

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

2015 CERN–Latin-American School of High-Energy Physics

Ibarra, Ecuador
5 – 17 March 2015

Proceedings

Editors: M. Mulders
G. Zanderighi

ISBN 978-92-9083-436-6 (paperback)


ISBN 978-92-9083-437-3 (PDF)

ISSN 0531-4283

DOI <https://doi.org/10.5170/CERN-2016-005>

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

Copyright © CERN, 2016

 Creative Commons Attribution 4.0

Knowledge transfer is an integral part of CERN's mission.

CERN publishes this report Open Access under the Creative Commons Attribution 4.0 license (<http://creativecommons.org/licenses/by/4.0/>) in order to permit its wide dissemination and use. The submission of a contribution to a CERN Yellow Report shall be deemed to constitute the contributor's agreement to this copyright and license statement. Contributors are requested to obtain any clearances that may be necessary for this purpose.

This report is indexed in: CERN Document Server (CDS), INSPIRE, Scopus.

This report should be cited as:

Proceedings of the 2015 CERN–Latin-American School of High-Energy Physics, Ibarra, Ecuador, 4 – 17 March 2015, edited by M. Mulders and G. Zanderighi, CERN-2016-005 (CERN, Geneva, 2016), <https://doi.org/10.5170/CERN-2016-005>

A contribution in this report should be cited as:

[Author name(s)], in Proceedings of the 2015 CERN–Latin-American School of High-Energy Physics, Ibarra, Ecuador, 4 – 17 March 2015, edited by M. Mulders and G. Zanderighi, CERN-2016-005 (CERN, Geneva, 2016), pp. [first page]–[last page], <https://doi.org/10.5170/CERN-2016-005>. [first page]

Abstract

The CERN–Latin-American 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 the Standard Model of electroweak interactions, flavour physics, neutrino physics, Higgs physics, new physics beyond the standard model, quantum chromodynamics under extreme conditions, cosmology, an introduction to experimental facilities at the high-energy frontier, and practical statistics for particle physicists.

Preface

The eighth School in the series of Latin-American Schools of High-Energy Physics took place from 4 to 17 March 2015 in Ibarra, Ecuador. It was organized by CERN with the support of local colleagues from Escuela Politécnica Nacional (EPN), Quito, and Universidad San Francisco de Quito (USFQ).

The School received financial support from: CERN; CIEMAT, Spain; the Ecuadorian funding agency Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT), EPN and USFQ. Our sincere thanks go to all of these sponsors for making it possible to organize the School with many young participants from Latin-American countries who otherwise would not have been able to attend. We would particularly like to acknowledge the support of SENESCYT that financed the attendance of 13 students from Ecuador, in addition to contributing to general expenses for the School.

The School was hosted in the beautiful and traditional Hacienda Chorlavi on the outskirts of the city of Ibarra. We are indebted to the hotel and its friendly staff for their help in making the event such a success. In particular, we would like to mention Andrea and Carolina Tobar who helped us greatly in preparing the School as well as during the event itself.

Edgar Carrera from USFQ acted as local director for the School, assisted by members of the local organising committee. We are extremely grateful to Edgar and his colleagues for their excellent work in organizing the School and for creating such a wonderful atmosphere for the participants. We would especially like to thank Andrea Ayala and Silvana Guitarra, for their invaluable help and presence throughout the School.

Sixty-nine students of 19 different nationalities attended the School. Following the tradition of the School the students shared twin rooms mixing nationalities, and in particular the Europeans mixed with Latin Americans.

The 11 lecturers came from Europe, Latin America and the USA. The lectures, which were given in English, were complemented by daily discussion sessions led by five physicists coming from Latin America. The lectures and the discussion sessions were all held using the conference facilities of the hotel. The students displayed their own research work in the form of posters in a special evening session during the first week. The posters were left on display until the end of the School. The students from each discussion group also performed a project, studying in detail the analysis of a published paper from an LHC experiment. A representative of each group presented a brief summary talk during a special evening session during the second week of the School.

