like $\frac{1}{\sqrt{N}}$, (3) applies whatever the form of the original $P(x)$.

Before proving this, it is helpful to see a demonstration to convince yourself that the implausible assertion in (3) actually does happen. Take a uniform distribution from 0 to 1, as shown in the top left subplot of Fig. 9. It is flat. Add two such numbers and the distribution is triangular, between 0 and 2, as shown in the top right.

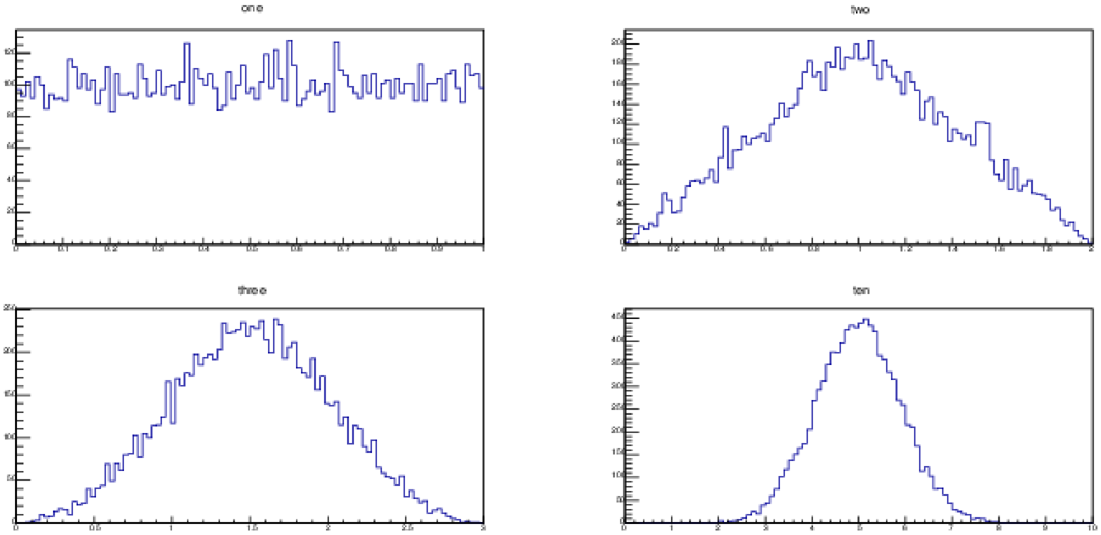


Fig. 9: Demonstration of the central limit theorem

With 3 numbers, at the bottom left, it gets curved. With 10 numbers, at the bottom right, it looks pretty Gaussian. The proof follows.

Proof. First, introduce the characteristic function $\langle e^{ikx}\rangle = \int e^{ikx} P(x)\,dx = \tilde{P}(k)$.

This can usefully be thought of as an expectation value and as a Fourier transform.

Expand the exponential as a series

$\langle e^{ikx}\rangle = \langle 1 + ikx + \frac{(ikx)^2}{2!} + \frac{(ikx)^3}{3!} \dots\rangle = 1 + ik\langle x\rangle + (ik)^2\frac{\langle x^2\rangle}{2!} + (ik^3)\frac{\langle x^3\rangle}{3!} \dots \quad .$

Take the logarithm and use the expansion $\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} \dots \quad .$ This gives a power series in (ik), where the coefficient $\frac{\kappa_r}{r!}$ of $(ik)^r$ is made up of expectation values of x of total power r:

$\kappa_1 = \langle x\rangle, \kappa_2 = \langle x^2\rangle - \langle x\rangle^2 =, \kappa_3 = \langle x^3\rangle - 3\langle x^2\rangle\langle x\rangle + 2\langle x\rangle^3 \dots \quad .$

These are called the semi-invariant cumulants of Thièle , under a change of scale α, $\kappa_r \to \alpha^r\kappa_r$. Under a change in location, only κ_1 changes.

If X is the sum of independent and identically distributed (i.i.d.) random variables, $x_1+x_2+x_3$..., then $P(X)$ is the convolution of $P(x)$ with itself N times.

The Fourier Transform of a convolution is the product of the individual Fourier Transforms,

the logarithm of a product is the sum of the logarithms, so $P(X)$ has cumulants $K_r = N\kappa_r$.

To make graphs commensurate, you need to scale the X axis by the standard deviation, which grows like $\sqrt{N}$. The cumulants of the scaled graph are $K'_r = N^{1-r/2}\kappa_r$.

As $N \to \infty$, these vanish for $r > 2$, leaving a quadratic.

If the log is a quadratic, the exponential is a Gaussian. So $\tilde{P}(X)$ is Gaussian.

And finally, the inverse Fourier Transform of a Gaussian is also a Gaussian. □

Even if the distributions are not identical, the CLT tends to apply, unless one (or two) dominates. Most 'errors' fit this, being compounded of many different sources.

4 Hypothesis testing

'Hypothesis testing' is another piece of statistical technical jargon. It just means 'making choices'—in a logical way—on the basis of statistical information.

- Is some track a pion or a kaon?
- Is this event signal or background?
- Is the detector performance degrading with time?
- Do the data agree with the Standard Model prediction or not?

To establish some terms: you have a *hypothesis* (the track is a pion, the event is signal, the detector is stable, the Standard Model is fine . . .). and an alternative hypothesis (kaon, background, changing, new physics needed . . .) Your hypothesis is usually *simple* i.e. completely specified, but the alternative is often *composite* containing a parameter (for example, the detector decay rate) which may have any non-zero value.

4.1 Type I and type II errors

As an example, let's use the signal/background decision. Do you accept or reject the event (perhaps in the trigger, perhaps in your offline analysis)? To make things easy we consider the case where both hypotheses are simple, i.e. completely defined.

Suppose you measure some parameter x which is related to what you are trying to measure. It may well be the output from a neural network or other machine learning (ML) systems. The expected distributions for x under the hypothesis and the alternative, S and B respectively, are shown in Fig. 10.

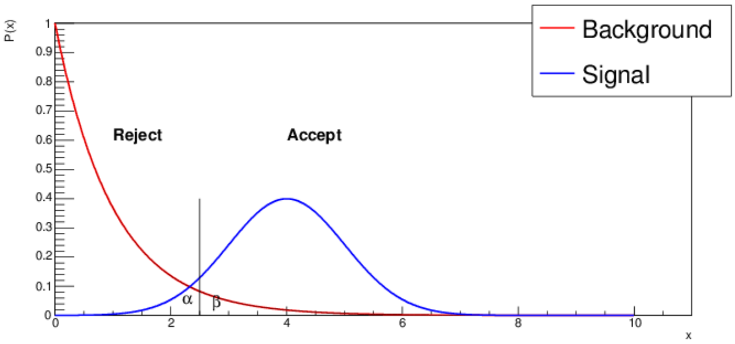


Fig. 10: Hypothesis testing example

You impose a cut as shown—you have to put one somewhere—accepting events above $x = x_{cut}$ and rejecting those below.

This means losing a fraction α of signal. This is called a *type I error* and α is known as the *significance*.

You admit a fraction β of background. This is called a *type II error* and $1 - \beta$ is the power.

You would like to know the best place to put the cut. This graph cannot tell you! The strategy for the cut depends on three things—hypothesis testing only covers one of them.

The second is the prior signal to noise ratio. These plots are normalized to 1. The red curve is (probably) MUCH bigger. A value of β of, say, 0.01 looks nice and small—only one in a hundred background events get through. But if your background is 10,000 times bigger than your signal (and it often is) you are still swamped.

The third is the cost of making mistakes, which will be different for the two types of error. You have a trade-off between efficiency and purity: what are they worth? In a typical analysis, a type II error is more serious than a type I: losing a signal event is regrettable, but it happens. Including background events in your selected pure sample can give a very misleading result. By contrast, in medical decisions, type I errors are much worse than type II. Telling healthy patients they are sick leads to worry and perhaps further tests, but telling sick patients they are healthy means they don't get the treatment they need.

4.2 The Neymann-Pearson lemma

In Fig. 10 the strategy is plain—you choose x_{cut} and evaluate α and β. But suppose the S and B curves are more complicated, as in Fig. 11? Or that x is multidimensional?

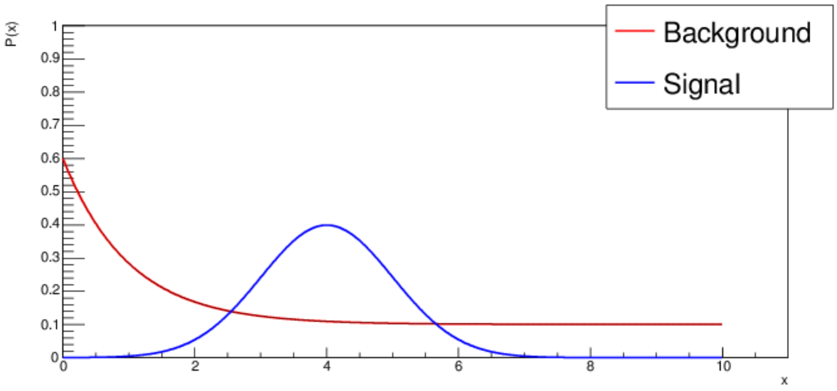


Fig. 11: A more complicated case for hypothesis testing

Neymann and Pearson say: your acceptance region just includes regions of greatest $\frac{S(x)}{B(x)}$ (the ratio of likelihoods). For a given α, this gives the smallest β ('Most powerful at a given significance')

The proof is simple: having done this, if you then move a small region from 'accept' to 'reject' it has to be replaced by an equivalent region, to balance α, which (by construction) brings more background, increasing β.

However complicated, such a problem reduces to a single monotonic variable $\frac{S}{B}$, and you cut on that.

4.3 Efficiency, purity, and ROC plots

ROC plots are often used to show the efficacy of different selection variables. You scan over the cut value (in x, for Fig. 10 or in S/B for a case like Fig. 11 and plot the fraction of background accepted (β) against fraction of signal retained ($1 - \alpha$), as shown in Fig. 12.

For a very loose cut all data is accepted, corresponding to a point at the top right. As the cut is tightened both signal and background fractions fall, so the point moves to the left and down, though hopefully the background loss is greater than the signal loss, so it moves more to the left than it does downwards. As the cut is increased the line moves towards the bottom left, the limit of a very tight cut where all data is rejected.

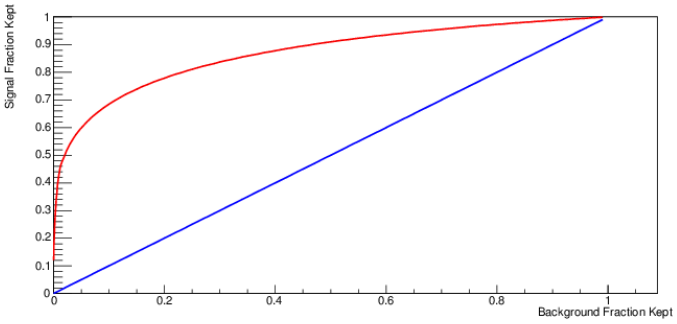


Fig. 12: ROC curves

A diagonal line corresponds to no discrimination—the S and B curves are identical. The further the actual line bulges away from that diagonal, the better.

Where you should put your cut depends, as pointed out earlier, also on the prior signal/background ratio and the relative costs of errors. The ROC plots do not tell you that, but they can be useful in comparing the performance of different discriminators.

The name 'ROC' stands for 'receiver operating characteristic', for reasons that are lost in history. Actually it is good to use this meaningless acronym, otherwise they get called 'efficiency-purity plots' even though they definitely do not show the purity (they cannot, as that depends on the overall signal/background ratio). Be careful, as the phrases 'background efficiency', 'contamination', and 'purity' are used ambiguously in the literature.

4.4 The null hypothesis

An analysis is often (but not always) investigating whether an effect is present, motivated by the hope that the results will show that it is:

- Eating broccoli makes you smart.
- Facebook advertising increases sales.
- A new drug increases patient survival rates.
- The data show Beyond-the-Standard-Model physics.

To reach such a conclusion you have to use your best efforts to try, and to fail, to prove the opposite: the *Null Hypothesis* H_0.

- Broccoli lovers have the same or smaller IQ than broccoli loathers.
- Sales are independent of the Facebook advertising budget.
- The survival rates for the new treatment is the same.
- The Standard Model (functions or Monte-Carlo) describe the data.

If the null hypothesis is not tenable, you've proved—or at least, supported—your point.

The reason for calling α the 'significance' is now clear. It is the probability that the null hypothesis will be wrongly rejected, and you'll claim an effect where there isn't any.

There is a minefield of difficulties. Correlation is not causation. If broccoli eaters are more intelligent, perhaps that's because it's intelligent to eat green vegetables, not that vegetables make you intelligent. One has to consider that if similar experiments are done, self-censorship will influence which results get published. This is further discussed in Section 9.

This account is perhaps unconventional in introducing the null hypothesis at such a late stage. Most treatments bring it in right at the start of the description of hypothesis testing, because they assume that all decisions are of this type.

5 Estimation

What statisticians call 'estimation', physicists would generally call 'measurement'.

Suppose you know the probability (density) function $P(x; a)$ and you take a set of data $\{x_i\}$. What is the best value for a? (Sometimes one wants to estimate a property (e.g. the mean) rather than a parameter, but this is relatively uncommon, and the methodology is the same.)

x_i may be single values, or pairs, or higher-dimensional. The unknown a may be a single parameter or several. If it has more than one component, these are sometimes split into 'parameters of interest' and 'nuisance parameters'.

The *estimator* is defined very broadly: an estimator $\hat{a}(x_1 \dots x_N)$ is a function of the data that gives a value for the parameter a. There is no 'correct' estimator, but some are better than others. A perfect estimator would be:

- Consistent. $\hat{a}(x_1 \dots x_N) \to a$ as $N \to \infty$,
- Unbiased: $\langle \hat{a} \rangle = a$,
- Efficient: $\langle (\hat{a} - a)^2 \rangle$ is as small as possible,
- Invariant: $\hat{f}(a) = f(\hat{a})$.

No estimator is perfect—these 4 goals are incompatible. In particular the second and the fourth; if an estimator $\hat{a}$ is unbiased for a then $\sqrt{\hat{a}}$ is not an unbiased estimator of $\sqrt{a}$.

5.1 Bias

Suppose we estimate the mean by taking the obvious[2] $\hat{\mu} = \overline{x}$

$\langle \hat{\mu} \rangle = \left\langle \frac{1}{N} \sum x_i \right\rangle = \frac{1}{N} \sum \mu = \mu.$

So there is no bias. This expectation value of this estimator of μ is just μ itself. By contrast suppose we estimate the variance by the apparently obvious $\hat{V} = \overline{x^2} - \overline{x}^2$.

Then $\left\langle \hat{V} \right\rangle = \left\langle \overline{x^2} \right\rangle - \left\langle \overline{x}^2 \right\rangle$.

The first term is just $\left\langle x^2 \right\rangle$. To make sense of the second term, note that $\langle x \rangle = \langle \overline{x} \rangle$ and add and subtract $\langle x \rangle^2$ to get

$\left\langle \hat{V} \right\rangle = \left\langle x^2 \right\rangle - \langle x \rangle^2 - (\left\langle \overline{x}^2 \right\rangle - \langle \overline{x} \rangle^2)$

$\left\langle \hat{V} \right\rangle = V(x) - V(\overline{x}) = V - \frac{V}{N} = \frac{N-1}{N} V.$

So the estimator is biased! $\hat{V}$ will, on average, give too small a value.

This bias, like any known bias, can be corrected for. Using $\hat{V} = \frac{N}{N-1}(\overline{x^2} - \overline{x}^2)$ corrects the bias. The familiar estimator for the standard deviation follows: $\hat{\sigma} = \sqrt{\frac{\sum_i (x_i - \overline{x})^2}{N-1}}$.

(Of course this gives a biased estimate of σ. But V is generally more important in this context.)

[2] Note the difference between $\langle x \rangle$ which is an average over a PDF and $\overline{x}$ which denotes the average over a particular sample: both are called 'the mean x'.

5.2 Efficiency

Somewhat surprisingly, there is a limit to the efficiency of an estimator: the *minimum variance bound* (MVB), also known as the *Cramér-Rao bound*.

For any unbiased estimator $\hat{a}(x)$, the variance is bounded

$$V(\hat{a}) \geq -\frac{1}{\left\langle \frac{d^2 \ln L}{da^2} \right\rangle} = \frac{1}{\left\langle \left(\frac{d \ln L}{da} \right)^2 \right\rangle} \quad . \tag{12}$$

L is the likelihood (as introduced in Section 2.6.2) of a sample of independent measurements, i.e. the probability for the whole data sample for a particular value of a. It is just the product of the individual probabilities:

$L(a; x_1, x_2, ...x_N) = P(x_1; a)P(x_2; a)...P(x_N; a)$.

We will write $L(a; x_1, x_2, ...x_N)$ as $L(a; x)$ for simplicity.

Proof. Proof of the MVB

Unitarity requires $\int P(x; a)\, dx = \int L(a; x)\, dx = 1$

Differentiate wrt a:

$$0 = \int \frac{dL}{da}\, dx = \int L \frac{d \ln L}{da}\, dx = \left\langle \frac{d \ln L}{da} \right\rangle \tag{13}$$

If $\hat{a}$ is unbiased $\langle \hat{a} \rangle = \int \hat{a}(x)P(x; a)\, dx = \int \hat{a}(x)L(a; x)\, dx = a$

Differentiate wrt a: $\qquad 1 = \int \hat{a}(x) \frac{dL}{da}\, dx = \int \hat{a} L \frac{d \ln L}{da}\, dx$

Subtract Eq. 13 multiplied by a, and get $\int (\hat{a} - a) \frac{d \ln L}{da} L dx = 1$

Invoke the Schwarz inequality $\int u^2\, dx \int v^2\, dx \geq \left(\int uv\, dx \right)^2$ with $u \equiv (\hat{a} - a)\sqrt{L}, v \equiv \frac{d \ln L}{da} \sqrt{L}$

Hence $\int (\hat{a} - a)^2 L\, dx \int \left(\frac{d \ln L}{da} \right)^2 L\, dx \geq 1$

$$\left\langle (\hat{a} - a)^2 \right\rangle \geq 1 / \left\langle \left(\frac{dlnL}{da} \right)^2 \right\rangle \tag{14}$$

□

Differentiating Eq. 13 again gives

$\frac{d}{da} \int L \frac{d \ln L}{da}\, dx = \int \frac{dL}{da} \frac{d \ln L}{da}\, dx + \int L \frac{d^2 \ln A}{da^2}\, dx = \left\langle \left(\frac{d \ln L}{da} \right)^2 \right\rangle + \left\langle \frac{d^2 \ln L}{da^2} \right\rangle = 0,$

hence $\left\langle \left(\frac{d \ln L}{da} \right)^2 \right\rangle = -\left\langle \frac{d^2 \ln L}{da^2} \right\rangle$.

This is the *Fisher information* referred to in Section 2.6.4. Note how it is intrinsically positive.

5.3 Maximum likelihood estimation

The *maximum likelihood* (ML) estimator just does what it says: a is adjusted to maximise the likelihood of the sample (for practical reasons one actually maximises the log likelihood, which is a sum rather than a product).

$$\text{Maximise} \ln L = \sum_i \ln P(x_i; a) \quad , \tag{15}$$

$$\left. \frac{d \ln L}{da} \right|_{\hat{a}} = 0 \quad . \tag{16}$$

The ML estimator is very commonly used. It is not only simple and intuitive, it has lots of nice properties.

- It is consistent.
- It is biased, but bias falls like $1/N$.
- It is efficient for the large N.
- It is invariant—doesn't matter if you reparametrize a.