Our 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 undoubtedly appreciated their personal contributions in answering questions and explaining points of theory.

The opening ceremony of the School was attended by high-level representatives of important organisations and universities in Ecuador. In particular, we would like to thank Rina Pazos, General Sub-Secretary of SENESCYT, Jaime Calderón, Rector of EPN, and Carlos Montúfar, President of USFQ, all of whom addressed the participants, as well as Fernando Albericio, Rector of Yachay Tech University and Daniel Larson, Chancellor of Yachay Tech University.

Hosting the School was an important event for the physics community in Ecuador and many outreach activities were arranged around it, benefitting from the presence of high-level scientists who were teaching at the School. In particular, a number of free public lectures were arranged in Quito and at Yachay, and there was coverage in the local, regional and national press.

We are very grateful to Kate Ross, the administrator for the CERN Schools of Physics, for her efforts in the lengthy preparations for the School and during the event itself. Her efficient work, friendly attitude, and continuous care of the participants and their needs were highly appreciated.

The participants will certainly remember the three interesting excursions: an afternoon visit to the spec-

tacular volcanic Cuicocha lake, and the towns of Otavalo and Cotacachi; a full-day excursion to the “Seven Waterfalls” reserve and the thermal springs at Chachimbiro; and a second afternoon excursion to the town of Ibarra followed by dinner at a restaurant with a beautiful panoramic view of the region. They also greatly appreciated evenings spent together in the hotel, especially the farewell party on the last night.

The success of the School was to a large extent due to the students themselves. Their poster session and group projects were very well prepared and highly appreciated, 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)





People in the photograph

1	Ilya Komarov	42	Guy De Teramond
2	Daniel Evangelho Vieira	43	Nerina Fidanza Romani
3	Joel Jones	44	Pawel Klimek
4	Peter Jenni	45	Sebastian Tapia Araya
5	Edgar Carrera	46	Julio Torres
6	Jean Iliopoulos	47	Eduardo Quintana Miranda
7	Daniel De Florian	48	Leandro Da Rold
8	Fabricio Jimenez Morales	49	Fabio Lucio Alves
9	Kate Ross	50	Melissa Cruz Torres
10	Giulia Zanderighi	51	Oscar Garcia
11	Nick Ellis	52	Jorge Fraga
12	Ma.Elena Tejeda Yeomans	53	Fernando Febres Cordero
13	Cesar Bernardes	54	Blanca Canas Orduz
14	Mauro Mendizabal	55	Gustavo Ramirez Hidalgo
15	Stephany Vargas	56	Christopher Bock
16	Oscar Taborda	57	Benjamin Grinstein
17	Santiago Paredes Saenz	58	Liliet Calero Diaz
18	Raquel Quishpe	59	Miqueias Melo De Almeida
19	Martijn Mulders	60	Henrique Correia Zanoli
20	David Romero Abad	61	Alvaro Chavarria
21	Sylvain Blunier	62	Arianna Batista Camejo
22	Xin Ran Liu	63	Luisa Fernanda Chaparro Sierra
23	Joaquin Grefa	64	Luis Manzanillas
24	Daniel Guerrero Ibarra	65	Christian Sarmiento
25	Barak Raimond Gruberg Cazon	66	Gabriela Bailas
26	David Hervas Aguilar	67	Yee Chinn Yap
27	Cristina Mantilla Suarez	68	Thiago Muhlbeier
28	Francisca Garay Walls	69	Jorge Castano Yepes
29	Rebecca Falla	70	Daniela Dominguez Damiani
30	Rocio Bermudez	71	Armando Bermudez Martinez
31	Liam Fulton	72	Jorge Nogueira
32	Roberto Morales Valbuena	73	Arturo Nunez
33	Sebastian Olivares Pino	74	Robinson Longas
34	Bruce Hoeneisen	75	Juan Pablo Carlomagno
35	Andres Baquero	76	Omar Roldan Garcia
36	Javier Mazzitelli	77	Yamila Rotstein Habarnau
37	Jaime Calderon	78	Jorge Lopez Lopez
38	Manfredi Ronzani		
39	Luis Otiniano		
40	Jose Camilo Diaz Bustamante	<i>Not in the photograph:</i>	
41	Adolfo Guevara	79	Gerardo Hernandez-Tome
		80	Julio Cesar Leal

PHOTOGRAPHS (MONTAGE)







Contents

Preface	
<i>N. Ellis</i>	v
Photograph of participants	vii
Photographs (montage)	x
Introduction to the Standard Model of Electro-Weak Interactions	
<i>J. Iliopoulos</i>	1
Flavor Physics and CP Violation	
<i>B. Grinstein</i>	43
Neutrino physics	
<i>P. Hernández</i>	85
Higgs Physics	
<i>C. Grojean</i>	143
Physics Beyond the Standard Model	
<i>R. Rosenfeld</i>	159
QCD under extreme conditions: an informal discussion	
<i>E.S. Fraga</i>	165
An introduction to cosmology	
<i>K.E. Kunze</i>	177
Experimental Facilities at the High Energy Frontier	
<i>P. Jenni</i>	213
Practical Statistics for Particle Physicists	
<i>L. Lyons</i>	245
Organizing Committee	271
Local Organizing Committee	271
List of Lecturers	271
List of Discussion Leaders	271
List of Students	272
List of Posters	273

Introduction to the Standard Model of the Electro-Weak Interactions

J. Iliopoulos

Laboratoire de Physique Théorique de L'Ecole Normale Supérieure, Paris, France

Abstract

These lectures notes cover the basic ideas of gauge symmetries and the phenomenon of spontaneous symmetry breaking which are used in the construction of the Standard Model of the Electro-Weak Interactions.

Keywords

Lectures; Standard Model; electroweak interaction; gauge theory; spontaneous symmetry breaking; field theory.

1 Introduction

These are the notes from a set of four lectures that I gave at the 2015 European Organization for Nuclear Research (CERN)–Latin-American School of High-Energy Physics as an introduction to more specialized lectures. With minor corrections, they follow the notes of the lectures I gave at the 2012 CERN Summer School. In both cases, the students were mainly young graduate students doing experimental high-energy physics. They were supposed to be familiar with the phenomenology of particle physics and to have a working knowledge of quantum field theory and the techniques of Feynman diagrams. The lectures were concentrated on the physical ideas underlying the concept of gauge invariance, the mechanism of spontaneous symmetry breaking, and the construction of the Standard Model. Although the methods of computing higher-order corrections and the theory of renormalization were not discussed at all in the lectures, the general concept of renormalizable versus non-renormalizable theories was supposed to be known. Nevertheless, for the benefit of the younger students, a special lecture on the physical principles of renormalization theory was included. It is given as an appendix in these notes. The plan of the notes follows that of the lectures with five sections:

- a brief summary of the phenomenology of the electromagnetic and the weak interactions;
- gauge theories, Abelian and non-Abelian;
- spontaneous symmetry breaking;
- the step-by-step construction of the Standard Model;
- the Standard Model and experiment.

It is generally accepted that progress in physics occurs when an unexpected experimental result contradicts the established theoretical beliefs. As Feynman put it “progress in physics is to prove yourself wrong as soon as possible”. This has been the rule in the past, but there are exceptions. The construction of the Standard Model is one of them. In the late 1960s, weak interactions were well described by the Fermi current \times current theory and there was no compelling experimental reason to want to change it: the problems were theoretical. It was only a phenomenological model which, in technical language, was non-renormalizable. In practice, this meant that any attempt to compute higher-order corrections in the standard perturbation theory would give meaningless, divergent results. So the motivation for changing the theory was for aesthetic rather than experimental reasons: it was the search for mathematical consistency and theoretical elegance. In fact, at the beginning, the data did not seem to support the theoretical speculations. Although the history of these ideas is a fascinating subject, I decided not to follow the historical evolution which would have taken more than four lectures to develop. I start instead from the experimental data known at present and show that they point unmistakably to what is known as the Standard Model. In the last section, I recall its many experimental successes.