A particular maximisation problem may be solved in 3 ways, depending on the complexity

1. Solve Eq. 16 algebraically,
2. Solve Eq. 16 numerically, and
3. Solve Eq. 15 numerically.

5.4 Least squares

Least squares estimation follows from maximum likelihood estimation. If you have Gaussian measurements of y taken at various x values, with measurement error σ, and a prediction $y = f(x; a)$ then the Gaussian probability

$$P(y; x, a) = \tfrac{1}{\sigma\sqrt{2\pi}} e^{-(y-f(x,a))^2/2\sigma^2}$$

gives the log likelihood

$$\ln L = -\sum \tfrac{(y_i - f(x_i;a))^2}{2\sigma_i^2} + \text{constants}.$$

To maximise $\ln L$, you minimise $\chi^2 = \sum \frac{(y_i - f(x_i;a))^2}{\sigma_i^2}$, hence the name 'least squares'.

Differentiating gives the *normal equations*: $\sum \frac{(y_i - f(x_i;a))}{\sigma_i^2} f'(x_i; a) = 0$.

If $f(x; a)$ is linear in a then these can be solved exactly. Otherwise an iterative method has to be used.

5.5 Straight line fits

As a particular instance of least squares estimation, suppose the function is $y = mx + c$, and assume all σ_i are the same (the extension to the general case is straightforward). The normal equations are then $\sum(y_i - mx_i - c)x_i = 0$ and $\sum(y_i - mx_i - c) = 0$, for which the solution, shown in Fig. 13, is $m = \frac{\overline{xy} - \overline{x}\ ,\overline{y}}{\overline{x^2} - \overline{x}^2}$, $c = \overline{y} - m\overline{x}$.

Statisticians call this *regression*. Actually there is a subtle difference, as shown in Fig. 14.

The straight line fit considers well-defined x values and y values with measurement errors—if it were not for those errors then presumably the values would line up perfectly, with no scatter. The scatter in regression is not caused by measurement errors, but by the fact that the variables are linked only loosely.

The history of regression started with Galton, who measured the heights of fathers and their (adult) sons. Tall parents tend to have tall children so there is a correlation. Because the height of a son depends not just on his paternal genes but on many factors (maternal genes, diet, childhood illnesses . . .), the points do not line up exactly—and using a high accuracy laser interferometer to do the measurements, rather than a simple ruler, would not change anything.

Galton, incidentally, used this to show that although tall fathers tend to have tall sons, they are not that tall. An outstandingly tall father will have (on average) quite tall children, and only tallish grandchildren. He called this 'Regression towards mediocrity', hence the name.

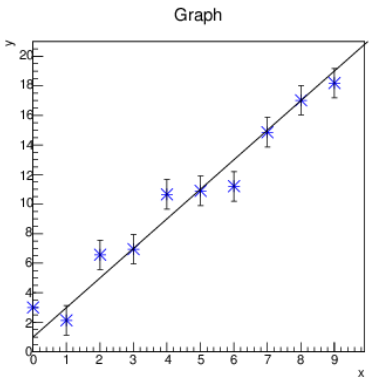


Fig. 13: A straight line fit

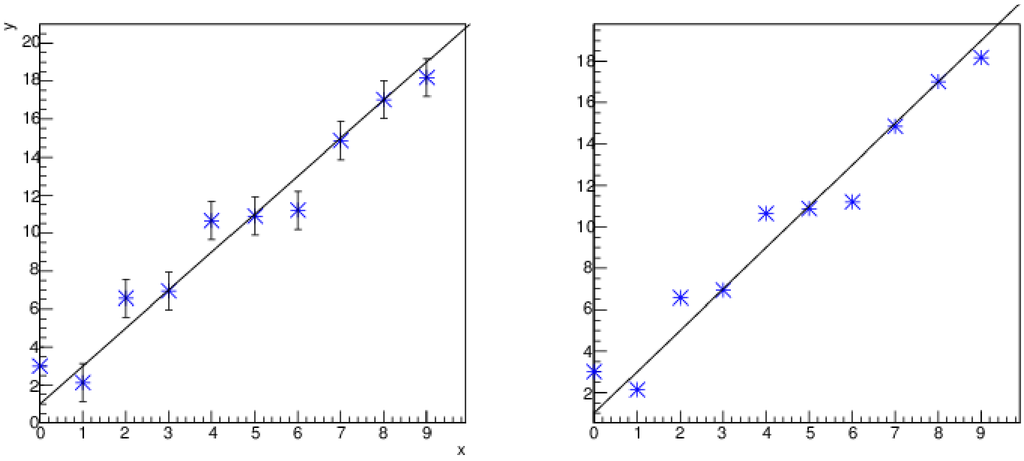


Fig. 14: A straight line fit (left) and linear regression (right)

It is also true that tall sons tend to have tall fathers—but not that tall—and only tallish grandfathers. Regress works in both directions!

Thus for regression there is always an ambiguity as to whether to plot x against y or y against x. For straight line fits as we usually meet them this does not arise: one variable is precisely specified and we call that one x, and the one with measurement errors is y.

5.6 Fitting histograms

When fitting a histogram the error is given by Poisson statistics for the number of events in each bin.

There are 4 methods of approaching this problem—in order of increasing accuracy and decreasing speed. It is assumed that the bin width W is narrow, so that $f(x_i, a) = \int_{x_i}^{x_i+W} P(x, a)\, dx$ can be approximated by $f_i(x_i; a) = P(x_i; a) \times W$. W is almost always the same for all bins, but the rare cases of variable bin width can easily be included.

1. Minimise $\chi^2 = \sum_i \frac{(n_i - f_i)^2}{n_i}$. This is the simplest but clearly breaks if $n_i = 0$.
2. Minimise $\chi^2 = \sum_i \frac{(n_i - f_i)^2}{f_i}$. Minimising the Pearson χ^2 avoids the division-by-zero problem, though it assumes that the Poisson distribution can be approximated by a Gaussian.
3. Maximise $\ln L = \sum \ln(e^{-f_i} f_i^{n_i} / n_i!) \sim \sum n_i \ln f_i - f_i$. This method, known as *binned maximum*

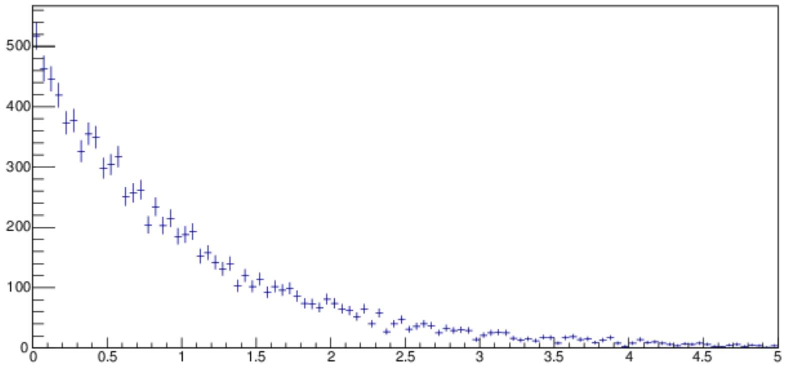


Fig. 15: Fitting a histogram

likelihood, remedies that assumption.

4. Ignore bins and maximise the total likelihood. Sums run over N_{events} not N_{bins}. So if you have large data samples this is much slower. You have to use it for sparse data, but of course in such cases the sample is small and the time penalty is irrelevant.

Which method to use is something you have to decide on a case by case basis. If you have bins with zero entries then the first method is ruled out (and removing such bins from the fit introduces bias so this should not be done). Otherwise, in my experience, the improvement in adopting a more complicated method tends to be small.

6 Errors

Estimation gives you a value for the parameter(s) that we have called a. But you also—presumably—want to know something about the uncertainty on that estimate. The maximum likelihood method provides this.

6.1 Errors from likelihood

For large N, the $\ln L(a, x)$ curve is a parabola, as shown in Fig. 16.

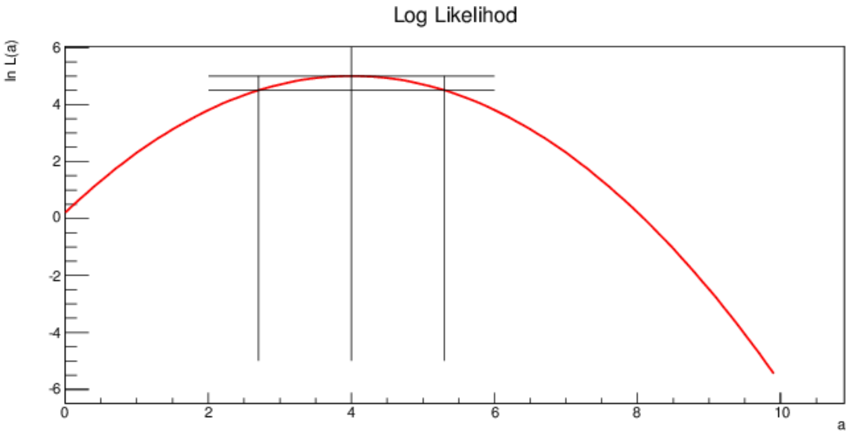


Fig. 16: Reading off the error from a Maximum Likelihood fit

At the maximum, a Taylor expansion gives $\ln L(a) = \ln L(\hat{a}) + \frac{1}{2}(a-\hat{a})^2 \frac{d^2 \ln L}{da^2} \ldots$
The maximum likelihood estimator saturates the MVB, so

$$V_{\hat{a}} = -1 / \left\langle \frac{d^2 \ln L}{da^2} \right\rangle \qquad \sigma_{\hat{a}} = \sqrt{-\frac{1}{\frac{d^2 \ln L}{da^2}}} \quad . \tag{17}$$

We approximate the expectation value $\left\langle \frac{d^2 \ln L}{da^2} \right\rangle$ by the actual value in this case $\left. \frac{d^2 \ln L}{da^2} \right|_{a=\hat{a}}$ (for a discussion of the introduced inaccuracy, see Ref. [12]).

This can be read off the curve, as also shown in Fig. 16. The maximum gives the estimate. You then draw a line $\frac{1}{2}$ below that (of course nowadays this is done within the code, not with pencil and ruler, but the visual image is still valid). This line $\ln L(a) = \ln L(\hat{a}) - \frac{1}{2}$ intersects the likelihood curve at the points $a = \hat{a} \pm \sigma_{\hat{a}}$. If you are working with χ^2, $L \propto e^{-\frac{1}{2}\chi^2}$ so the line is $\Delta\chi^2 = 1$.

This gives σ, or 68% errors. You can also take $\Delta \ln L = -2$ to get 2 sigma or 95% errors, or -4.5 for 3 sigma errors as desired. For large N these will all be consistent.

6.2 Combining errors

Having obtained—by whatever means—errors σ_x, σ_y... how does one combine them to get errors on derived quantities $f(x, y...), g(x, y, ...)$?

Suppose $f = Ax + By + C$, with A, B and C constant. Then it is easy to show that

$$\begin{aligned} V_f &= \left\langle (f - \langle f \rangle)^2 \right\rangle \\ &= \left\langle (Ax + By + C - \langle Ax + By + C \rangle)^2 \right\rangle \\ &= A^2(\left\langle x^2 \right\rangle - \langle x \rangle^2) + B^2(\left\langle y^2 \right\rangle - \langle y \rangle^2) + 2AB(\langle xy \rangle - \langle x \rangle \langle y \rangle) \\ &= A^2 V_x + B^2 V_y + 2AB\, \mathrm{Cov}_{xy} \quad . \end{aligned} \tag{18}$$

If f is not a simple linear function of x and y then one can use a first order Taylor expansion to approximate it about a central value $f_0(x_0, y_0)$

$$f(x, y) \approx f_0 + \left(\frac{\partial f}{\partial x}\right)(x - x_0) + \left(\frac{\partial f}{\partial y}\right)(y - y_0) \tag{19}$$

and application of Eq. 18 gives

$$V_f = \left(\frac{\partial f}{\partial x}\right)^2 V_x + \left(\frac{\partial f}{\partial y}\right)^2 V_y + 2\left(\frac{\partial f}{\partial x}\right)\left(\frac{\partial f}{\partial y}\right) \mathrm{Cov}_{xy} \tag{20}$$

writing the more familiar σ^2 instead of V this is equivalent to

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + 2\rho \left(\frac{\partial f}{\partial x}\right)\left(\frac{\partial f}{\partial y}\right) \sigma_x \sigma_y \quad . \tag{21}$$

If x and y are independent, which is often but not always the case, this reduces to what is often known as the ‘combination of errors’ formula

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 \quad . \tag{22}$$

Extension to more than two variables is trivial: an extra squared term is added for each and an extra covariance term for each of the variables (if any) with which it is correlated.

This can be expressed in language as *errors add in quadrature*. This is a friendly fact, as the result is smaller than you would get from arithmetic addition. If this puzzles you, it may be helpful to think of this as allowing for the possibility that a positive fluctuation in one variable may be cancelled by a negative fluctuation in the other.

There are a couple of special cases we need to consider. If f is a simple product, $f = Axy$, then Eq. 22 gives

$$\sigma_f^2 = (Ay)^2\sigma_x^2 + (Ax)^2\sigma_y^2 \, ,$$

which, dividing by f^2, can be written as

$$\left(\frac{\sigma_f}{f}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2 . \tag{23}$$

Furthermore this also applies if f is a simple quotient, $f = Ax/y$ or Ay/x or even $A/(xy)$.

This is very elegant, but it should not be overemphasised. Equation 23 is not fundamental: it only applies in certain cases (products or quotients). Equation 22 is the fundamental one, and Eq. 23 is just a special case of it.

A full error analysis has to include the treatment of the covariance terms—if only to show that they can be ignored. Why should the x and y in Eq. 20 be correlated? For direct measurements very often (but not always) they will not be. However the interpretation of results is generally a multistage process. From raw numbers of events one computes branching ratios (or cross sections...), from which one computes matrix elements (or particle masses...). Many quantities of interest to theorists are expressed as ratios of experimental numbers. And in this interpretation there is plenty of scope for correlations to creep into the analysis.

For example, an experiment might measure a cross section $\sigma(pp \to X)$ from a number of observed events N in the decay channel $X \to \mu^+\mu^-$. One would use a formula

$$\sigma = \frac{N}{B\eta\mathcal{L}} \, ,$$

where η is the efficiency for detecting and reconstructing an event, B is the branching ratio for $X \to \mu^+\mu^-$, and $\mathcal{L}$ is the integrated luminosity. These will all have errors, and the above prescription can be applied.

However it might also use the $X \to e^+e^-$ channel and then use

$$\sigma' = \frac{N'}{B'\eta'\mathcal{L}} \, .$$

Now σ and σ' are clearly correlated; even though N and N' are independent, the same $\mathcal{L}$ appears in both. If the estimate of $\mathcal{L}$ is on the high side, that will push both σ and σ' downwards, and vice versa.

On the other hand, if a second experiment did the same measurement it would have its own N, η and $\mathcal{L}$, but would be correlated with the first through using the same branching ratio (taken, presumably, from the Particle Data Group).

To calculate correlations between results we need the equivalent of Eq. 18

$$\begin{aligned} \mathrm{Cov}_{fg} &= \langle (f - \langle f \rangle)(g - \langle g \rangle) \rangle \\ &= \left(\frac{\partial f}{\partial x}\right)\left(\frac{\partial g}{\partial x}\right)\sigma_x^2 \quad , \end{aligned} \tag{24}$$

This can all be combined in the general formula which encapsulates all of the ones above

$$\mathbf{V_f} = \mathbf{G V_x \tilde{G}} \quad , \tag{25}$$

where $\mathbf{V_x}$ is the covariance matrix of the primary quantities (often, as pointed out earlier, this is diagonal), $\mathbf{V_f}$ is the covariance matrix of secondary quantities, and

$$G_{ij} = \frac{\partial f_i}{\partial x_j} \quad . \tag{26}$$

The $\mathbf{G}$ matrix is rectangular but need not be square. There may be more—or fewer—derived quantities than primary quantities. The matrix algebra of $\mathbf{G}$ and its transpose $\mathbf{\tilde{G}}$ ensures that the numbers of rows and columns match for Eq. 25.

As an example, consider a simple straight line fit, $y = mx + c$. Assuming that all the N y values are measured with the same error σ, least squares estimation gives the well known results

$$m = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} \qquad c = \frac{\overline{y}\,\overline{x^2} - \overline{xy}\,\overline{x}}{\overline{x^2} - \overline{x}^2} \quad . \tag{27}$$

For simplicity we write $D = 1/(\overline{x^2} - \overline{x}^2)$. The differentials are

$$\frac{\partial m}{\partial y_i} = \frac{D}{N}(x_i - \overline{x}) \qquad \frac{\partial c}{\partial y_i} = \frac{D}{N}(\overline{x^2} - x_i\overline{x}) \quad ,$$

from which, remembering that the y values are uncorrelated,

$$V_m = \sigma^2 \left(\frac{D}{N}\right)^2 \sum (x_i - \overline{x})^2 = \sigma^2 \frac{D}{N}$$
$$V_c = \sigma^2 \left(\frac{D}{N}\right)^2 \sum (\overline{x^2} - x_i\overline{x})^2 = \sigma^2 \overline{x^2} \frac{D}{N}$$
$$\mathrm{Cov}_{mc} = \sigma^2 \left(\frac{D}{N}\right)^2 \sum (x_i - \overline{x})(\overline{x^2} - x_i\overline{x}) = -\sigma^2 \overline{x} \frac{D}{N}$$

from which the correlation between m and c is just $\rho = -\overline{x}/\sqrt{\overline{x^2}}$.

This makes sense. Imagine you're fitting a straight line through a set of points with a range of positive x values (so $\overline{x}$ is positive). If the rightmost point happened to be a bit higher, that would push the slope m up and the intercept c down. Likewise if the leftmost point happened to be too high that would push the slope down and the intercept up. There is a negative correlation between the two fitted quantities.

Does it matter? Sometimes. Not if you're just interested in the slope—or the constant. But suppose you intend to use them to find the expected value of y at some extrapolated x. Equation 21 gives

$$y = mx + c \pm \sqrt{x^2\sigma_m^2 + \sigma_c^2 + 2x\rho\sigma_m\sigma_c}$$

and if, for a typical case where $\overline{x}$ is positive so ρ is negative, you leave out the correlation term you will overestimate your error.

This is an educational example because this correlation can be avoided. Shifting to a co-ordinate system in which $\overline{x}$ is zero ensures that the quantities are uncorrelated. This is equivalent to rewriting the well-known $y = mx + c$ formula as $y = m(x - \overline{x}) + c'$, where m is the same as before and $c' = c + m\overline{x}$. m and c' are now uncorrelated, and error calculations involving them become a lot simpler.

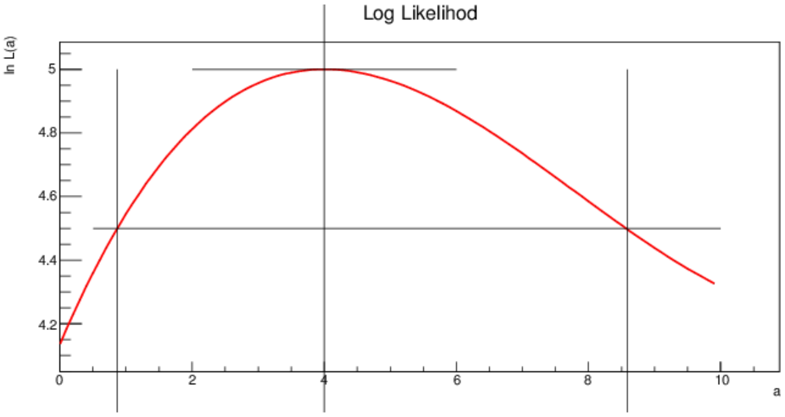


Fig. 17: An asymmetric likelihood curve

6.3 Asymmetric errors

So what happens if you plot the likelihood function and it is not symmetric like Fig. 16 but looks more like Fig. 17? This arises in many cases when numbers are small. For instance, in a simple Poisson count suppose you observe one event. $P(1;\lambda) = \lambda e^{-\lambda}$ is not symmetric: $\lambda = 1.5$ is more likely to fluctuate down to 1 than $\lambda = 0.5$ is to fluctuate up to 1.