Table 1: Our present ideas on the structure of matter. Quarks and gluons do not exist as free particles and the graviton has not yet been observed.

Table of elementary particles		
<i>Quanta of radiation</i>		
Strong interactions		Eight gluons
Electromagnetic interactions		Photon (γ)
Weak interactions		Bosons W^+ , W^- , Z^0
Gravitational interactions		Graviton (?)
<i>Matter particles</i>		
	Leptons	Quarks
1st family	ν_e, e^-	$u_a, d_a, a = 1, 2, 3$
2nd family	ν_μ, μ^-	$c_a, s_a, a = 1, 2, 3$
3rd family	ν_τ, τ^-	$t_a, b_a, a = 1, 2, 3$
Higgs boson		

2 Phenomenology of the electro-weak interactions: a reminder

2.1 The elementary particles

The notion of an ‘elementary particle’ is not well defined in high-energy physics. It evolves with time following progress in experimental techniques which, by constantly increasing the resolution power of our observations, have shown that systems that were believed to be ‘elementary’ are in fact composed of smaller constituents. So, in the last century we went through the chain:

molecules \rightarrow atoms \rightarrow electrons + nuclei \rightarrow electrons + protons + neutrons \rightarrow electrons + quarks
 \rightarrow ???

There is no reason to believe that there is an end to this series and, even less, that this end has already been reached. Table 1 summarizes our present knowledge, and the following remarks can be made.

- All interactions are produced by the exchange of virtual quanta. For the strong, electromagnetic, and weak interactions they are vector (spin-one) fields, whereas the graviton is assumed to be a tensor, spin-two field. We shall see in these lectures that this property is well understood in the framework of gauge theories.
- The constituents of matter appear to all be spin one-half particles. They are divided into quarks, which are hadrons, and ‘leptons’ which have no strong interactions. No deep explanation is known either for their number (why three families?) or for their properties, such as their quantum numbers. We shall come back to this point when we discuss the gauge-theory models. In the framework of some theories that go beyond the Standard Model, such as supersymmetric theories, we can find particles of zero spin among the matter constituents.
- Each quark species, called ‘flavour’, appears in three forms, often called ‘colours’ (no relation to the ordinary sense of either word).
- Quarks and gluons do not appear as free particles. They form a large number of bound states, known as the hadrons. This property of ‘confinement’ is one of the deep unsolved problems in particle physics.
- Quarks and leptons seem to fall into three distinct groups, or ‘families’. No deep explanation is known.
- The mathematical consistency of the theory, known as ‘the cancellation of the triangle anomalies’, requires that the sum of all electric charges inside any family is equal to zero. This property has strong predictive power.

2.2 The electromagnetic interactions

All experimental data are well described by a simple Lagrangian interaction in which the photon field interacts with a current created from the fields of charged particles.

$$\mathcal{L}_i \sim e A_\mu(x) j^\mu(x). \quad (1)$$

For the spinor matter fields of Table 1, the current takes the simple form

$$j^\mu(x) = \sum_i q_i \bar{\Psi}_i(x) \gamma^\mu \Psi_i(x), \quad (2)$$

where q_i is the charge of the field Ψ_i in units of e .

This simple Lagrangian has some remarkable properties, all of which are verified by experiment.

- j is a vector current. The interaction separately conserves P , C and T .
- The current is diagonal in flavour space.
- More complex terms, such as $j^\mu(x) j_\mu(x)$ and $\partial A(x) \bar{\Psi}(x) \dots \Psi(x), \dots$ are absent, although they do not seem to be forbidden by any known property of the theory. All these terms, as well as all others we can write, share one common property: in a four-dimensional space–time, their canonical dimension is larger than four. We can easily show that the resulting quantum field theory is *non-renormalizable*. For some reason, nature does not like non-renormalizable theories.

Quantum electrodynamics (QED), the quantum field theory described by the Lagrangian in Eq. (1) and supplemented with the programme of renormalization, is one of the most successful physical theories. Its agreement with experiment is spectacular. For years it was the prototype for all other theories. The Standard Model is the result of the efforts to extend the ideas and methods of electromagnetic interactions to all other forces in physics.

2.3 The weak interactions

Weak interactions are mediated by massive vector bosons. When the Standard Model was proposed, their very existence as well as their number were unknown. But today we know that three massive vector bosons exist; two which are electrically charged and one which is neutral: W^+ , W^- and Z^0 . Like the photon, their couplings to matter are described by current operators:

$$\mathcal{L}_i \sim V_\mu(x) j^\mu(x); \quad V_\mu : W_\mu^+, W_\mu^-, Z_\mu^0, \quad (3)$$

where the weak currents are again bi-linear in the fermion fields: $\bar{\Psi} \dots \Psi$. Depending on the corresponding vector boson, we distinguish two types of weak currents: *the charged current*, coupled to W^+ and W^- and the *neutral current* coupled to Z^0 , which have different properties.

The charged current:

- contains only left-handed fermion fields

$$j_\mu \sim \bar{\Psi}_L \gamma_\mu \Psi_L \sim \bar{\Psi} \gamma_\mu (1 + \gamma_5) \Psi; \quad (4)$$

- is non-diagonal in the quark flavour space;
- the coupling constants are complex.

The neutral current:

- contains both left- and right-handed fermion fields

$$j_\mu \sim C_L \bar{\Psi}_L \gamma_\mu \Psi_L + C_R \bar{\Psi}_R \gamma_\mu \Psi_R ; \quad (5)$$

- is diagonal in the quark flavour space.

With these currents, weak interactions have some properties which differ from those of the electromagnetic ones.

- Weak interactions violate P , C and T .
- In contrast to the photon, the weak vector bosons are self-coupled. The nature of these couplings is predicted theoretically in the framework of gauge theories and it has been determined experimentally.
- A new element has been added recently to the experimental landscape. It is a new scalar particle, compatible with what theorists have called *the Higgs boson*. Although all its properties have not yet been studied in detail, the existing evidence points towards the Higgs boson predicted by the Standard Model.

It is this kind of interaction that the Standard Model is supposed to describe.

3 Gauge symmetries

3.1 The concept of symmetry

In physics the concept of a symmetry follows from the assumption that a certain quantity is not measurable. As a result, the equations of motion should not depend on this quantity. We know from the general properties of classical mechanics that this implies the existence of conserved quantities. This relation between symmetries and conservation laws, epitomized by Noether's theorem, has been one of the most powerful tools in deciphering the properties of physical theories.

Some simple examples are given by the symmetries of space and time. The assumption that the position of the origin of the coordinate system is not physically measurable implies the invariance of the equations under space translations and the conservation of momentum. In the same way that we obtain the conservation laws of energy (time translations) and angular momentum (rotations), we can also distinguish between symmetries in *continuous transformations*, such as translations and rotations, and *discrete* symmetries, such as space or time inversions. Noether's theorem applies to the first. All symmetries of space and time are *geometrical* in the common sense of the word, and are easy to understand and visualize. During the last century we were led to consider two abstractions, each one of which has had a profound influence on our way of thinking about the fundamental interactions. Reversing the chronological order, we shall introduce first the idea of *internal* symmetries and second, that of local or *gauge* symmetries.