You can read off σ_+ and σ_- from the two $\Delta \ln L = -\frac{1}{2}$ crossings, but they are different. The result can then be given as $a^{+\sigma_+}_{-\sigma_-}$. What happens after that?

The first advice is to avoid this if possible. If you get $\hat{a} = 4.56$ with $\sigma_+ = 1.61, \sigma_- = 1.59$ then quote this as 4.6 ± 1.6 rather than $4.56^{+1.61}_{-1.59}$. Those extra significant digits have no real meaning. If you can convince yourself that the difference between σ_+ and σ_- is small enough to be ignored then you should do so, as the alternative brings in a whole lot of trouble and it's not worth it.

But there will be some cases where the difference is too great to be swept away, so let's consider that case. There are two problems that arise: combination of measurements and combination of errors.

6.3.1 Combination of measurements with asymmetric errors

Suppose you have two measurements of the same parameter a: $\hat{a}_1{}^{+\sigma_1^+}_{-\sigma_1^-}$ and $\hat{a}_2{}^{+\sigma_2^+}_{-\sigma_2^-}$ and you want to combine them to give the best estimate and, of course, its error. For symmetric errors the answer is well established to be $\hat{a} = \frac{\hat{a}_1/\sigma_1^2 + \hat{a}_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}$.

If you know the likelihood functions, you can do it. The joint likelihood is just the sum. This is shown in Fig. 18 where the red and green curves are measurements of a. The log likelihood functions just add (blue), from which the peak is found and the $\Delta \ln L = -\frac{1}{2}$ errors read off.

But you don't know the full likelihood function: just 3 points (and that it had a maximum at the second). There are, of course, an infinite number of curves that could be drawn, and several models have been tried (cubics, constrained quartic...) on likely instances—see Ref. [13] for details. Some do better than others. The two most plausible are

$$\ln L = -\frac{1}{2}\left(\frac{a - \hat{a}}{\sigma + \sigma'(a - \hat{a})}\right)^2 \quad \text{and} \tag{28}$$

$$\ln L = -\frac{1}{2}\frac{(a - \hat{a})^2}{V + V'(a - \hat{a})} \quad . \tag{29}$$

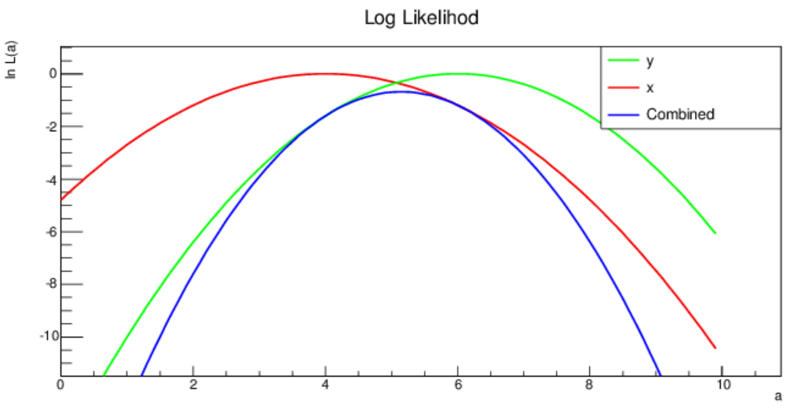


Fig. 18: Combination of two likelihood functions (red and green) to give the total (blue)

These are similar to the Gaussian parabola, but the denominator is not constant. It varies with the value of a, being linear either in the standard deviation or in the variance. Both are pretty good. The first does better with errors on $\log a$ (which are asymmetric if a is symmetric: such asymmetric error bars are often seen on plots where the y axis is logarithmic), the second does better with Poisson measurements.

From the 3 numbers given one readily obtains

$$\sigma = \frac{2\sigma^+\sigma^-}{\sigma^+ + \sigma^-} \qquad \sigma' = \frac{\sigma^+ - \sigma^-}{\sigma^+ + \sigma^-} \tag{30}$$

or, if preferred

$$V = \sigma^+\sigma^- \qquad V' = \sigma^+ - \sigma^- \quad . \tag{31}$$

From the total likelihood you then find the maximum of sum, numerically, and the $\Delta \ln L = -\frac{1}{2}$ points.

Code for doing this is available on GitHub[3] in both R and Root.

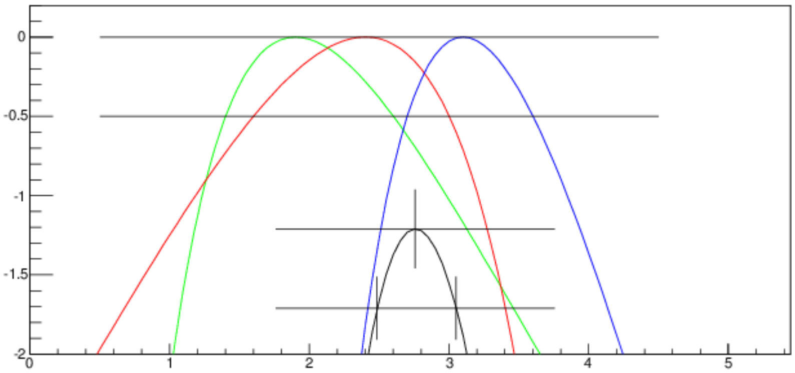


Fig. 19: Combining three asymmetric measurements

An example is shown in Fig. 19. Combining $1.9^{+0.7}_{-0.5}$, $2.4^{+0.6}_{-0.8}$ and $3.1^{+0.5}_{-0.4}$ gives $2.76^{+0.29}_{-0.27}$.

[3] https://github.com/RogerJBarlow/Asymmetric-Errors

6.3.2 Combination of errors for asymmetric errors

For symmetric errors, given $x \pm \sigma_x, y \pm \sigma_y$, (and $\rho_{xy} = 0$) the error on $f(x, y)$ is the sum in quadrature: $\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2$. What is the equivalent for the error on $f(x, y)$ when the errors are asymmetric, $x^{+\sigma_x^+}_{-\sigma_x^-}, y^{+\sigma_y^+}_{-\sigma_y^-}$? Such a problem arises frequently at the end of an analysis when the systematic errors from various sources are all combined.

The standard procedure—which you will see done, though it has not, to my knowledge, been written down anywhere—is to add the positive and negative errors in quadrature separately: ${\sigma_f^+}^2 = {\sigma_x^+}^2 + {\sigma_y^+}^2, \quad {\sigma_f^-}^2 = {\sigma_x^-}^2 + {\sigma_y^-}^2$. This looks plausible, but it is *manifestly wrong* as it breaks the central limit theorem.

To see this, suppose you have to average N i.i.d. variables each with the same errors which are asymmetric: $\sigma^+ = 2\sigma^-$. The standard procedure reduces both σ^+ and σ^- by a factor $1/\sqrt{N}$, but the skewness remains. The positive error is twice the negative error. This is therefore not Gaussian, and never will be, even as $N \to \infty$.

You can see what's happening by considering the combination of two of these measurements. They both may fluctuate upwards, or they both may fluctuate downwards, and yes, the upward fluctuation will be, on average, twice as big. But there is a 50% chance of one upward and one downward fluctuation, which is not considered in the standard procedure.

For simplicity we write $z_i = \frac{\partial f}{\partial x_i}(x_i - x_i^0)$, the deviation of the parameter from its nominal value, scaled by the differential. The individual likelihoods are again parametrized as Gaussian with a linear dependence of the standard deviation or of the variance, giving

$$\ln L(\vec{z}) = -\frac{1}{2}\sum_i \left(\frac{z_i}{\sigma_i + \sigma_i' z_i}\right)^2 \qquad \text{or} \qquad -\frac{1}{2}\sum_i \frac{z_i^2}{V_i + V_i' z_i} \quad , \tag{32}$$

where σ, σ', V, V' are obtained from Eqs. 30 or 31.

The z_i are nuisance parameters (as described later) and can be removed by profiling. Let $u = \sum z_i$ be the total deviation in the quoted f arising from the individual deviations. We form $\hat{L}(u)$ as the maximum of $L(\vec{z})$ subject to the constraint $\sum_i z_i = u$. The method of undetermined multipliers readily gives the solution

$$z_i = u\frac{w_i}{\sum_j w_j} \quad , \tag{33}$$

where

$$w_i = \frac{(\sigma_i + \sigma_i' z_i)^3}{2\sigma_i} \qquad \text{or} \qquad \frac{(V_i + V_i' z_i)^2}{2V_i + V_i' z_i} \quad . \tag{34}$$

The equations are nonlinear, but can be solved iteratively. At $u = 0$ all the z_i are zero. Increasing (or decreasing) u in small steps, Eqs. 33 and 34 are applied successively to give the z_i and the w_i: convergence is rapid. The value of u which maximises the likelihood should in principle be applied as a correction to the quoted result.

Programs to do this are also available on the GitHub site.

As an example, consider a counting experiment with a number of backgrounds, each determined by an ancillary Poisson experiment, and that for simplicity each background was determined by running the apparatus for the same time as the actual experiment. (In practice this is unlikely, but scale factors can easily be added.)

Suppose two backgrounds are measured, one giving four events and the other five. These would be reported, using $\Delta lnL = -\frac{1}{2}$ errors, as $4^{+2.346}_{-1.682}$ and $5^{+2.581}_{-1.916}$. The method, using linear V, gives the combined error on the background count as ${}^{+3.333}_{-2.668}$.

In this simple case we can check the result against the total background count of nine events, which has errors $^{+3.342}_{-2.676}$. The agreement is impressive. Further examples of the same total, partitioned differently, are shown in table 1.

Table 1: Various combinations of Poisson errors. The target value is $\sigma^- = 2.676$, $\sigma^+ = 3.342$

Inputs	Linear σ		Linear V	
	σ^-	σ^+	σ^-	σ^+
4+5	2.653	3.310	2.668	3.333
3+6	2.653	3.310	2.668	3.333
2+7	2.653	3.310	2.668	3.333
2+7	2.653	3.310	2.668	3.333
3+3+3	2.630	3.278	2.659	3.323
1+1+1+1+1+1+1+1+1	2.500	3.098	2.610	3.270

6.4 Errors in 2 or more dimensions

For 2 (or more) dimensions, one plots the log likelihood and defines regions using contours in $\Delta \ln L$ (or $\Delta\chi^2 \equiv -2\Delta \ln L$). An example is given in Fig. 20.

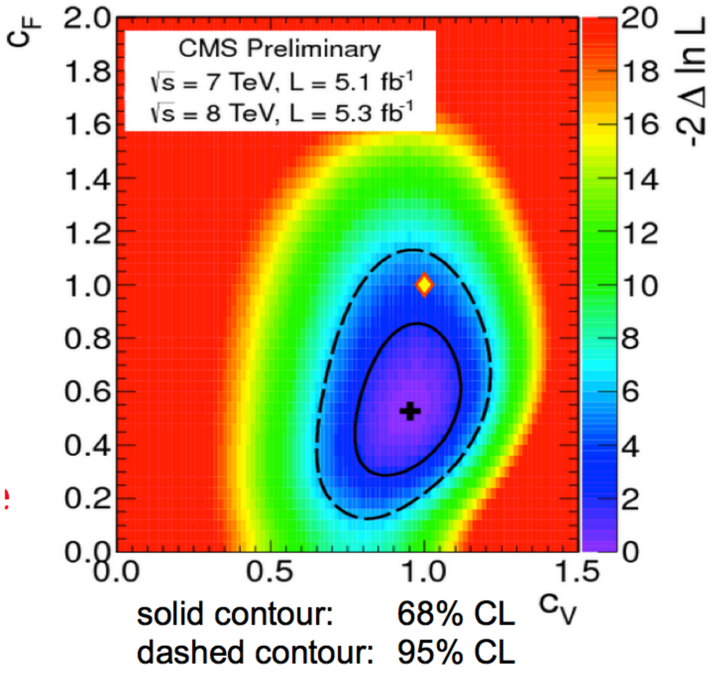


Fig. 20: CMS results on C_V and C_F, taken from Ref. [14]

The link between the $\Delta \ln L$ values and the significance changes. In 1D, there is a 68% probability of a measurement falling within 1 σ. In 2D, a 1σ square would give a probability $0.68^2 = 47\%$. If one rounds off the corners and draws a 1σ contour at $\Delta \ln L = -\frac{1}{2}$ this falls to 39%. To retrieve the full 68% one has to draw a contour at $\Delta \ln L = -1.14$, or equivalently $\Delta\chi^2 = 2.27$. For 95% use $\Delta\chi^2 = 5.99$ or $\Delta \ln L = -3.00$.

The necessary value is obtained from the χ^2 distribution—described later. It can be found by the R function `qchisq(p,n)` or the Root function `TMath::ChiSquareQuantile(p,n)`, where the desired probability `p` and number of degrees of freedom `n` are the arguments given.

6.4.1 Nuisance parameters

In the example of Fig. 20, both C_V and C_F are interesting. But in many cases one is interested only in one (or some) of the quantities and the others are 'nuisance parameters' that one would like to remove, reducing the dimensionality of the quoted result. There are two methods of doing this, one (basically) Frequentist and one Bayesian.

The Frequentist uses the *profile likelihood* technique. Suppose that there are two parameters, a_1 and a_2, where a_2 is a nuisance parameter, and so one wants to reduce the joint likelihood function $L(x; a_1, a_2)$ to some function $\hat{L}(a_1)$. To do this one scans across the values of a_1 and inserts $\hat{\hat{a}}_2(a_1)$, the value of a_2 which maximises the likelihood for that particular a_1

$$\hat{L}(x, a_1) = L(a_1, \hat{\hat{a}}_2(a_1)) \tag{35}$$

and the location of the maximum and the $\Delta \ln L = \frac{1}{2}$ errors are read off as usual.

To see why this works—though this is not a very rigorous motivation—suppose one had a likelihood function as shown in Fig. 21.

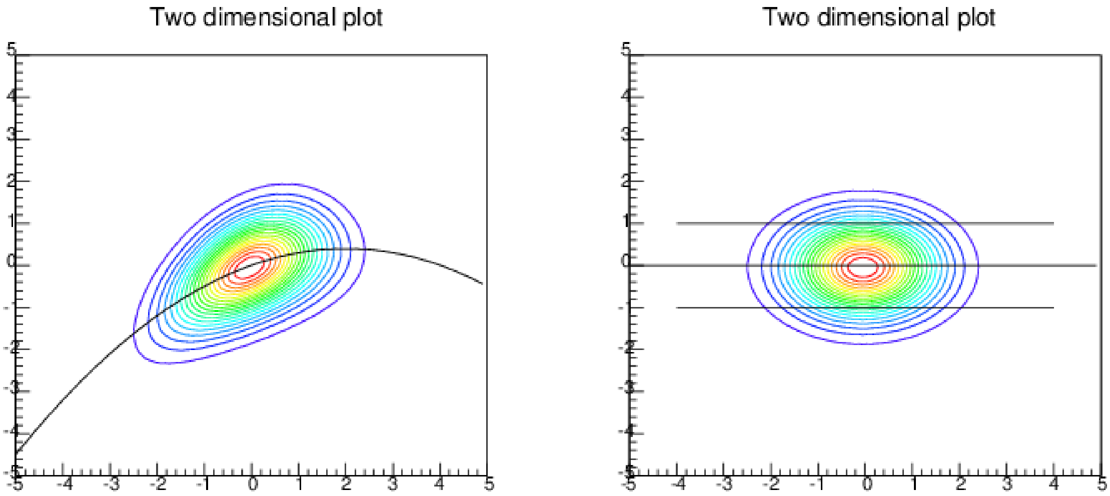


Fig. 21: Justification of the likelihood profile method

The horizontal axis is for the parameter of interest, a_1, and the vertical for the nuisance parameter a_2.

Different values of a_2 give different results (central and errors) for a_1.

If it is possible to transform to $a_2'(a_1, a_2)$ so that L factorises, then we can write $L(a_1, a_2') = L_1(a_1)L_2(a_2')$: this is shown in the plot on the right. We suppose that this is indeed possible. In the case here, and other not-too-complicated cases, it clearly is, although it will not be so in more complicated topologies with multiple peaks.

Then using the transformed graph, whatever the value of a_2', one would get the same result for a_1. Then one can present this result for a_1, independent of anything about a_2'.

There is no need to factorise explicitly: the path of central a_2' value as a function of a_1 (the central of the 3 lines on the right hand plot) is the path of the peak, and that path can be located in the first plot (the transformation only stretches the a_2 axis, it does not change the heights).

The Bayesian method uses the technique called *marginalisation*, which just integrates over a_2. Frequentists cannot do this as they are not allowed to integrate likelihoods over the parameter: $\int P(x; a)\,dx$ is fine, but $\int P(x; a)\,da$ is off limits. Nevertheless this can be a very helpful alternative to profiling, specially for many nuisance parameters. But if you use it you must be aware that this is strictly

Bayesian. Reparametrizing a_2 (or choosing a different prior) will give different results for a_1. In many cases, where the effect of the nuisance parameter is small, this does not have a big effect on the result.

6.5 Systematic errors

This can be a touchy subject. There is a lot of bad practice out there. Muddled thinking and following traditional procedures without understanding. When statistical errors dominated, this didn't matter much. In the days of particle factories and big data samples, it does.

6.5.1 What is a systematic error?

Consider these two quotations, from eminent and widely-read authorities.

R. Bevington defines

> 'Systematic error: reproducible inaccuracy introduced by faulty equipment, calibration, or technique.' [15],

whereas J. Orear writes

> 'Systematic effects is a general category which includes effects such as background, scanning efficiency, energy resolution, variation of counter efficiency with beam position, and energy, dead time, etc. The uncertainty in the estimation of such a systematic effect is called a systematic error.' [16].

Read these carefully and you will see that they are contradictory. They are not talking about the same thing. Furthermore, Orear is RIGHT and Bevington is WRONG—as are a lot of other books and websites.

We teach undergraduates the difference between measurement *errors*, which are part of doing science, and *mistakes*. They are not the same. If you measure a potential of 12.3 V as 12.4 V, with a voltmeter accurate to 0.1V, that is fine. Even if you measure 12.5 V. If you measure it as 124 V, that is a mistake.

In the quotes above, Bevington is describing *systematic mistakes* (the word 'faulty' is the key) whereas Orear is describing *systematic uncertainties*—which are 'errors' in the way we use the term.

There is a case for saying one should avoid the term 'systematic error' and always use 'uncertainty' or 'mistake'. This is probably impossible. But you should always know which you mean.

Restricting ourselves to uncertainties (we will come back to mistakes later) here are some typical examples:

- Track momenta from $p_i = 0.3B\rho_i$ have statistical errors from ρ and systematic errors from B,
- Calorimeter energies from $E_i = \alpha D_i + \beta$ have statistical errors from the digitised light signal D_i and systematic errors from the calibration α, β, and
- Branching ratios from $Br = \frac{N_D - B}{\eta N_T}$ have statistical errors from N_D and systematic errors from efficiency η, background B, total N_T .

Systematic uncertainties can be either Bayesian or Frequentist. There are clearly frequentist cases where errors have been determined by an *ancillary experiment* (real or simulated), such as magnetic field measurements, calorimeter calibration in a testbeam, and efficiencies from Monte Carlo simulations. (Sometimes the ancillary experiment is also the main experiment—e.g. in estimating background from sidebands.) There are also uncertainties that can only be Bayesian, e.g. when a theorist tells you that their calculation is good to 5% (or whatever) or an experimentalist affirms that the calibration will not have shifted during the run by more than 2% (or whatever).

6.5.2 How to handle them: correlations

Working with systematic errors is actually quite straightforward. They obey the same rules as statistical uncertainties.