3.2 Internal symmetries

Internal symmetries are those with transformation parameters that do not affect the point of space and time x . The concept of such symmetries can be seen in classical physics, but it becomes natural in quantum mechanics and quantum field theory. The simplest example is the phase of the wave function. We know that it is not a measurable quantity, so the theory must be invariant under a change of phase. This is true for both relativistic or non-relativistic quantum mechanics. The equations of motion (Dirac or Schrödinger), as well as the normalization condition, are invariant under the transformation:

$$\Psi(x) \rightarrow e^{i\theta} \Psi(x) . \quad (6)$$

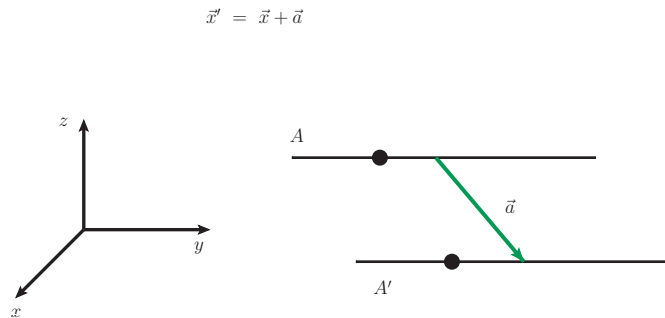


Fig. 1: A space translation by a constant vector \vec{a}

The transformation leaves the space–time point invariant, so it is an internal symmetry. Through Noether’s theorem, invariance under Eq. (6) implies the conservation of the probability current.

The phase transformation in Eq. (6) corresponds to the Abelian group $U(1)$. In 1932 Werner Heisenberg enlarged the concept to a non-Abelian symmetry with the introduction of isospin. The assumption is that strong interactions are invariant under a group of $SU(2)$ transformations in which the proton and the neutron form a doublet $N(x)$:

$$N(x) = \begin{pmatrix} p(x) \\ n(x) \end{pmatrix}; \quad N(x) \rightarrow e^{i\vec{\tau} \times \vec{\theta}} N(x), \quad (7)$$

where $\vec{\tau}$ are proportional to the Pauli matrices and $\vec{\theta}$ are the three angles of a general rotation in a three-dimensional Euclidean space. Again, the transformations do not apply on the points of ordinary space.

Heisenberg’s iso-space is three dimensional and isomorphic to our physical space. With the discovery of new internal symmetries the idea was generalized to multi-dimensional internal spaces. The space of physics, i.e. the space in which all symmetry transformations apply, became an abstract mathematical concept with non-trivial geometrical and topological properties. Only a part of it, the three-dimensional Euclidean space, is directly accessible to our senses.

3.3 Gauge symmetries

The concept of a local, or gauge, symmetry was introduced by Albert Einstein in his quest for the theory of general relativity¹. Let us come back to the example of space translations, as shown in Fig. 1.

Figure 1 shows that if A is the trajectory of a free particle, then its image, after a translation of the form $\vec{x} \rightarrow \vec{x} + \vec{a}$, A' , is also a possible trajectory of a free particle. The dynamics of free particles is invariant under space translations by a constant vector. It is a *global* invariance, in the sense that the parameter \vec{a} is independent of the space–time point x . Is it possible to extend this invariance to a *local* one, namely one in which \vec{a} is replaced by an arbitrary function of x ; $\vec{a}(x)$? One usually calls the transformations in which the parameters are functions of the space–time point x *gauge transformations*². There may be various, essentially aesthetic, reasons for which one may wish to extend a global invariance to a gauge one. In physical terms, it can be argued that the formalism should allow for a local definition

¹It is also present in classical electrodynamics if one considers the invariance under the change of the vector potential $A_\mu(x) \rightarrow A_\mu(x) - \partial_\mu \theta(x)$ with θ an arbitrary function, but before the introduction of quantum mechanics, this aspect of the symmetry was not emphasized.

²This strange terminology is due to Hermann Weyl. In 1918 he attempted to enlarge diffeomorphisms to local scale transformations and he called them, correctly, *gauge transformations*. The attempt was unsuccessful but, when he developed the theory for the Dirac electron in 1929, he still used the term gauge invariance, a term which has survived ever since, although the theory is no longer scale invariant.