We write $x = 12.2 \pm 0.3 \pm 0.4$ 'where the first error is statistical and the second is systematic', but it would be valid to write $x = 12.2 \pm 0.5$. For single measurement the extra information given by the two separate numbers is small. (In this case it just tells you that there is little to be gained by increasing the size of the data sample). For multiple measurements e.g. $x_a = 12.2 \pm 0.3, x_b = 17.1 \pm 0.4, all \pm 0.5$ the extra information is important, as results are correlated. Such cases arise, for example, in cross section measurements with a common luminosity error, or branching ratios with common efficiency.

Such a correlation means that taking more measurements and averaging does not reduce the error. Also there is no way to estimate σ_{sys} from the data—hence no check on the goodness of fit from a χ^2 test.

6.5.3 Handling systematic errors in your analysis

It is useful to consider systematic errors as having three types:

1. Uncertainty in an explicit continuous parameter. For example an uncertainty in efficiency, background and luminosity in determining a branching ratio or cross section. For these the standard combination of errors formula and algebra are usable, just like undergraduate labs.
2. Uncertainty in an implicit continuous parameter. For example: MC tuning parameters (σ_{p_T}, polarisation . . .). These are not amenable to algebra. Instead one calculates the result for different parameter values, typically at $\pm\sigma$, and observes the variation in the result, as illustrated in Fig. 22.

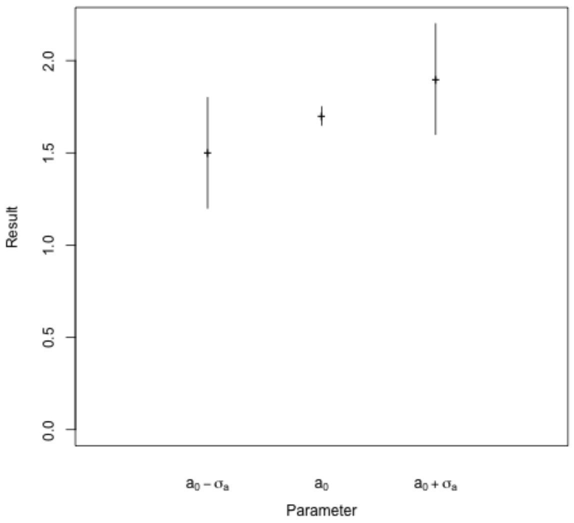


Fig. 22: Evaluating the effect of an implicit systematic uncertainty

Hopefully the effect is equal but opposite—if not then one can reluctantly quote an asymmetric error. Also your analysis results will have errors due to finite MC statistics. Some people add these in quadrature. This is wrong. The technically correct thing to do is to subtract them in quadrature, but this is not advised.

3. Discrete uncertainties:
   These typically occur in model choices. Using a different Monte Carlo for background—or signal—gives you a (slightly) different result. How do you include this uncertainty?

The situation depends on the status of the models. Sometimes one is preferred, sometimes they are all equal (more or less).
With 1 preferred model and one other, quote $R_1 \pm |R_1 - R_2|$.
With 2 models of equal status, quote $\frac{R_1+R_2}{2} \pm |\frac{R_1-R_2}{\sqrt{2}}|$.
With N models: take $\overline{R} \pm \sqrt{\frac{N}{N-1}(\overline{R}^2 - \overline{R}^2)}$ or similar mean value.
2 extreme models: take $\frac{R_1+R_2}{2} \pm \frac{|R_1-R_2|}{\sqrt{12}}$.
These are just ballpark estimates. Do not push them too hard. If the difference is not small, you have a problem—which can be an opportunity to study model differences.

6.5.4 Checking the analysis

"As we know, there are known knowns. There are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know."

Donald H. Rumsfeld

Errors are not mistakes—but mistakes still happen. Statistical tools can help find them. Check your result by repeating the analysis with changes which *should* make no difference:

- Data subsets,
- Magnet up/down,
- Different selection cuts,
- Changing histogram bin size and fit ranges,
- Changing parametrization (including order of polynomial),
- Changing fit technique,
- Looking for impossibilities,
- . . .

The more tests the better. You cannot prove the analysis is correct. But the more tests it survives the more likely your colleagues[4] will be to believe the result.

For example: in the paper reporting the first measurement of CP violation in B mesons the BaBar Collaboration [17] reported

> '. . . consistency checks, including separation of the decay by decay mode, tagging category and B_{tag} flavour . . . We also fit the samples of non-CP decay modes for $\sin 2\beta$ with no statistically significant difference found.'

If your analysis passes a test then *tick the box and move on*. Do not add the discrepancy to the systematic error. Many people do—and your supervisor and your review committee may want you to do so. Do not give in.

- It's illogical,
- It penalises diligence, and
- Errors get inflated.

If your analysis fails a test then worry!

[4] and eventually even you

– Check the test. Very often this turns out to be faulty.
– Check the analysis. Find mistake, enjoy improvement.
– Worry. Consider whether the effect might be real. (E.g. June's results are different from July's. Temperature effect? If so can (i) compensate and (ii) introduce implicit systematic uncertainty).
– Worry harder. Ask colleagues, look at other experiments.

Only as a last resort, add the term to the systematic error. Remember that this could be a hint of something much bigger and nastier.

6.5.5 Clearing up a possible confusion

What's the difference between?

Evaluating implicit systematic errors: vary lots of parameters, see what happens to the result, and include in systematic error.

Checks: vary lots of parameters, see what happens to the result, and don't include in systematic error.

If you find yourself in such a situation there are actually two ways to tell the difference.

(1) Are you expecting to see an effect? If so, it's an evaluation, if not, it's a check.

(2) Do you clearly know how much to vary them by? If so, it's an evaluation. If not, it's a check.

These cover even complicated cases such as a trigger energy cut where the energy calibration is uncertain—and it may be simpler to simulate the effect by varying the cut rather than the calibration.

6.5.6 So finally:

1. Thou shalt never say 'systematic error' when thou meanest 'systematic effect' or 'systematic mistake'.
2. Thou shalt know at all times whether what thou performest is a check for a mistake or an evaluation of an uncertainty.
3. Thou shalt not incorporate successful check results into thy total systematic error and make thereby a shield to hide thy dodgy result.
4. Thou shalt not incorporate failed check results unless thou art truly at thy wits' end.
5. Thou shalt not add uncertainties on uncertainties in quadrature. If they are larger than chickenfeed thou shalt generate more Monte Carlo until they shrink.
6. Thou shalt say what thou doest, and thou shalt be able to justify it out of thine own mouth; not the mouth of thy supervisor, nor thy colleague who did the analysis last time, nor thy local statistics guru, nor thy mate down the pub.

Do these, and thou shalt flourish, and thine analysis likewise.

7 Goodness of fit

You have the best fit model to your data—but is it good enough? The upper plot in Fig. 23 shows the best straight line through a set of points which are clearly not well described by a straight line. How can one quantify this?

You construct some measure of agreement—call it t—between the model and the data. Convention: $t \geq 0$, $t = 0$ is perfect agreement. Worse agreement implies larger t. The null hypothesis H_0 is that the model did indeed produce this data. You calculate the $p-$value: the probability under H_0 of getting a t this bad, or worse. This is shown schematically in the lower plot. Usually this can be done using known algebra—if not one can use simulation (a so-called 'Toy Monte Carlo').

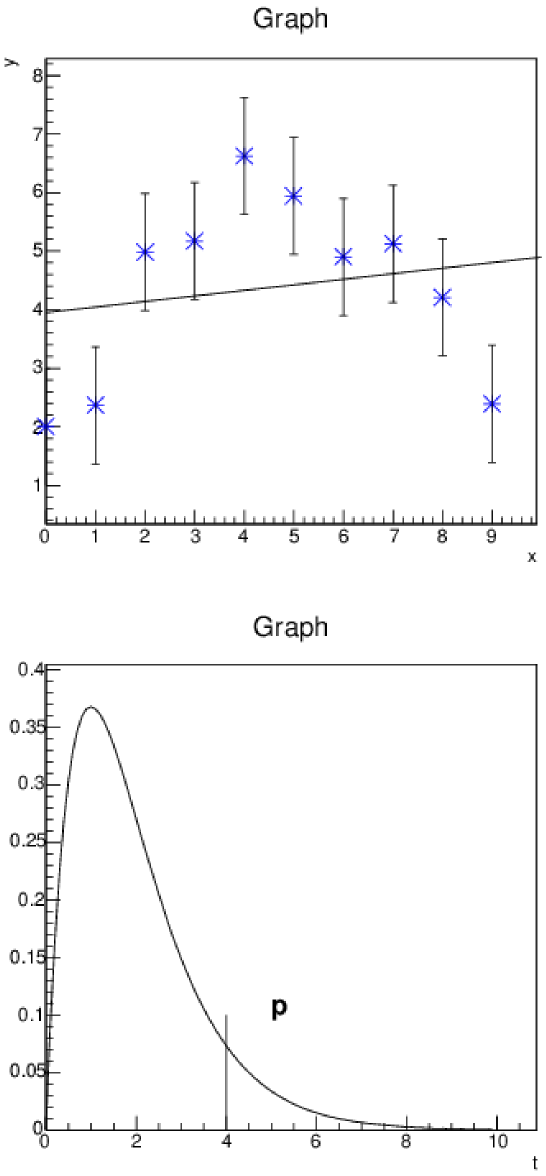


Fig. 23: The best fit to the data may not be good enough

7.1 The χ^2 distribution

The overwhelmingly most used such measure of agreement is the quantity χ^2

$$\chi^2 = \sum_1^N \left(\frac{y_i - f(x_i)}{\sigma_i}\right)^2 \quad . \tag{36}$$

In words: the total of the squared differences between prediction and data, scaled by the expected error. Obviously each term will be about 1, so $\left\langle \chi^2 \right\rangle \approx N$, and this turns out to be exact.

The distribution for χ^2 is given by

$$P(\chi^2; N) = \frac{1}{2^{N/2}\Gamma(N/2)} \chi^{N-2} e^{-\chi^2/2} \tag{37}$$

shown in Fig. 24, though this is in fact not much used: one is usually interested in the $p-$value, the probability (under the null hypothesis) of getting a value of χ^2 as large as, or larger than, the one observed. This can be found in ROOT with `TMath::Prob(chisquared,ndf)`, and in R from `1-pchisq(chisquared,ndf)`.

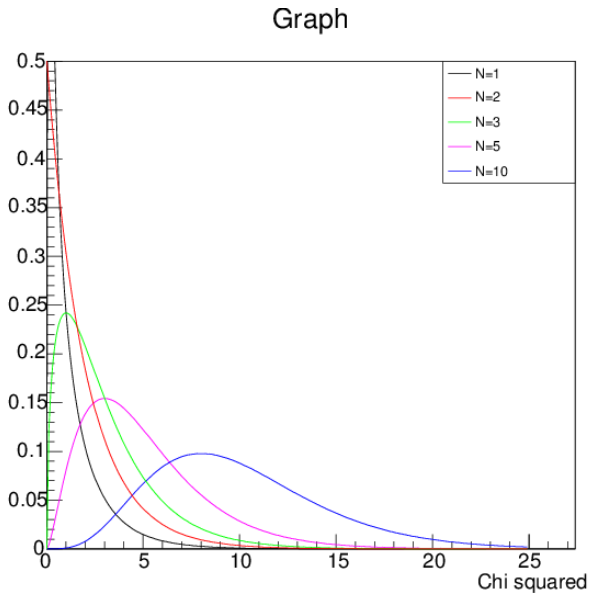


Fig. 24: The χ^2 distribution for various N

Thus for example with $N = 10, \chi^2 = 15$ then $p = 0.13$. This is probably OK. But for $N = 10, \chi^2 = 20$ then $p = 0.03$, which is probably not OK.

If the model has parameters which have been adjusted to fit the data, this clearly reduces χ^2. It is a very useful fact that the result also follows a χ^2 distribution for $NDF = N_{data} - N_{parameters}$ where NDF is called the 'number of degrees of freedom'.

If your χ^2 is suspiciously big, there are 4 possible reasons:

1. Your model is wrong,
2. Your data are wrong,
3. Your errors are too small, or
4. You are unlucky.

If your χ^2 is suspiciously small there are 2 possible reasons:

1. Your errors are too big, or
2. You are lucky.

7.2 Wilks' theorem

The Likelihood on its own tells you *nothing*. Even if you include all the constant factors normally omitted in maximisation. This may seem counter-intuitive, but it is inescapably true.

There is a theorem due to Wilks which is frequently invoked and appears to link likelihood and χ^2, but it does so only in very specific circumstances. Given two nested models, for large N the improvement in $\ln L$ is distributed like χ^2 in $-2\Delta \ln L$, with NDF the number of extra parameters.

So suppose you have some data with many (x, y) values and two models, Model 1 being linear and Model 2 quadratic. You maximise the likelihood using Model 1 and then using Model 2: the Likelihood increases as more parameters are available ($NDF = 1$). If this increase is significantly more than N

that justifies using Model 2 rather than Model 1. So it may tell you whether or not the extra term in a quadratic gives a meaningful improvement, but not whether the final quadratic (or linear) model is a good one.

Even this has an important exception. it does not apply if Model 2 contains a parameter which is meaningless under Model 1. This is a surprisingly common occurrence. Model 1 may be background, Model 2 background plus a Breit-Wigner with adjustable mass, width and normalization ($NDF = 3$). The mass and the width are meaningless under Model 1 so Wilks' theorem does not apply and the improvement in likelihood cannot be translated into a χ^2 for testing.

7.3 Toy Monte Carlos and likelihood for goodness of fit

Although the likelihood contains no information about the goodness of fit of the model, an obvious way to get such information is to run many simulations of the model, plot the spread of fitted likelihoods and use it to get the $p-$value.

This may be obvious, but it is wrong [18]. Consider a test case observing decay times where the model is a simple exponential $P(t) = \frac{1}{\tau}e^{-t/\tau}$, with τ an adjustable parameter. Then you get the Log Likelihood $\sum(-t_i/\tau - \ln\tau) = -N(\overline{t}/\tau + \ln\tau)$ and maximum likelihood gives $\hat{t} = \overline{t} = \frac{1}{N}\sum_i t_i$, so $\ln L(\hat{t}; x) = -N(1 + \ln\overline{t})$. This holds whatever the original sample $\{t_i\}$ looks like: any distribution with the same $\overline{t}$ has the same likelihood, after fitting.

8 Upper limits

Many analyses are 'searches for...' and most of these are unsuccessful. But you have to say something! Not just 'We looked, but we didn't see anything'. This is done using the construction of Frequentist confidence intervals and/or Bayesian credible intervals.

8.1 Frequentist confidence

Going back to the discussion of the basics, for Frequentists the probability that it will rain tomorrow is meaningless: there is only one tomorrow, it will either rain or it will not, there is no ensemble. The probability $N_{\text{rain}}/N_{\text{tomorrows}}$ is either 0 or 1. To talk about P_{rain} is "unscientific" [10].

This is unhelpful. But there is a workaround.

Suppose some forecast says it will rain and studies show this forecast is correct 90% of the time. We now have an ensemble of statements, and can say: 'The statement 'It will rain tomorrow' has a 90% probability of being true'. We shorten this to 'It will rain tomorrow, with 90% confidence'. We state X with confidence P if X is a member of an ensemble of statements of which at least P are true.

Note the 'at least' which has crept into the definition. There are two reasons for it:

1. Higher confidences embrace lower ones. If X at 95% then X at 90%, and
2. We can cater for composite hypotheses which are not completely defined.

The familiar quoted error is in fact a confidence statement. Consider as an illustration the Higgs mass measurement (current at the time of writing) $M_H = 125.09 \pm 0.24$ GeV. This does not mean that the probability of the Higgs mass being in the range $124.85 < M_H < 125.33$ GeV is 68%: the Higgs mass is a single, unique, number which either lies in this interval or it does not. What we are saying is that M_H has been measured to be 125.09 GeV with a technique that will give a value within 0.24 GeV of the true value 68% of the time. We say: $124.85 < M_H < 125.33\ GeV$ with 68% confidence. The statement is either true or false (time will tell), but it belongs to a collection of statements of which (at least) 68% are true.

So we construct *confidence regions* also known as confidence intervals $[x_-, x_+]$ such that $\int_{x_-}^{x_+} P(x)\,dx = CL$. We have not only a choice of the probability content (68%, 90%, 95%, 99%...) to work with but also of strategy. Common options are:

1. Symmetric: $\hat{x} - x_- = x_+ - \hat{x}$,
2. Shortest: Interval that minimises $x_+ - x_-$,
3. Central: $\int_{-\infty}^{x_-} P(x)\,dx = \int_{x_+}^{\infty} P(x)\,dx = \frac{1}{2}(1 - CL)$,
4. Upper Limit: $x_- = -\infty$, $\int_{x_+}^{\infty} P(x)\ ,dx = 1 - CL$, and
5. Lower Limit: $x_+ = \infty$, $\int_{-\infty}^{x_-} P(x)\ ,dx = 1 - CL$.

For the Gaussian (or any symmetric PDF) 1-3 are the same.

We are particularly concerned with the upper limit: we observe some small value x. We find a value x_+ such that for values of x_+ or more the probability of getting a result as small as x, or even less, is $1 - CL$, or even less.

8.2 Confidence belts

We have shown that a simple Gaussian measurement is basically a statement about confidence regions. $x = 100 \pm 10$ implies that [90,110] is the 68% central confidence region.

We want to extend this to less simple scenarios. As a first step, we consider a proportional Gaussian. Suppose we measure $x = 100$ from Gaussian measurement with $\sigma = 0.1x$ (a 10% measurement—which is realistic). If the true value is 90 the error is $\sigma = 9$ so $x = 100$ is more than one standard deviation, whereas if the true value is 110 then $\sigma = 11$ and it is less than one standard deviation. 90 and 110 are not equidistant from 100.

This is done with a technique called a confidence belt. The key point is that they are are constructed horizontally and read vertically, using the following procedure (as shown in Fig. 25). Suppose that a is the parameter of interest and x is the measurement.

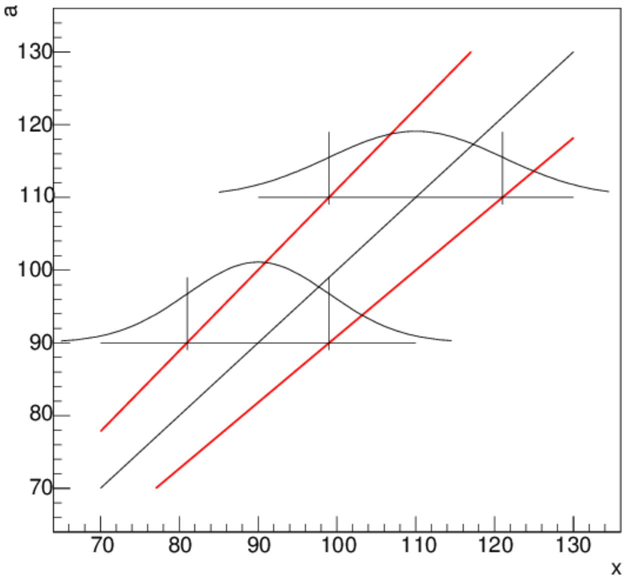


Fig. 25: A confidence belt for a proportional Gaussian

1. For each a, construct desired confidence interval (here 68% central).
2. The result (x, a) lies inside the belt (the red lines), with 68% confidence.
3. Measure x.
4. The result (x, a) lies inside the belt, with 68% confidence. And now we know x.
5. Read off the belt limits a_+ and a_- at that x: in this case they are 111.1, 90.9. So we can report that a lies in [90.9,111.1] with 68% confidence.
6. Other choices for the confidence level value and for the strategy are available.

This can be extended to the case of a Poisson distribution, Fig. 26.

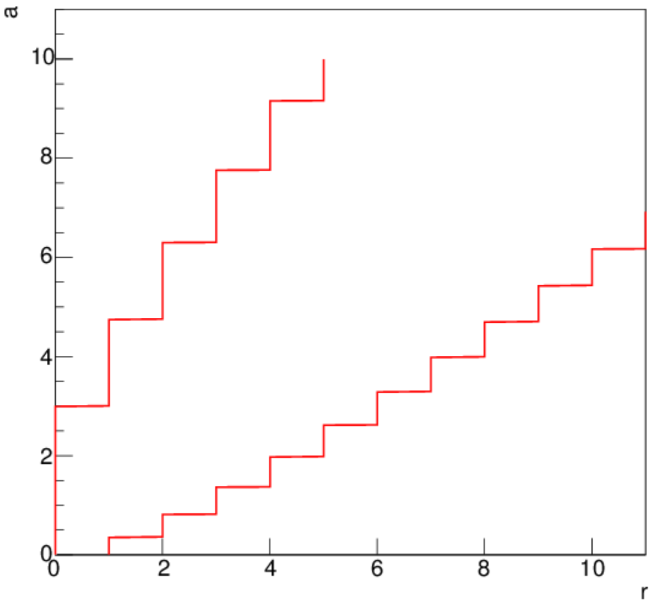


Fig. 26: A confidence belt for a Poisson

The only difference is that the horizontal axis is discrete as the number observed, x, is integer. In constructing the belt (horizontally) there will not in general be x values available to give $\sum_{x_-}^{x_+} = CL$ and we call, again, on the 'at least' in the definition and allow it to be $\sum_{x_-}^{x_+} \geq CL$.

Thus for a central 90% confidence we require for each a the largest integer x_{lo} and smallest x_{hi} for which $\sum_{x=0}^{x_{lo}-1} e^{-a} \frac{a^x}{x!} \leq 0.05$ and $\sum_{x=x_{hi}+1}^{\infty} e^{-a} \frac{a^x}{x!} \leq 0.05$. For the second sum it is easier to calculate $\sum_{x=0}^{x_{hi}} e^{-a} \frac{a^x}{x!} \geq 0.95$.

Whatever the value of a, the probability of the result falling in the belt is 90% or more. We proceed as for the Gaussian.

8.3 Coverage

This is an appropriate point to introduce *coverage*: the probability, given a, that the statement '$a_{lo} \leq a \leq a_{hi}$' will be true. Ideally this would be the same as the confidence level, however it may (because of the 'at least' clauses) exceed it ('overcover'); this is allowed though in principle inefficient. It should never be less ('undercover').

For example: suppose we have a Poisson process with $a = 3.5$ and we want a 90% central limit.

There is a probability $e^{-3.5} = 3\%$ of getting zero events, leading to $a_+ = 3.0$, which would be wrong as $3.0 < 3.5$.

Continuing in sequence, there is a probability $3.5e^{-3.5}$ =11% of getting one event, leading to $a_+ = 4.7$, which would be right.

Right answers continue up to seven events (with probability $3.5^7e^{-3.5}/7!$ =4%): this gives a safely large value for a_+ and $a_- = 3.3$, which is right as $3.3 < 3.5$, though only just, The next outcome, eight events (probability 2%) gives $a_- = 4.0$ which is wrong, as are all subsequent results.

Adding up the probabilities for the outcomes 1 thru 7 that give a true answer totals 94%, so there is 4% overcoverage.

Note that coverage is a function of the true value of the parameter on which limits are being placed. Values of a other than 3.5 will give different coverage numbers—though all are over 90%.

8.4 Upper limits

The one-sided upper limit—option 4 in the list above—gives us a way of quantifying the outcome of a null experiment. 'We saw nothing (or nothing that might not have been background), so we say $a \leq a_+$ at some confidence level'.

One simple and enlightening example occurs if you see no events, and there is no expected background. Now $P(0; 2.996) = 0.05$ and $2.996 \sim 3$. So if you see zero events, you can say with 95% confidence that the true value is less than 3.0. You can then directly use this to calculate a limit on the branching fraction, cross section, or whatever you're measuring.

8.5 Bayesian 'credible intervals'

A Bayesian has no problems saying 'It will probably rain tomorrow' or 'The probability that $124.85 < M_H < 125.33$ GeV is 68%'. The downside, of course, is that another Bayesian can say 'It will probably not rain tomorrow' and 'The probability that $124.85 < M_H < 125.33\ GeV$ is 86%' with equal validity and the two cannot resolve their subjective difference in any objective way.

A Bayesian has a prior belief PDF $P(a)$ and defines a region R such that $\int_R P(a)\, da = CL$. There is the same ambiguity regarding choice of content (68%, 90%, 95%...) and strategy (central, symmetric, upper limit...). So Bayesian credible intervals look a lot like Frequentist confidence intervals even if their meaning is different.

There are two happy coincidences.

The first is that Bayesian credible intervals on Gaussians, with a flat prior, are the same as Frequentist confidence intervals. If F quotes 68% or 95% or ... confidence intervals and B quotes 68% or 95% or ... credible interval, their results will agree.

The second is that although the Frequentist Poisson upper limit is given by $\sum_{r=0}^{r=r_{data}} e^{-a_{hi}} a_{hi}^r / r!$ whereas the Bayesian Poisson flat prior upper limit is given by $\int_0^{a_{hi}} e^{-a} a^{r_{data}} / r_{data}!\, da$, integration by parts of the Bayesian formula gives a series which is same as the Frequentist limit. A Bayesian will also say : 'I see zero events—the probability is 95% that the true value is 3.0 or less.' This is (I think) a coincidence—it does not apply for lower limits. But it does avoid heated discussions as to which value to publish.

8.6 Limits in the presence of background

This is where it gets tricky. Typically an experiment may observe N_D events, with an expected background N_B and efficiency η, and wants to present results for $N_S = \frac{N_D - N_B}{\eta}$. Uncertainties in η and N_B are handled by profiling or marginalising. The problem is that the *actual number* of background events is not N_B but Poisson in N_B.

So in a straightforward case, if you observe twelve events, with expected background 3.4 and $\eta = 1$ it is obviously sensible to say $N_S = 8.6$ (though the error is $\sqrt{12}$ not $\sqrt{8.6}$)

But suppose, with the same background, you see four events, three events or zero events. Can you say $N_S = 0.6$? Or -0.4? Or -3.4???

We will look at three methods of handling this, considering as an example the observation of three events with expected background 3.40, for which we want to present a 95% CL upper limit on N_S.

8.6.1 Method 1: Pure frequentist

$N_D - N_B$ is an unbiased estimator of N_S and its properties are known. Quote the result. Even if it is non-physical.

The argument for doing so is that this is needed for balance: if there is really no signal, approximately half of the experiments will give positive values and half negative. If the negative results are not published, but the positive ones are, the world average will be spuriously high. For a 95% confidence limit one accepts that 5% of the results can be wrong. This (unlikely) case is clearly one of them. So what?

A counter-argument is that if $N_D < N_B$, we *know* that the background has fluctuated downwards. But this cannot be incorporated into the formalism.

Anyway, the upper limit from 3 is 7.75, as $\sum_0^3 e^{-7.75} 7.75^r/r! = 0.05$, and the 95% upper limit on $N_S = 7.75 - 3.40 = 4.35$.

8.6.2 Method 2: Go Bayesian

Assign a uniform prior to N_S, for $N_S > 0$, zero for $N_S < 0$. The posterior is then just the likelihood, $P(N_S|N_D, N_B) = e^{-(N_S+N_B)} \frac{(N_S+N_B)^{N_D}}{N_D!}$. The required limit is obtained from integrating $\int_0^{N_{hi}} P(N_S)\, dN_S = 0.95$ where $P(N_S) \propto e^{-(N_s+3.40)} \frac{(N_s+3.4)^3}{3!}$; this is illustrated in Fig. 27 and the value of the limit is 5.21.

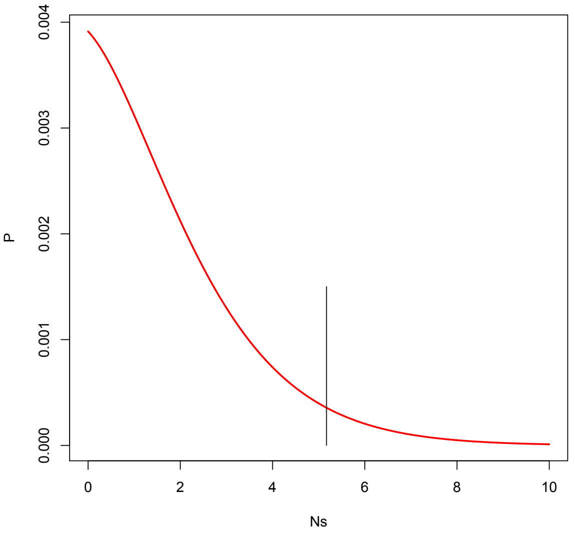


Fig. 27: The Bayesian limit construction

8.6.3 Method 3: Feldman-Cousins

This—called ‘the unified approach’ by Feldman and Cousins [19]—takes a step backwards and considers the ambiguity in the use of confidence belts.

In principle, if you decide to work at, say, 90% confidence you may choose to use a 90% central or a 90% upper limit, and in either case the probability of the result lying in the band is at least 90%. This is shown in Fig. 28.

In practice, if you happen to get a low result you would quote an upper limit, but if you get a high result you would quote a central limit. This, which they call 'flip-flopping', is illustrated in the plot by a break shown here for $r = 10$.

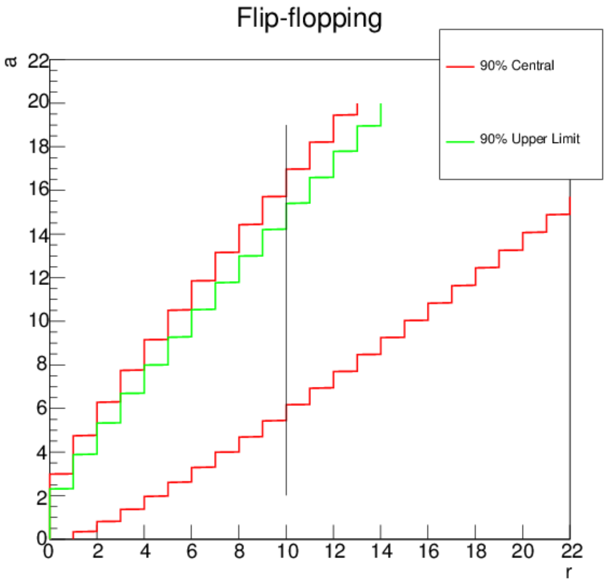


Fig. 28: The flip-flopping problem

Now the confidence belt is the green one for $r < 10$ and the red one for $r \geq 10$. The probability of lying in the band is no longer 90%! Flip-flopping invalidates the Frequentist construction, leading to undercoverage.

They show how to avoid this. You draw the plot slightly differently: $r \equiv N_D$ is still the horizontal variable, but as the vertical variable you use N_S. (This means a different plot for any different N_B, whereas the previous Poisson plot is universal, but this is not a problem.) This is to be filled using $P(r; N_s) = e^{-(N_s+N_B)} \frac{(N_S+N_B)^r}{r!}$.

For each N_S you define a region R such that $\sum_{r \epsilon R} P(r; N_s) \geq 90\%$. You have a choice of strategy that goes beyond 'central' or 'upper limit': one plausible suggestion would be to rank r by probability and take them in order until the desired total probability content is achieved (which would, incidentally, give the shortest interval). However this has the drawback that outcomes with $r < N_B$ will have small probabilities and be excluded for all N_S, so that, if such a result does occur, one cannot say anything constructive, just 'This was unlikely'.

An improved form of this suggestion is that for each N_S, considering each r you compare $P(r; N_S)$ with the largest possible value obtained by varying N_S. This is easier than it sounds because this highest value is either at $N_S = r - N_B$ (if $r \geq N_B$) or $N_S = 0$ (if $r \leq N_B$). Rank on the ratio $P(r; N_S)/P(r; N_S^{best})$ and again take them in order till their sum gives the desired probability.

This gives a band as shown in Fig. 29, which has $N_B = 3.4$. You can see that 'flip-flopping' occurs naturally: for small values of r one just has an upper limit, whereas for larger values, above $r = 7$, one obtains a lower limit as well. Yet there is a single band, and the coverage is correct (i.e. it does not undercover). In the case we are considering, $r = 3$, just an upper limit is given, at 4.86.

Like other good ideas, this has not found universal favour. Two arguments are raised against the

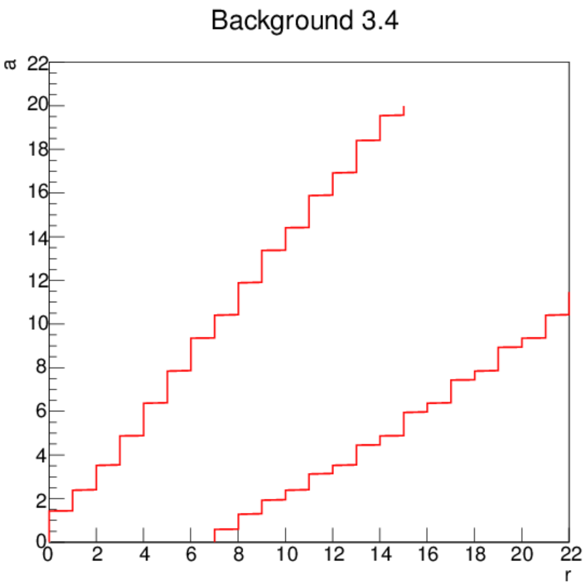


Fig. 29: A Feldman-Cousins confidence band

method.

First, that it deprives the physicist of the choice of whether to publish an upper limit or a range. It could be embarrassing if you look for something weird and are 'forced' to publish a non-zero result. But this is actually the point, and in such cases one can always explain that the limits should not be taken as implying that the quantity actually is nonzero.

Secondly, if two experiments with different N_B get the same small N_D, the one with the higher N_B will quote a smaller limit on N_S. The worse experiment gets the better result, which is clearly unfair! But this is not comparing like with like: for a 'bad' experiment with large background to get a small number of events is much less likely than it is for a 'good' low background experiment.

8.6.4 *Summary so far*

Given three observed events, and an expected background of 3.4 events, what is the 95% upper limit on the 'true' number of events? Possible answers are shown in table 2.

Table 2: Upper limits from different methods

Strict Frequentist	4.35
Bayesian (uniform prior)	5.21
Feldman-Cousins	4.86

Which is 'right'? Take your pick! All are correct. (Well, not wrong.). The golden rule is to say what you are doing, and if possible give the raw numbers.

8.6.5 *Extension: not just counting numbers*

These examples have used simple counting experiments. But a simple number does not (usually) exploit the full information.

Consider the illustration in Fig. 30. One is searching for (or putting an upper limit on) some broad resonance around 7 GeV. One could count the number of events inside some window (perhaps 6

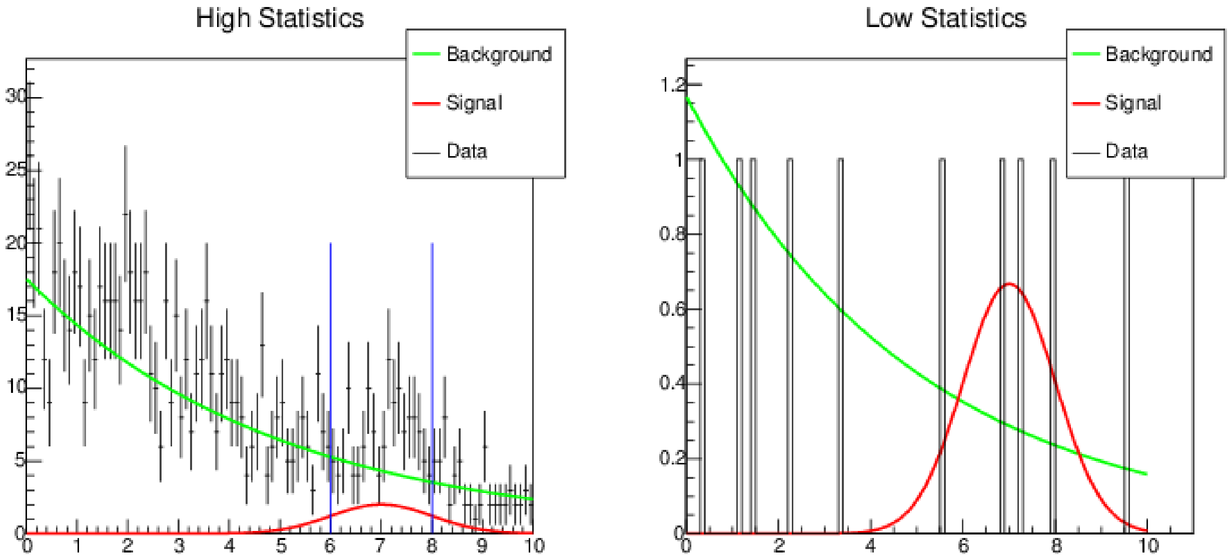


Fig. 30: Just counting numbers may not give the full information

to 8 GeV?) and subtract the estimated background. This might work with high statistics, as in the left, but would be pretty useless with small numbers, as in the right. It is clearly not optimal just to count an event as 'in', whether it is at 7.0 or 7.9, and to treat an event as 'out', if it is at 8.1 or 10.1.

It is better to calculate the Likelihood $\ln L_{s+b} = \sum_i \ln N_s S(x_i) + N_b B(x_i) \quad ; \quad \ln L_b = \sum_i \ln N_b B(x_i)$. Then, for example using CL_s, you can work with L_{s+b}/L_b, or $-2\ln\left(L_{s+b}/L_b\right)$. The confidence/probability quantities can be found from simulations, or sometimes from data.

8.6.6 Extension: CL_s

This is a modification of the standard frequentist approach to include the fact, as mentioned above, that a small observed signal implies a downward fluctuation in background [20]. It can also be understood as an extension of the Bayesian method described in Section 8.6.2 for simple counting. The integrals over Poisson likelihoods that occur in can be done by parts

$$\int e^{-a}\frac{a^r}{r!}\,da = \left[e^{-a}\frac{a^{r-1}}{(r-1)!}\right] + \int e^{-a}\frac{a^{r-1}}{(r-1)!}\,da \tag{38}$$

which, repeated, turns the integral into a series. The requirement for a 95% credible interval, $\int_0^{N_{hi}} P(N_S)\,dN_S = CL$, including the normalisation, becomes

$$1 - CL = e^{-N_{hi}} \frac{\sum_0^{N_D} \frac{(N_{hi}+B)^r}{r!}}{\sum_0^{N_D} \frac{B^r}{r!}} \tag{39}$$

This is known as Helène's formula [21]. Looking at it, it is the probability, with the signal strength at the upper limit N_{hi}, of getting the observed result of N_D or smaller (which is what the Frequentist recipe brings) divided by the probability of getting such a result from pure background.