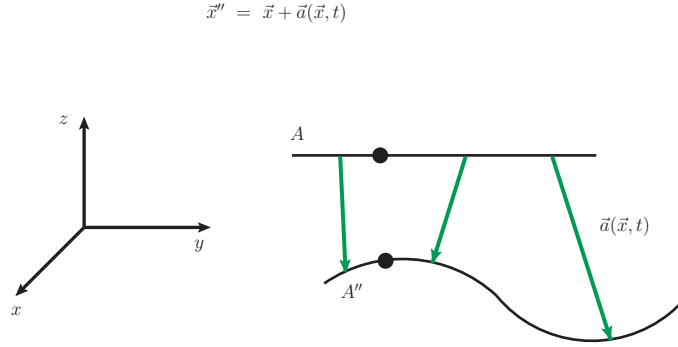


Fig. 2: A space translation by a vector $\vec{a}(x)$

of the origin of the coordinate system, since the latter is an unobservable quantity. From the mathematical point of view, local transformations produce a much richer and more interesting structure. Whichever one's motivations may be, physical or mathematical, it is clear that the free-particle dynamics is not invariant under translations in which \vec{a} is replaced by $\vec{a}(x)$. This is shown schematically in Fig. 2.

We see that no free particle would follow the trajectory A'' . This means that for A'' to be a trajectory, the particle must be subject to external forces. Can we determine these forces? The question sounds purely geometrical without any obvious physical meaning, so we expect a mathematical answer with no interest for physics. The great surprise is that the resulting theory, which is invariant under local translations, turns out to be classical general relativity, one of the four fundamental forces in nature. Gravitational interactions have such a geometric origin. In fact, the mathematical formulation of Einstein's original motivation to extend the principle of equivalence to accelerated frames is precisely the requirement of local invariance. Historically, many mathematical techniques which are used in today's gauge theories were developed in the framework of general relativity.

The gravitational forces are not the only ones that have a geometrical origin. Let us come back to the example of the quantum mechanical phase. It is clear that neither the Dirac nor the Schrödinger equation are invariant under a local change of phase $\theta(x)$. To be precise, let us consider the free Dirac Lagrangian,

$$\mathcal{L} = \bar{\Psi}(x)(i\cancel{\partial} - m)\Psi(x) . \quad (8)$$

It is not invariant under the transformation

$$\Psi(x) \rightarrow e^{i\theta(x)}\Psi(x) . \quad (9)$$

The reason behind this is the presence of the derivative term in Eq. (8) which gives rise to a term proportional to $\partial_\mu\theta(x)$. In order to restore invariance, one must modify Eq. (8), in which case it will no longer describe a free Dirac field; invariance under gauge transformations leads to the introduction of interactions. Both physicists and mathematicians know the answer to the particular case of Eq. (8): one introduces a new field $A_\mu(x)$ and replaces the derivative operator ∂_μ by a 'covariant derivative' D_μ given by

$$D_\mu = \partial_\mu + ieA_\mu , \quad (10)$$

where e is an arbitrary real constant. D_μ is said to be 'covariant' because it satisfies

$$D_\mu[e^{i\theta(x)}\Psi(x)] = e^{i\theta(x)}D_\mu\Psi(x) , \quad (11)$$

valid if, at the same time, $A_\mu(x)$ undergoes the transformation

$$A_\mu(x) \rightarrow A_\mu(x) - \frac{1}{e} \partial_\mu \theta(x). \quad (12)$$

The Dirac Lagrangian density now becomes

$$\mathcal{L} = \bar{\Psi}(x)(i\mathcal{D} - m)\Psi(x) = \bar{\Psi}(x)(i\cancel{\partial} - e\mathcal{A} - m)\Psi(x). \quad (13)$$

It is invariant under the gauge transformations of Eqs. (9) and (12) and describes the interaction of a charged spinor field with an external electromagnetic field! Replacing the derivative operator by the covariant derivative turns the Dirac equation into the same equation in the presence of an external electromagnetic field. Electromagnetic interactions give the same geometrical interpretation³. We can complete the picture by including the degrees of freedom of the electromagnetic field itself and add to Eq. (13) the corresponding Lagrangian density. Again, gauge invariance determines its form uniquely and we are led to the well-known result

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) + \bar{\Psi}(x)(i\mathcal{D} - m)\Psi(x) \quad (14)$$

with

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x). \quad (15)$$

The constant e we introduced is the electric charge, the coupling strength of the field Ψ with the electromagnetic field. Notice that a second field Ψ' will be coupled with its own charge e' .