CL_S generalises this prescription to apply to likelihoods rather than simple counting. Denote the (strict Frequentist) probability of getting a result this small (or less) from $s+b$ events as CL_{s+b}, and the equivalent probability from pure background as CL_b (so $CL_b = CL_{s+b}$ for $s=0$). Then introduce

$$CL_s = \frac{CL_{s+b}}{CL_b} \quad . \tag{40}$$

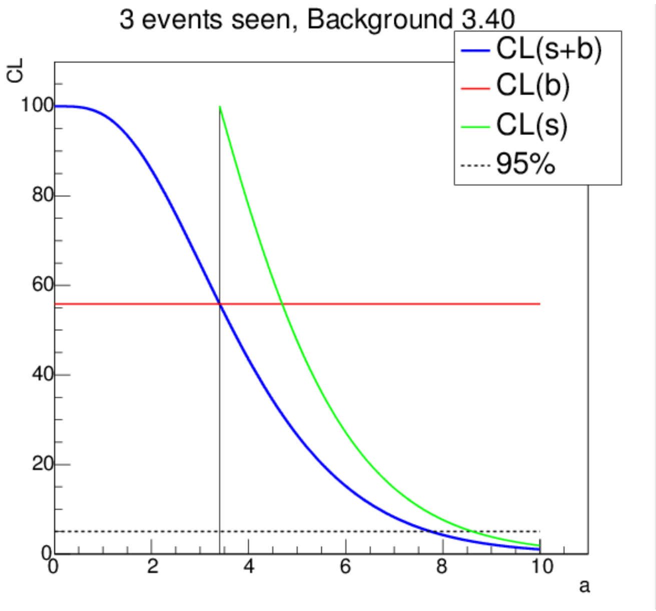


Fig. 31: The CL_s construction

Looking at Fig. 31, the CL_{s+b} curve shows that if $s+b$ is small then the probability of getting three events or less is high, near 100%. As $s+b$ increases this probability falls, and at $s+b=7.75$ the probability of only getting three events or less is only 5%. This, after subtraction of $b=3.4$, gives the strict Frequentist value. The point $s+b=3.4$ corresponds to $s=0$, at which the probability CL_b is 56% Dividing the (blue) CL_{s+b} curve by 0.56 gives the (green) CL_S curve, which has a maximum of 100% in the physically sensible region. This is treated in the same way as the CL_{s+b} curve, reading off the point at $s+b=8.61$ where it falls to 5%. This is a limit on $s+b$ so we subtract 3.4 to get the limit on s as 5.21(the same here as obtained by the Bayesian method 2: the only difference is in the way the integrals are done). This is larger than the strict Frequentist limit: the method over-covers (which, as we have seen, is allowed if not encouraged) and is, in this respect 'conservative'[5]. CL_s is not Frequentist, just 'Frequentist inspired'. In terms of statistics there is perhaps little in its favour. But it has an intuitive appeal, and is widely used.

8.6.7 Extension: From numbers to masses

Limits on numbers of events can readily be translated into limits on branching ratios, $BR = \frac{N_s}{N_{total}}$, or limits on cross sections, $\sigma = \frac{N_s}{\int \mathcal{L}dt}$.

These may translate to limits on other, theory, parameters.

In the Higgs search (to take an example) the cross section depends on the mass, M_H—and so does the detection efficiency—which may require changing strategy (hence different backgrounds). This leads to the need to basically repeat the analysis for all (of many) M_H values. This can be presented in two ways.

The first is shown in Fig. 32, taken from Ref. [22]. For each M_H (or whatever is being studied) you search for a signal and plot the CL_s (or whatever limit method you prefer) significance in a *Significance Plot*. Small values indicate that it is unlikely to get a signal this large just from background.

One often also plots the expected (from MC) significance, assuming the signal hypothesis is true. This is a measure of a 'good experiment'. In this case there is a discovery level drop at $M_H \approx 125$ GeV, which exceeds the expected significance, though not by much: ATLAS were lucky but not incredibly

[5] 'Conservative' is a misleading word. It is used by people describing their analyses to imply safety and caution, whereas it usually entails cowardice and sloppy thinking.

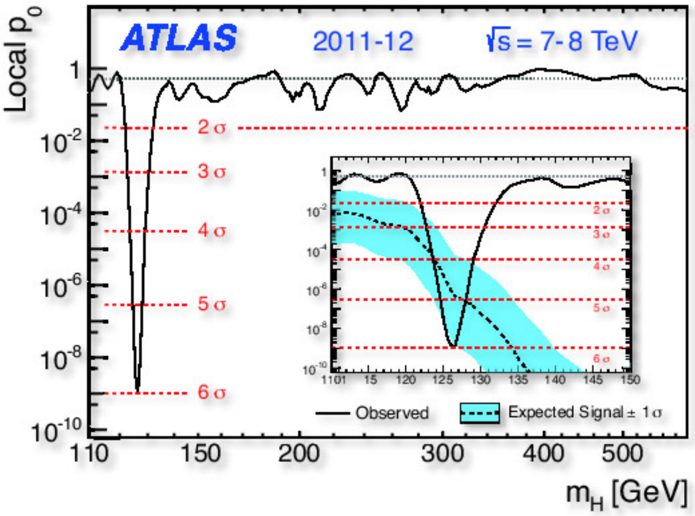


Fig. 32: Significance plot for the Higgs search

lucky.

The second method is—for some reason—known as the green-and-yellow plot. This is basically the same data, but fixing CL at a chosen value: in Fig. 33 it is 95%. You find the limit on signal strength, at this confidence level, and interpret it as a limit on the cross section σ/σ_{SM}. Again, as well as plotting the actual data one also plots the expected (from MC) limit, with variations. If there is no signal, 68% of experiments should give results in the green band, 95% in the yellow band.

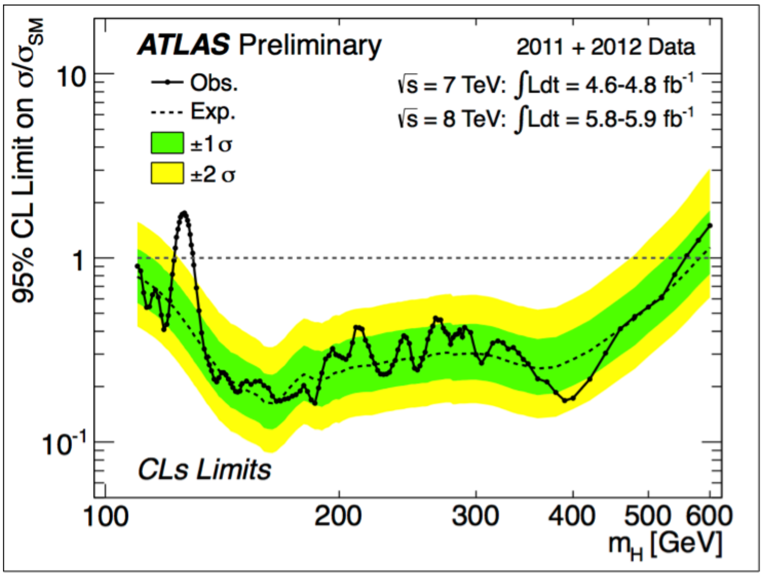


Fig. 33: Green and yellow plot showing the Higgs discovery

So this figure shows the experimental result as a black line. Around 125 GeV the 95% upper limit is more than the Standard Model prediction indicating a discovery. There are peaks between 200 and 300 GeV, but they do not approach the SM value, indicating that they are just fluctuations. The value rises at 600 GeV, but the green (and yellow) bands rise also, showing that the experiment is not sensitive for such high masses: basically it sees nothing but would expect to see nothing.

9 Making a discovery

We now turn from setting limits, to say what you did not see, to the more exciting prospect of making a discovery.

Remembering hypothesis testing, in claiming a discovery you have to show that your data can't be explained without it. This is quantified by the $p-$value: the probability of getting a result this extreme (or worse) under the null hypothesis/Standard Model. (This is *not* 'The probability that the Standard Model is correct', but it seems impossible for journalists to understand the difference.)

Some journals (particularly in psychology) refuse to publish papers giving $p-$values. If you do lots of studies, some will have low $p-$values (5% below 0.05 etc.). The danger is that these get published, but the unsuccessful ones are binned.

Is p like the significance α? Yes and no. The formula is the same, but α is a property of the test, computed before you see the data. p is a property of the data.

9.1 Sigma language

The probability ($p-$value) is often translated into Gaussian-like language: the probability of a result more than 3σ from the mean is 0.27% so a $p-$value of 0.0027 is a '3 σ effect' (or 0.0013 depending on whether one takes the 1-tailed or 2-tailed option. Both are used.) In reporting a result with a significance of 'so many σ' there is no actual σ involved: it is just a translation to give a better feel for the size of the probability.

By convention, 3 sigma, $p = 0.0013$ is reported as 'Evidence for' whereas a full 5 sigma
$p = 0.0000003$ is required for 'discovery of'.

9.2 The look-elsewhere effect

You may think that the requirement for 5 σ is excessively cautious. Its justification comes from history—too many 3- and 4- sigma 'signals' have gone away when more data was taken.

This is partly explained by the 'look-elsewhere effect'. How many peaks can you see in the data in Fig. 34?

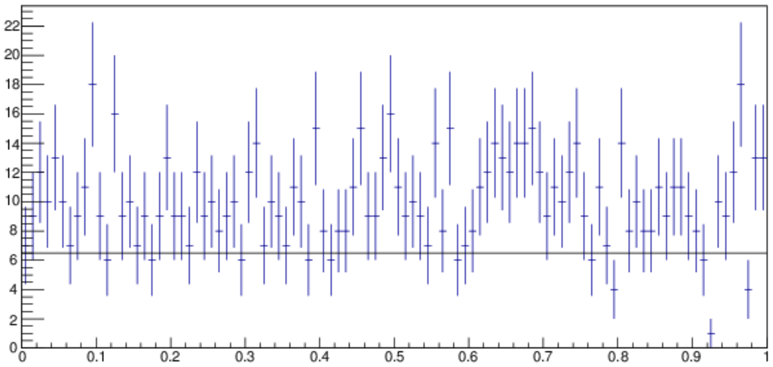


Fig. 34: How many peaks are in this data?

The answer is that there are none. The data is in fact purely random and flat. But the human eye is very good at seeing features.

With 100 bins, a $p-$value below 1% is pretty likely. This can be factored in, to some extent, using pseudo-experiments, but this does not allow for the sheer number of plots being produced by hard-working physicists looking for something. Hence the need for caution.

This is not just ancient history. ATLAS and CMS recently observed a signal in the $\gamma\gamma$ mass around 750 GeV, with a significance of 3.9σ (ATLAS) and 3.4σ (CMS), which went away when more data was taken.

9.3 Blind analysis

It is said[6] that when Michelangelo was asked how he created his masterpiece sculpture ‘David’ he replied ‘It was easy—all I did was get a block of marble and chip away everything that didn’t look like David’. Such creativity may be good for sculpture, but it’s bad for physics. If you take your data and devise cuts to remove all the events that don’t look like the signal you want to see, then whatever is left at the end will look like that signal.

Many/most analyses are now done ‘blind’. Cuts are devised using Monte Carlo and/or non-signal data. You only ‘open the box’ once the cuts are fixed. Most collaborations have a formal procedure for doing this.

This may seem a tedious imposition, but we have learnt the hard way that it avoids embarrassing mistakes.

10 Conclusions

Statistics is a tool for doing physics. Good physicists understand their tools. Don’t just follow without understanding, but read books and conference proceedings, go to seminars, talk to people, experiment with the data, and understand what you are doing. Then you will succeed. And you will have a great time!

References

[1] R. J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, (Wiley, Chichester, 1989).

[2] G. Cowan, *Statistical Data Analysis*, (Oxford Univ. Press, Oxford,1998).

[3] I. Narsky and F. C. Porter, *Statistical Analysis Techniques in Particle Physics*, (Wiley, Weinheim, 2014), doi:10.1002/9783527677320.

[4] O. Behnke *et al* (Eds.) *Data Analysis in High Energy Physics*, (Wiley, Weinheim, 2013), doi:10.1002/9783527653416.

[5] L. Lyons, *Statistics for Nuclear and Particle Physicists*, Cambridge Univ. Press, Cambridge, 1986).

[6] G. Bohm and G. Zech, *Introduction to Statistics and Data Analysis for Physicists*, 3rd revised ed. (Verlag Deutsches Elektronen-Synchrotron, Hamburg, 2017), doi:10.3204/PUBDB-2017-08987.

[7] M. R. Whalley and L. Lyons (Eds) *Advanced Statistical Techniques in Particle Physics*, Durham report IPPP/02/39, 2002, `https://inspirehep.net/literature/601052`.

[8] L. Lyons, R. Mount and R. Reitmayer (Eds.) *Proceedings of PHYSTAT03*, SLAC-R-703, 2003, `https://www.slac.stanford.edu/econf/C030908/`.

[9] L. Lyons and M. K. Unel (Eds.) *Proceedings of PHYSTAT05*, (Imperial College Press, London, 2006), doi:10.1142/p446.

[10] R. von Mises, *Probability, Statistics and Truth*, reprint of the second revised 1957 English edition (Dover, Mineola, NY, 1981).

[11] H. Jeffreys, *Theory of Probability*, 3rd ed. (Oxford Univ. Press, Oxford, 1961).

[12] R. J. Barlow, *A note on* $\Delta \ln L = -\frac{1}{2}$ *Errors*, arXiv:physics/0403046, 2004.

[6] This story is certainly not historically accurate, but it’s still a good story (*quoteinvestigator.com*: `https://`

[13] R. J. Barlow, *Asymmetric Statistical Errors*, Proceedings of PHYSTAT05, Eds. L. Lyons and M. K. Unel (Imperial College Press, London, 2006), p.56, doi:10.1142/9781860948985_0013, arXiv:physics/0406120, 2004.

[14] CMS Collaboration, *CMS 2D Cf-Cv Likelihood Profile*, `https://root.cern.ch/cms-2d-cf-cv-likelihood-profile`, accessed 26 May 2019.

[15] P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, 3rd ed. (McGraw Hill, New York, NY, 2003).

[16] J. Orear, *Notes on Statistics for Physicists*, UCRL-8417, 1958, `http://nedwww.ipac.caltech.edu/level5/Sept01/Orear/frames.html`

[17] B. Aubert *et al.* [BaBar Collaboration], *Phys. Rev. Lett.* **86** (2001) 2515, doi:10.1103/PhysRevLett.86.2515, arXiv:hep-ex/0102030.

[18] J. G. Heinrich, CDF internal note CDF/MEMO/BOTTOM.CDFR/5639 (Many thanks to Jonas Rademacker for pointing this out); L. Lyons, R. Mount and R. Reitmayer (Eds.) *Proceedings of PHYSTAT03*, SLAC-R-703, 2003, p.52, `https://www.slac.stanford.edu/econf/C030908/`.

[19] G. J. Feldman and R. D. Cousins, *Phys. Rev.* **D57** (1998) 3873, doi:10.1103/PhysRevD.57.3873, arXiv:physics/9711021.

[20] A. L. Read, *J. Phys.* **G28** (2002), 2693, doi:10.1088/0954-3899/28/10/313.

[21] O. Helène *et al.*, *Nucl. Instr. Meth. Res. A* **212** (1983), 319, doi:10.1016/0167-5087(83)90709-3.

[22] ATLAS Collaboration, *ATLAS Higgs Search Update*, `https://atlas.cern/updates/atlas-news/atlas-higgs-search-update`, accessed 26 May 2019.

Proceedings of the 2019 European School of High-Energy Physics, St Petersburg, Russia, 14–27 September 2019, edited by C. Duhr and M. Mulders, CERN-2021-005 (CERN, Geneva, 2021)

LHC Run 2 and future prospects

J. T. Boyd
CERN, Geneva, Switzerland

Abstract

The lecture discusses both the current status of the Large Hadron Collider as well as its future running scenarios. In addition, a selection of the latest physics results from the experiments ATLAS, CMS and LHCb is presented.

Keywords

Particle accelerators, Particle physics, LHC Run 2; Lectures.

1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the highest-energy particle collider in the world. The accelerator sits in a 27 km long tunnel, originally constructed for the Large Electron–Positron Collider (LEP), 100 metres underground at CERN, on the Franco-Swiss border near Geneva, Switzerland. It is an extremely sophisticated machine, using super conducting 8 T dipole magnets to steer the high-energy proton beams. The magnets are cooled to an operating temperature of 1.9 K by using superfluid liquid helium. Given the extreme energy of the beams, the LHC needs a complex machine protection system, relying on a large number of beam instrumentation devices to monitor the beam position and beam losses.

The two key parameters for a collider are the collision energy and the luminosity L, which is a measure of the number of collisions. The number of events for a specific process (N) is given by $N = \sigma \times L$, where σ is the production cross-section for that process and L is the integrated luminosity.

The luminosity at a collider is given by the formula: $L = n_\mathrm{b} N_1 N_2 F / 4\pi\epsilon\beta^*$ and can be increased by augmenting the number of protons per bunch (N_1, N_2), the number of colliding bunches (n_b), or reducing the transverse size of the beam at the collision point. This can be done by using a lower emittance (ϵ) beam, or by squeezing the beam more with the focusing magnets (reducing β^*). The crossing-angle between the beams, needed to avoid parasitic collisions due to the short distance between bunches, reduces the luminosity, and is encapsulated in the geometric factor F in the equation.

The main machine parameters for the LHC are shown in Table 1, for the design, Run 1, Run 2, as well as the expectation for Run 3 and the high-luminosity upgrade (HL-LHC). It can be seen that all of the design parameters have been exceeded, except the collision energy, and the number of colliding bunches. The LHC experts have continually improved the running scenario to increase the luminosity, and during Run 2 the design luminosity of 10^{34} cm^{-2}s^{-1}, was achieved and surpassed by a factor of two at the end of Run 2. As well as improving the instantaneous luminosity, the availability of the machine was dramatically improved from 2016 to 2018, which led to a large physics dataset. The machine provided collisions during 50% of the allocated physics time—a very impressive performance for a super conducting collider. An important parameter for the LHC experiments is the pileup, which is determined by the luminosity per bunch, and is a measure of the number of inelastic pp interactions that occur per bunch crossing. Higher pileup gives more luminosity (for a fixed number of bunches), but makes physics analysis more difficult due to the signals in the detector from the additional interactions.

Construction of the HL-LHC should be completed in 2026 and will be followed by at least ten years of operation, with the goal of reaching 3000 fb^{-1} in 2037 (an increase of a factor of ten compared to the expected dataset at that time). In order to achieve this, the injector needs to be upgraded to provide a higher intensity beam, the focusing magnets will be replaced to squeeze the beam further, and various components will be upgraded to cope with the increased radiation and stored energy. The pileup in ATLAS and CMS will increase significantly (to a maximum of 200 interactions per bunch crossing) and

http://doi.org/10.23730/CYRSP-2021-005.247

the detectors will need large upgrades to be able to make physics measurements at this large pileup, as well as to cope with the associated radiation.

Table 1: Summary of main accelerator parameters for the LHC, showing the design values, and those used during Run 1 and Run 2, as well as the expected parameters for Run 3 and the HL-LHC.

Parameter	Design	Run 1	Run 2	Run 3	HL-LHC
Energy [TeV]	14	7/8	13	14	14
Bunch spacing [ns]	25	50	25	25	25
Bunch intensity [10^{11} ppb]	1.15	1.6	1.2	up to 1.8	2.2
Number of bunches	2800	1400	2500	2800	2800
Emittance [μm]	3.5	2.2	2.2	2.5	2.5
β^* [cm]	55	80	$30 \rightarrow 25$	$30 \rightarrow 25$	down to 15
Crossing angle [μrad]	285	-	$300 \rightarrow 260$	$300 \rightarrow 260$	TBD
Peak luminosity [10^{34} cm^{-2}s^{-1}]	1.0	0.8	2.0	2.0	7.5
Peak pileup	25	45	60	55	200

2 Run 2 physics highlights and future prospects

2.1 The LHC detectors

ATLAS and CMS are the general-purpose detectors at the LHC with the same physics goals. There are significant differences in the detector designs, but despite these they have very similar physics performance.

The main differences in the detectors relate to the magnet design. ATLAS is equipped with a 2 T solenoid to provide the magnetic field to bend charged particles in the central detector region, with three toroidal magnets (one barrel and two endcap toroid systems) to bend muons in the muon spectrometer. CMS uses a single large solenoid with field of 3.8 T for both of these roles. Following on from this, the ATLAS calorimeters are placed outside the thin solenoid, whereas the CMS calorimeters are placed inside the solenoid.

2.2 The Run 2 dataset

During the LHC Run 2 period, from 2015 to 2018, the ATLAS and CMS experiments collected, each of them, a sample of pp collisions at a center of mass energy of 13 TeV corresponding to an integrated luminosity of around 140 fb^{-1}. These large data samples could be collected thanks to the exceptionally good LHC operation efficiency and to instantaneous luminosities exceeding the design value (10^{34} cm^{-2}s^{-1}). The downside of running at such high instantaneous luminosities is that the interesting pp collision (the one that gives the trigger for the readout of the data) occurs together with many other pp collisions in the same bunch crossing, the so-called pileup. For example, the average pileup in the ATLAS experiment increased from around 13 collisions in 2015 to around 36 collisions in 2018, as shown in the left panel of Fig. 1, where we also see that some fraction of the collected events include almost 70 “extra" simultaneous pp collisions. To cope with the challenges induced by such large pileup values, the experiments developed improved procedures, at all levels, from the trigger to the offline data reconstruction and analysis. Examples of successful outcomes of those improvements are shown in the middle and right panels of Fig. 1.

2.3 Higgs physics

The main modes for Higgs production at the LHC are (in order of decreasing cross-section): gluon-fusion (ggF), vector-boson fusion (VBF), production in association with a vector boson (VH) and production in association with a pair of top-quarks (ttH). In VBF production, the scattered quarks are likely to form

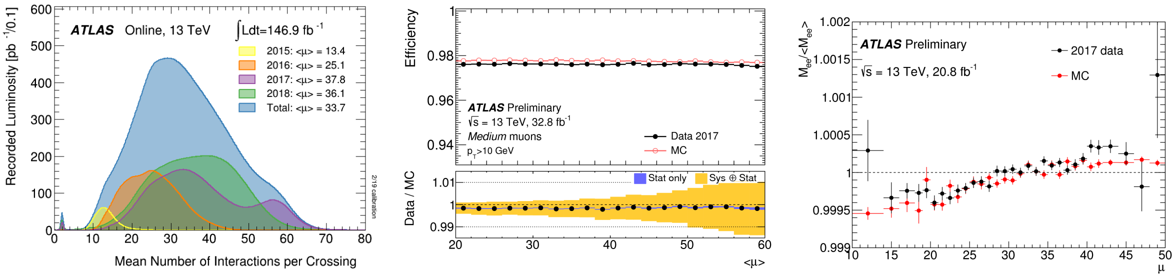


Fig. 1: (left) The Run 2 pileup distribution. An example of the pileup robustness of (middle) the reconstructed muon efficiency and (right) the electron energy scale.

forward jets on the two sides of the detector, which can be used to tag such events. The main decay modes for the Higgs boson are shown in Table 2. Experimentally the modes with the best mass resolution are important, as this allows to separate the signal from the background in a much more reliable way. The $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ ($\ell = e/\mu$) have both excellent mass resolution of ≈1–2%. These were the modes used for the Higgs discovery in 2012, despite the fact they have very low branching fractions (BF). Figure 2 shows the mass distributions for these two channels for the full Run 2 dataset. For $H \to \gamma\gamma$ the signal to background (S/B) is low, but the total number of selected Higgs events is a few thousand, whereas for $H \to 4\ell$ the S/B is high. The total number of signal events is an order of magnitude less.

Table 2: Summary of Higgs decay modes (BFs and resolutions) for the 125 GeV mass SM Higgs boson. For the good mass resolution channels involving a Z-boson, the resolution is only good for leptonic decays of the Z.

Poor mass resolution channels		Good mass resolution channels	
Decay mode	BF (%)	Decay mode	BF (%)
$H \to b\bar{b}$	58.2	$H \to ZZ^*$	2.6 (0.012 e,μ)
$H \to WW^*$	21.4 (1.1 e,μ)	$H \to \gamma\gamma$	0.23
$H \to gg$	8.2	$H \to Z\gamma$	0.15 (0.008 e,μ)
$H \to \tau^+\tau^-$	6.3	$H \to \mu^+\mu^-$	0.02
$H \to c\bar{c}$	2.9		

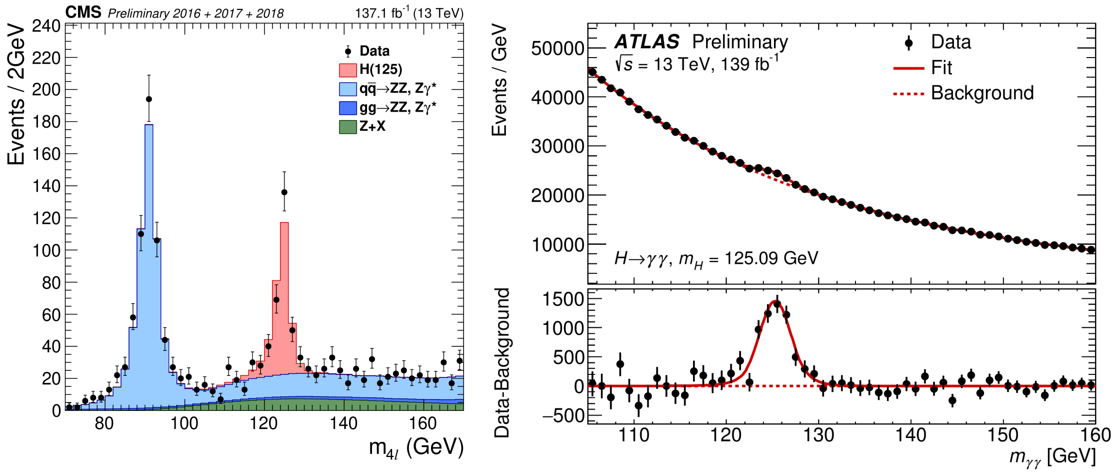


Fig. 2: Reconstructed Higgs candidate mass distributions in the $H \to \gamma\gamma$ (left) and $H \to 4\ell$ (right) channels.

Table 3 shows the status of the main Higgs production and decay modes, in terms of the significance of the measured signal. It shows that all of the main production and decay modes have been established, although many of these were only observed in the last year.

Table 3: Status of the measured significance for the main Higgs production and decay modes. Here Obs./Evid. means the significance is at the level of an observation/evidence, UL stands for 'upper limit' and '-' implies this mode has not been studied yet.

	$\gamma\gamma$	ZZ^*	WW^*	$b\bar{b}$	$c\bar{c}$	$\tau^+\tau^-$	$\mu^+\mu^-$	**Combined**
ggF	Obs.	Obs.	Obs.	-	-	UL	UL	Obs.
VBF	UL	UL	UL	UL	-	Evid.	UL	Obs.
VH	UL	UL	UL	Obs.	UL	-	-	Obs.
ttH	Evid.	UL	Evid.	UL	-	Evid.	-	Obs.
Combined	Obs.	Obs.	Obs.	Obs.	UL	Obs.	UL	-

The Higgs coupling to fermions was established with the observation of the $H \to \tau^+\tau^-$ decay. The analysis selects events with two τs (that either can decay hadronically or leptonically), and uses selections targeting either VBF Higgs production or high-p_T ggF Higgs production to reduce the backgrounds. The main background is $Z \to \tau^+\tau^-$ which has the same final state, with $\approx$1000$\times$ higher cross-section and with a similar di-τ mass (the mass resolution is not sufficient to be able to resolve the two processes). Figure 3 shows the di-τ mass distribution from the Run 2 CMS analysis [1] where a tiny signal can be seen on top of the large $Z \to \tau\tau$ background. The analysis measured the $H \to \tau\tau$ rate to be compatible with the SM expectation with a precision of $\approx$30%, corresponding to a 5.9 σ observation of the process. Searches for $H \to \mu^+\mu^-$ have found no evidence of a signal (as expected in the SM for the current dataset), which when combined with the $H \to \tau^+\tau^-$ result, demonstrates that the Higgs couplings do not obey lepton flavour conservation.

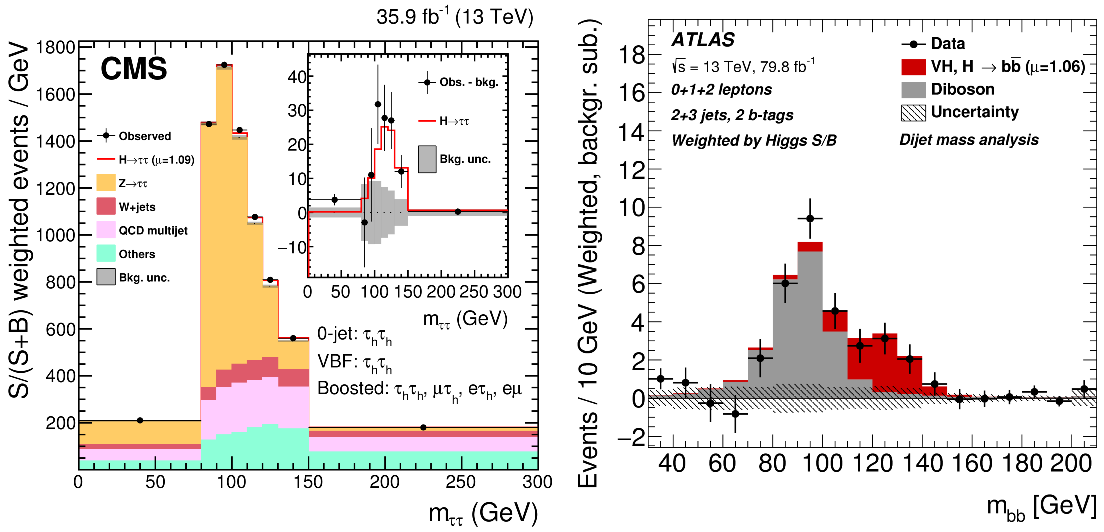


Fig. 3: (left) The di-tau mass distribution from the CMS $H \to \tau^+\tau^-$ analysis; (right) The di-b-jet mass distribution from the ATLAS $H \to b\bar{b}$ analysis.

The Higgs coupling to quarks was established with the observation of $H \to b\bar{b}$. Although this has the largest Higgs decay BF, it is experimentally challenging due to the large background from QCD $b\bar{b}$ production, and the poor di-b-jet mass resolution. In order to reduce the background and to trigger on the events, the analysis targets VH production where V is a Z or W boson decaying leptonically, so the final state can have 2 leptons, 1 lepton or 0 leptons (but large missing transverse momentum (MET) from the $Z \to \nu\bar{\nu}$ decay), and selects two b-jets with a mass close to the Higgs mass. As seen in Fig. 3, which shows the ATLAS analysis [2], the VZ, $Z \to b\bar{b}$ decay acts as an important validation of the analysis. This has the same final state, with a similar rate. The figure shows that VZ is observed with the expected rate (grey), and the $H \to b\bar{b}$ signal can be seen as a high mass shoulder on the Z peak. The analysis finds the expected SM rate with a precision of ≈30%.

The Higgs is too light to decay to top quarks, so the top-Higgs (tH) coupling can only be directly probed through ttH production. In the SM the ggF production process is dominated by a top-quark in the ggF loop, and so the tH coupling can also be extracted indirectly from ggF production rates. The direct and indirect measurement of the coupling then allow to constrain possible new-particles that could enter the ggF loop. Within the current precision the direct and indirect measurements of the coupling are compatible.

With the increased luminosity at the HL-LHC, the Higgs physics goals are:

- Improve the precision on the Higgs couplings to the few-% level (where they can be sensitive to effects beyond the SM (BSM);
- Establish the coupling to 2nd generation fermions through the $H \to \mu^+\mu^-$ and $H \to c\bar{c}$ decays;
- Improve the constraints on forbidden Higgs decays such as $H \to$ invisible and lepton-flavour violating Higgs decays;
- Make more precise differential measurements of Higgs production in more extreme regions of phase-space, which could be sensitive to new physics;
- Observe the very rare di-Higgs production process.

Studying di-Higgs production is needed to understand the Higgs self-coupling, and to probe the Higgs potential term of the SM Lagrangian. However, it is doubtful that this will be possible at the HL-LHC. Current projections [3, 4] suggest that evidence for di-Higgs production can be achieved by combining the ATLAS and CMS HL-LHC results.

2.4 Searches for physics beyond the Standard Model

One of the primary goals of the LHC is to search for the direct production of BSM physics. ATLAS and CMS have carried out a huge number of searches, but to date no significant excess of events over the SM expectation has been observed. A few example searches are discussed below.

A search for a new gauge boson (Z') that is similar to the SM Z boson but with much higher mass, looks for an excess of events in the di-lepton mass spectra at high mass. Figure 4 shows the di-electron and di-muon mass distributions from the ATLAS search [5]. No significant deviation from the expected background (dominated by SM Drell-Yan production) is observed. Examples signals are shown in the figures, which show that the mass resolution is significantly better at high mass for electrons than for muons, as the energy resolution improves for calorimeters, but deteriorates for tracking detectors, at higher energy. The main experimental challenge for this search is to have good efficiency and resolution for very high transverse momentum leptons (up to 2 TeV).

At the other end of the spectrum is a search for Higgsino production where very low-momentum leptons are expected. The ATLAS search [6] uses leptons with p_{T} down to 3 GeV (muons) and 4.5 GeV (electrons) in order to improve the sensitivity, and allows to exclude Higgsinos with masses up to 150 GeV for certain mass splittings.

Searching for dark matter (DM) production in LHC collisions can be done by taking advantage of initial-state-radiation, which can be used to tag events where DM particles are pair produced through an s-channel mediator particle but escape the detector without interacting with it. This can lead to a detector signature of a high-p_{T} jet + MET. Figure 5 shows the MET spectrum for such events from the CMS search [7], also showing the expected background, dominated by $Z \to \nu\bar{\nu}$ + jets (≈60%) and $W \to \ell\nu$ + jets (where the lepton is not reconstructed) (≈30%). The signal has a slightly harder MET-spectra than the background, but is much smaller than the background, meaning the background needs

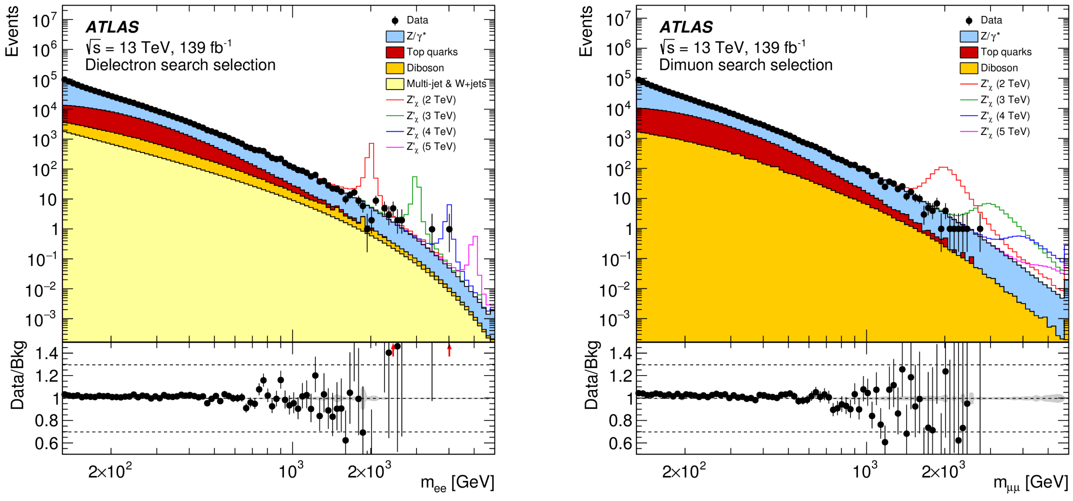


Fig. 4: The di-electron and di-muon mass distributions from the ATLAS Z' search.

to be controlled at the few-% level to allow to have sensitivity. The background is estimated from data control regions with $Z \rightarrow \ell^+\ell^-$ + jets, $W \rightarrow \ell\nu$ + jets and γ + jets but accurate theoretical predictions are needed on the ratio of Z + jets/γ + jets and Z + jets/W + jets; in order to achieve the needed precision NNLO electroweak corrections need to be taken into account.

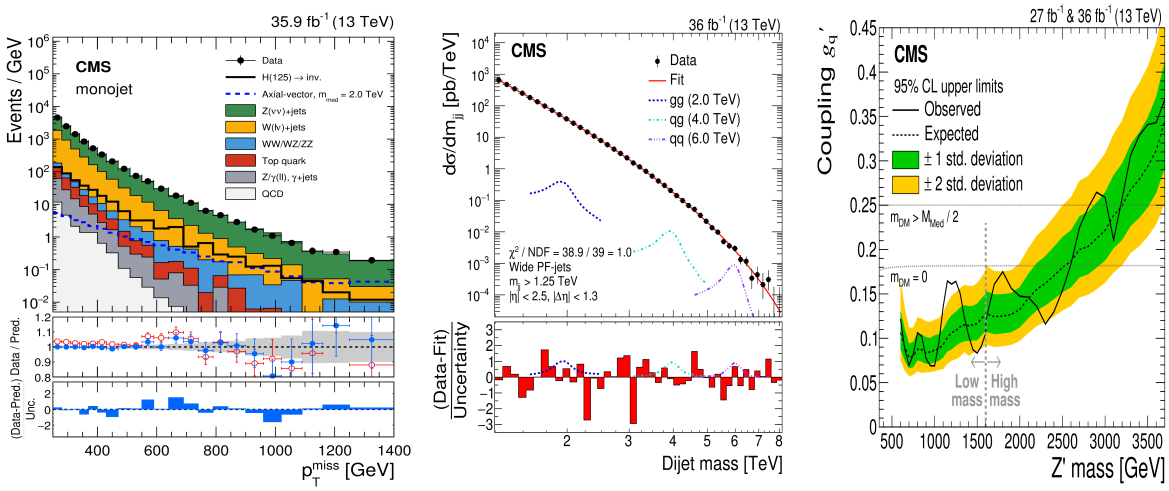


Fig. 5: (left) The MET spectrum in the DM search; (middle) The di-jet mass spectra in the mediator search; (right) The exclusion limit in the search for the mediator showing results for both the high mass search, and the low mass search that uses the *Data Scouting* technique.

As well as searching for the DM particle, we can also search for a mediator that can be produced in the LHC collisions, but decays back to SM particles (for example to two quarks). This could show up as a resonance in the di-jet mass spectra. Figure 5 shows the di-jet mass distribution for such a CMS search [8], showing a smoothly falling distribution with no sign of a resonance in the range 1 TeV to 8 TeV in the di-jet mass. The 1 TeV lower limit in the probed mass range comes from the trigger thresholds applied to the jets used in the search. Going to lower p_T-jets would increase the trigger rate leading to a too high bandwidth when reading out the detector. In order to search for possible resonances at lower mass a new technique called *Data Scouting* or *Trigger Level Analysis* was developed, in which just the trigger level jets are written out for certain triggers. These trigger level jets are much smaller than the full event data (less than 5% of the size), and can therefore be read out at a much higher rate without hitting bandwidth limitations. Thus lower thresholds can be applied. This technique allows for setting limits on di-jet mass resonances down to lower masses, as can be seen in Fig. 5 (CMS analysis [8]).

2.5 Precise Standard Model measurements

The LHC experiments carry out a large number of precise measurements of SM processes, measuring cross-sections, masses and other SM parameters. Cross-section measurements are normalized by the luminosity, which is measured by dedicated luminosity detectors in the experiments that are calibrated by dedicated van-der-Meer scans that are typically carried out each year. The precision of the luminosity measurements in ATLAS and CMS for the Run 2 dataset is an impressive ≈2.5%, which is far better than had thought to be possible before LHC running.

An example of a very precise cross-section measurement is the W and Z inclusive production cross-section measurement from ATLAS [9] with the 2011 7 TeV dataset. The precision is limited by systematic uncertainties, and the total experimental uncertainty is ≈0.5% dominated by uncertainties related to the lepton reconstruction, the background (for the W) and theoretical modeling uncertainties (for the Z). The luminosity uncertainty is 1.8%, but this cancels in ratios such as $\sigma(W \to e\nu)/\sigma(W \to \mu\nu)$ or $\sigma(W)/\sigma(Z)$ allowing very precise tests of lepton flavour conservation, and parton distribution functions (PDFs).

The measurement of the W-boson mass by ATLAS [10], with a precision of 19 MeV, represents one of the most precise measurements at the LHC and has a precision equal to the best single-experiment measurement. The W-mass is a fundamental parameter of the SM, and has important sensitivity in the electroweak fit. The ATLAS analysis measures the mass using a template, which fits to the transverse mass (formed from the lepton and the reconstructed hadronic recoil) and to the lepton transverse momentum. A very precise knowledge of experimental effects related to lepton reconstruction and the hadronic recoil reconstruction is needed, where the later deteriorates significantly with pileup. The current measurement utilizes the 2011 7 TeV data set which has an average pileup of around 9. Theoretical uncertainties also play an important role, in particular related to the modelling of the W-boson p_{T} which is derived from the measured Z-boson p_{T} spectra, as well as from Parton Distribution Function (PDF) uncertainties. Utilizing low-pileup data taken in 2017 and 2018 at 13 TeV there is the prospect of improving the precision of the measurement to the 10–15 MeV level.

Measurements of the top-quark mass are carried out in a number of different channels. A recent example from CMS [11] utilizes the lepton + jets final state to measure the mass using a kinematic fit (including the W-mass constraint on the hadronic W decay) to improve the resolution and to reduce the fraction of incorrect assignments of jets to the two top-quarks. The dominant systematic uncertainty is related to the jet energy scale which is constrained in the fit. The final result of 172.25 ± 0.08(stat.)± 0.62(syst.) GeV is the most precise single measurement to date.

2.6 Flavour physics

The $B_S \to \mu^+\mu^-$ rare decay is theoretically clean, and has a large sensitivity to many new physics models (for example MSSM scenarios with large tan β). Because of this, there is a long history of searches for this decay that started over 30 years ago. Sensitivity to the SM branching ratio of $(3.3 \pm 0.3) \times 10^{-9}$ was reached with a combination between LHCb and CMS [12]. Despite a much smaller dataset, LHCb has the best sensitivity due to the excellent track resolution, as well as an optimized trigger for low p_{T} physics; CMS has better sensitivity than ATLAS, due to the higher magnetic field in the inner tracker, which gives a better mass resolution. Current measurements from all three experiments are consistent with the SM estimate with an uncertainty from 20 to 30%.

LHCb searches for lepton flavour violation in B meson decays by measuring the ratio $R_{K^{(*)}} \equiv$

$\mathrm{BF}(B \to K^{(*)}\mu^+\mu^-)/\mathrm{BF}(B \to K^{(*)}e^+e^-)$. In the SM this is precisely predicted and is close to unity, modulo phase-space effects. Bremsstrahlung represents an experimental complication as the mass resolution is much worse in the di-electron channel than in the di-muon channel (as can be seen in Fig. 6). This is corrected for by normalizing by the measured ratio $\mathrm{BF}(B \to K^{(*)}J/\psi(\mu^+\mu^-))$ / $\mathrm{BF}(B \to K^{(*)}J/\psi(e^+e^-))$. For R_K, the reconstructed B meson mass distribution is shown in Fig. 6 in both the $\mu^+\mu^-$ and e^+e^- channels. The measured values from the LHCb measurements [13, 14] are shown in Table 4, and for R_{K^*} in Fig. 6. It is shown that the three measurements are between 2 and 2.5 σ lower than the SM prediction. This is currently one of the most intriguing anomalies observed by the LHC experiments, with many theoretical models proposed to explain the results. More data and measurements from Belle-2 should shed further light onto the situation.

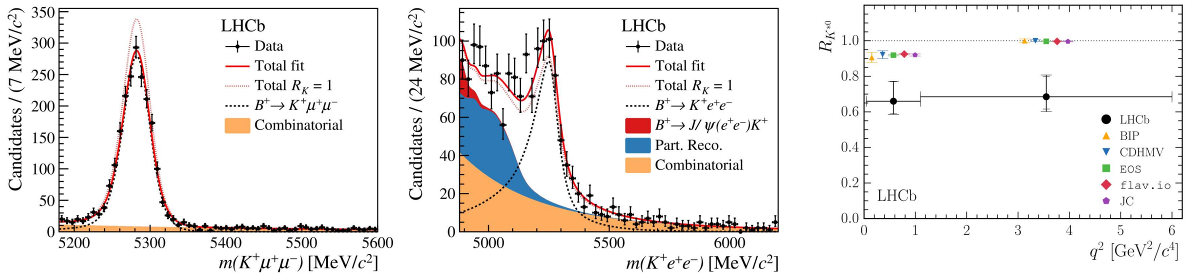


Fig. 6: Left/middle: The reconstructed B meson mass in the R_K analysis for the $\mu^+\mu^-$ / e^+e^- channels; Right: The measured R_{K^*} values in two bins of q^2 (the di-lepton mass) compared with various theoretical predictions.

Table 4: LHCb results on lepton flavour violation measurements R_K and R_{K^*}, where the latter is measured in two regions of q^2 (the di-lepton mass).

Measurement	Dataset	Measured value	Compatibility with SM
R_K	Run 1 + Run 2	$0.85^{+0.06}_{-0.05} \pm 0.015$	2.5σ
R_{K^*} low-q^2	Run 1	$0.66^{+0.11}_{-0.07} \pm 0.03$	2.2σ
R_{K^*} high-q^2	Run 1	$0.69^{+0.11}_{-0.07} \pm 0.05$	2.4σ

3 Summary

The LHC machine and the experiments performed extremely well in Run 2. A large and high-quality dataset was produced by the experiments leading to a huge number of physics results. A leading challenge for the experiments was the high pileup in the data, but they have coped very well with this situation.

The large dataset has allowed a more and more precise probing of the Higgs boson, where all major production modes and decay channels accessible at the LHC have been established. A huge number of direct searches for BSM physics have been carried out, with no significant excess of events over the SM prediction observed, such that increasingly stringent exclusion limits have been set on BSM model parameters. In addition, the experiments have been able to make very precise measurements of cross-sections and SM parameters, as well as measuring extremely rare processes, but again no discrepancy with the SM expectations have been observed. An intriguing set of results from lepton-flavour violation measurements by LHCb show a 2–2.5 standard deviation discrepancy with the SM in a few channels and q^2-bins.

The increased dataset that will be produced with Run 3, and then with the HL-LHC, along with the upgraded detector functionality, and innovations in triggering, reconstruction and physics analysis will allow to probe further the SM in the coming years.

References

[1] A.M. Sirunyan *et al.* [CMS Collaboration], *Phys. Lett.* **B779** (2018) 283–316, doi:10.1016/j.physletb.2018.02.004.

[2] M. Aaboud *et al.* [ATLAS Collaboration], *Phys. Lett.* **B786** (2018) 59–86, doi:10.1016/j.physletb.2018.09.013.

[3] CMS Collaboration, Sensitivity projections for Higgs boson properties measurements at the HL-LHC, CMS-PAS-FTR-18-011 (2018), https://cds.cern.ch/record/2647699.

[4] ATLAS Collaboration, Projections for measurements of Higgs boson cross sections, branching ratios, coupling parameters and mass with the ATLAS detector at the HL-LHC, ATL-PHYS-PUB-2018-054 (2018), http://cdsweb.cern.ch/record/2652762.

[5] G. Aad *et al.* [ATLAS Collaboration], *Phys. Lett.* **B796** (2019) 68–87, doi:10.1016/j.physletb.2019.07.016.

[6] G. Aad *et al.* [ATLAS Collaboration], *Phys. Rev.* **D101** (2020) 052005, doi:10.1103/PhysRevD.101.052005.

[7] A.M. Sirunyan *et al.* [CMS Collaboration], *Phys. Rev.* **D97** (2018) 092005, doi:10.1103/PhysRevD.97.092005.

[8] A.M. Sirunyan *et al.* [CMS Collaboration], *JHEP* **08** (2018) 130, doi:10.1007/JHEP08(2018)130.

[9] M. Aaboud *et al.* [ATLAS Collaboration], *Eur. Phys. J.* **C77** (2017) 367, doi:10.1140/epjc/s10052-017-4911-9.

[10] M. Aaboud *et al.* [ATLAS Collaboration], *Eur. Phys. J.* **C78** (2018) 110, doi:10.1140/epjc/s10052-017-5475-4
Erratum: *Eur. Phys. J.* **C78** (2018) 898, doi:10.1140/epjc/s10052-018-6354-3.

[11] A.M. Sirunyan *et al.* [CMS Collaboration], *Eur. Phys. J.* **C78** (2018) 891, doi:10.1140/epjc/s10052-018-6332-9.

[12] V. Khachatryan *et al.* [CMS and LHCb Collaborations], *Nature* **522** (2015) 68–72, doi:10.1038/nature14474.

[13] R. Aaij *et al.* [LHCb Collaboration], *JHEP* **08** (2017) 055, doi:10.1007/JHEP08(2017)055.

[14] R. Aaij *et al.* [LHCb Collaboration], *Phys. Rev. Lett.* **122** (2019) 191801, doi:10.1103/PhysRevLett.122.191801.

Scientific programme [1]

Practical statistics
Roger Barlow, University of Huddersfield, UK

Field theory and the E-W Standard Model
Alexander Bednyakov, JINR

LHC Run-2 and future prospects
Jamie Boyd, CERN

Special lecture on gravitational waves
Jo van den Brand, Nikhef, Netherlands

Higgs Physics
John Ellis, King's College London, UK and CERN

Neutrino physics
Concha Gonzalez-Garcia, Stony Brook, USA and University of Barcelona, Spain

Cosmology and dark matter
Valery Rubakov, INR and Moscow University, Russia

Physics beyond the Standard Model
Veronica Sanz, University of Sussex, UK

Flavour physics and CP violation
Mikhail Vysotsky, ITEP, Russia

QCD
Giulia Zanderighi, Max Planck Institute, Germany

Heavy-ion physics
Korinna Zapp, Lund University, Sweden

The CERN scientific programme
Fabiola Gianotti, CERN

The JINR scientific programme
Victor Matveev, JINR

[1] Slides available at https://indico.cern.ch/event/798971/timetable/?daysPerRow=5&view=nicecompact

Organizing committee

T. Donskova (Schools administrator, JINR)
C. Duhr (CERN & UCLouvain)
N. Ellis (CERN)
M. Mulders (CERN)
A. Olchevsky (JINR)
K. Ross (Schools administrator, CERN)

Local organizing committee

T. Donskova (JINR)
V. Kim (NRC Kurchatov Institute - PNPI & SPbPU, Russia)
V. Matveev (JINR)
A. Olchevsky (JINR)

International advisors

F. Gianotti (CERN)
M. Kovalchuk (NRC Kurchatov Institute)
V. Matveev (JINR)
G. Trubnikov (Ministry of Science & Higher Education, Russia)

Lecturers

R. Barlow (University of Huddersfield)
A. Bednyakov (JINR)
J. Boyd (CERN)
J. van den Brand (Nikhef)
J. Ellis (King’s College London & CERN)
C. Gonzalez-Garcia, (Stony Brook & University of Barcelona)
V. Rubakov (INR & Moscow University)
V. Sanz (University of Sussex)
M. Vysotsky (ITEP)
G. Zanderighi (Max Planck Institute)
K. Zapp (Lund University)

Discussion leaders

J.R. Gaunt (CERN)
A. Gladyshev (JINR)
A. Huss (CERN)
D. Levkov (INR)
E. Nugaev (INR)
E. Vryonidou (CERN)

Students

Deshan Kavishka ABHAYASINGHE
Jonatan ADOLFSSON
Pepijn Johannes BAKKER
Giovanni BARTOLINI
Alexander BOOTH
Sebastian BYSIAK
Ryan Bernard CALLADINE
Terry WS CHAN
Pu-Sheng CHEN
Aleksei CHUBYKIN
Valerio D'AMICO
Alessandro DA ROLD
Agostino DE IORIO
Maurizio DE SANTIS
Mariia DIDENKO
Michal DRAGOWSKI
Anatolii EGOROV
Yassine EL GHAZALI
Lorenz Konrad EMBERGER
Feruzjon ERGASHEV
Luis Ignacio ESTEVEZ BANOS
Mohammed FARAJ
Armin FEHR
Alexandra FELL
Tobias FITSCHEN
Guglielmo FRATTARI
Egor FROLOV
Beatriz GARCIA PLANA
Mario GRANDI
Eva Brottmann HANSEN
Nicole Michelle HARTMAN
Eirik HATLEN
Daniel HEUCHEL
Lesya HORYN
Simona ILIEVA
Antonio IULIANO
Karolina JURASKOVA
Jakub KANDRA
Philip Daniel KEICHER
Dias KEREIBAY
Petr KHARLAMOV
Martin KLASSEN
Vasilis KONSTANTINIDES
Anastasia KUROVA
King Wai KWOK
Roman LAVICKA
Konstantin LEHMANN
Clara Elisabeth LEITGEB
Jindrich LIDRYCH
Iacopo LONGARINI
Peter MAJOR
Cristina Ana MANTILLA SUAREZ
Laura MARTIKAINEN
Luca MARTINELLI
James MEAD
Simone MELONI
Mehrnoosh MOALLEMI
Vyacheslav MOISEEV
Denise MÜLLER
Jose Luis MUNOZ MARTINEZ
Yvonne NG
Yuval NISSAN
Jakob NOVAK
Michael William O'KEEFE
Isabella OCEANO
Vitalii OKHOTNIKOV
Gogita PAPALASHVILI
Botho PASCHEN
Vasilije PEROVIC
Henriette PETERSEN
Krystsina PETUKHOVA
Vasilii PLOTNIKOV
Louis PORTALES
Daria PROKHOROVA
Michael Philipp REICHMANN
Viktor ROMANOVSKII
Luigi SABETTA
Cristina SÁNCHEZ GRAS
Vladislav SANDUL
Valerie SCHEURER
Lara Katharina SCHILDGEN
Patrick SCHWENDIMANN
Anton SHUMAKOV
Viktor SINETCKII
Rafael Eduardo SOSA RICARDO
Dmitry SOSNOV
Amanda Lynn STEINHEBEL
Joel SWALLOW
Sergey SERGEY
Federico VAZZOLER
Christos VERGIS
Michele VERONESI
Janik VON AHNEN
Hendrik WINDEL
Ioannis XIOTIDIS
Hanlin XU
Milosz ZDYBAL
Jean-Philippe ZOPOUNIDIS
Davide ZUOLO

Posters

Poster title	Presenter
Study of strangeness enhancement in small systems through Ξ-baryon correlations in pp collisions at 13 TeV	ADOLFSSON, J.
Search for the associated production of Higgs boson and top quark pairs in multijet events with the ATLAS detector	BARTOLINI, G.
Probing lepton universality with $b \to sll$ decays	CALLADINE, R.
Search for lepton-flavour-violating decays of the Z boson into a τ-lepton and a light lepton with the ATLAS detector	CHAN, W. S.
Beyond Standard Model top-antitop resonance in the dilepton channel	DE SANTIS, M.
X-ray alignment test of the small-strip thin gap chambers for the ATLAS muon new small wheel upgrade	DIDENKO, M.
Search for di-boson resonance in semi-leptonic final state with the ATLAS detector at 13 TeV	EL GHAZALI, Y.
Study of the nucleaon transfer reactions in ^{10}B+^{12}C and ^{10}B+^{16}O interaction at the energies near the Coulomb barrier for nuclear astrophysics	ERGASHEV, F. KH.
Four-top-quarks analysis in the single-lepton and opposite-sign dilepton final states in pp collisions at 13 TeV with the ATLAS detector	FARAJ, M.
Calculating the dissociative contribution for exclusive WW production	FELL, A.
ITk Pixel Planar Sensor Market Survey	FITSCHEN, T.
Unveiling dark matter with the LHC	FRATTARI, G.
Lepton flavour universality tests using semitauonic decays at LHCb	GARCÍA, B.
Detecting dark matter from the galactic center with Gaussian processes	HATLEN, E. S.
The CALICE highly granular SiPM-on-tile hadron calorimeter prototype	HEUCHEL, D.
Searh for displaced leptons in the ATLAS detector	HORYN, L.

Poster title	**Presenter**
Hadron production measurements for neutrino experiments with the NA61/SHINE spectrometer	ILIEVA, S.
Neutrino detection in the SHiP experiment	IULIANO, A.
Precision and stability of the Belle II vertex detector	KANDRA, J.
Measurement of Higgs-boson production in association with a top quark-antiquark pair in the H→ $b\bar{b}$ channel	KEICHER, PH.
Straw leading time stability check in NA62	KEREIBAY, D.
Testing and quality assurance of BM@N silicon tracking modules	KHARLAMOV, P.
Search for Lepton-flavor Violation in Different flavor, High-mass Final States in *pp* Collisions	KING WAI, K.
The ATLAS level-1 calorimeter trigger energy calibration in Run 2	KLASSEN, M.
Misidentified particles in the ATLAS $H \to WW^* \to \ell\nu\ell\nu$ analysis	LEHMANN, K.
Search for direct stau production with ATLAS	LEITGEB, C.
Dark photon decaying to displaced lepton jets – a search for unconventional signatures with the ATLAS detector	LONGARINI, I.
Studies of the factorizability of bunch proton densities	MAJOR, P.
Boosting low mass resonances with ISR	MANTILLA SUAREZ, C.
Top antitop cross section measurements with lepton kinematics using full Run II dataset of ATLAS	MARTINELLI, L.
Top charge asymmetry	MEAD, J. V.
Study FCNC interactions at the FCC-ee via top-Z-jet production	MOALLEMI, M.
Search for the s-channel single top quark production at the CMS experiment	MÜLLER, D.
Low mass dijet resonances search using ISR with 80 fb^{-1} 13 TeV ATLAS data	NG, Y.
Search for compressed mass Higgsino production at CMS using tracks	NISSAN, Y.

Poster title	**Presenter**
Search for gauge-boson resonances in events with a charged lepton and missing transverse momentum using the full Run-2 ATLAS dataset	O'Keefe, M.
The PADME experiment	Oceano, I.
Overview of the CMS BCML system and the potential of pCVD diamond detectors surface modification	Okhotnikov, V.
The KM3NeT project	Papalashvili, G.
Module production tests and integration of the Belle II pixel detector	Paschen, B.
Serial powering for the CMS pixel detector	Perovic, V.
Measurements of multi-differential cross section for $t\bar{t}$ production at 13 TeV with the CMS experiment	Petersen, H.
Higgs boson mass reconstruction in the ATLAS $H \to \tau\tau$ analysis	Petukhova, K.
Analysis of K$^+$/π^+ at BM@N for argon run	Plotnikov, V.
Triggering a muon: Machine learning at work	Sabetta, L.
Central exclusive production of J/ψ mesons in pp collisions at LHCb	Sánchez Gras, C.
Calibration of the BCM1F detector	Scheurer, V.
Study of future 3D calorimetry based on LYSO or LaBr:Ce crystals for high precision physics	Schwendimann, P.
Tracking detectors based on CVD diamonds	Reichmann, M.
Exploiting tau decay mode information in the $H \to \tau\tau$ coupling measurement	Schildgen, L.
Search for Higgs $\to$ invisible with the ATLAS detector	Steinhebel, A.
Searches for $K^+ \to \pi\mu e$ Decays at NA62	Swallow, J.
Study of the process e$^+$e$^-$ $\to \pi^+\pi^-\gamma$ with the CMD-3 detector at the VEPP-2000 e$^+$e$^-$ collider	Tolmachev, S.

Poster title	**Presenter**
Search for high-mass resonances decaying to $\tau\nu$ in pp collisions at 13 TeV with the ATLAS detector	VERGIS, C.
Mixing and CP violation in B to open-charm decays	VERONESI, M.
Track reconstruction for the SPASCHARM	VYACHESLAV, M.
CLAWS–Scintillators with SiPM readout Monitoring Injection Backgrounds at SuperKEKB	WINDEL, H.
A hardware tracking system for the trigger at the High-Luminosity LHC	XIOTIDIS, I.
R&D of new silicon pixel sensors for the High Luminosity upgrade of the CMS experiment at LHC	ZUOLO, D.