Let us summarize: we started with a theory invariant under a group $U(1)$ of global phase transformations. The extension to a local invariance can be interpreted as a $U(1)$ symmetry at each point x . In a qualitative way we can say that gauge invariance induces an invariance under $U(1)^\infty$. We saw that this extension, a purely geometrical requirement, implies the introduction of new interactions. The surprising result here is that these ‘geometrical’ interactions describe the well-known electromagnetic forces.

The extension of the formalism of gauge theories to non-Abelian groups is not trivial and was first discovered by trial and error. Here we shall restrict ourselves to internal symmetries which are simpler to analyse and they are the ones we shall apply to particle physics outside gravitation.

Let us consider a classical field theory given by a Lagrangian density \mathcal{L} . It depends on a set of N fields $\psi^i(x)$, $i = 1, \dots, r$, and their first derivatives. The Lorentz transformation properties of these fields will play no role in this discussion. We assume that the ψ transform linearly according to an r -dimensional representation, not necessarily irreducible, of a compact, simple Lie group, G , which does not act on the space-time point x .

$$\Psi = \begin{pmatrix} \psi^1 \\ \vdots \\ \psi^r \end{pmatrix}, \quad \Psi(x) \rightarrow U(\omega)\Psi(x), \quad \omega \in G, \quad (16)$$

where $U(\omega)$ is the matrix of the representation of G . In fact, in these lectures we shall be dealing only with perturbation theory and it will be sufficient to look at transformations close to the identity in G .

$$\Psi(x) \rightarrow e^{i\Theta}\Psi(x), \quad \Theta = \sum_{a=1}^m \theta^a T^a \quad (17)$$

³The same applies to the Schrödinger equation. In fact, this was done first by V. Fock in 1926, immediately after Schrödinger’s original publication.

where the θ^a are a set of m constant parameters, and the T^a are $m \times r \times r$ matrices representing the m generators of the Lie algebra of G . They satisfy the commutation rules

$$[T^a, T^b] = if^{abc}T^c . \quad (18)$$

The f are the structure constants of G and a summation over repeated indices is understood. The normalization of the structure constants is usually fixed by requiring that, in the fundamental representation, the corresponding matrices of the generators t^a are normalized such as

$$\text{Tr}(t^a t^b) = \frac{1}{2} \delta^{ab} . \quad (19)$$

The Lagrangian density $\mathcal{L}(\Psi, \partial\Psi)$ is assumed to be invariant under the global transformations of Eqs. (16) or (17). As was done for the Abelian case, we wish to find a new \mathcal{L} , invariant under the corresponding gauge transformations in which the θ^a of Eq. (17) are arbitrary functions of x . In the same qualitative sense, we look for a theory invariant under G^∞ . This problem, stated the way we present it here, was first solved by trial and error for the case of SU(2) by C.N. Yang and R.L. Mills in 1954. They gave the underlying physical motivation and these theories are called since ‘Yang–Mills theories’. The steps are direct generalizations of the ones followed in the Abelian case. We need a gauge field, the analogue of the electromagnetic field, to transport the information contained in Eq. (17) from point to point. Since we can perform m independent transformations, the number of generators in the Lie algebra of G , we need m gauge fields $A_\mu^a(x)$, $a = 1, \dots, m$. It is easy to show that they belong to the adjoint representation of G . Using the matrix representation of the generators we can cast $A_\mu^a(x)$ into an $r \times r$ matrix:

$$\mathcal{A}_\mu(x) = \sum_{a=1}^m A_\mu^a(x) T^a . \quad (20)$$

The covariant derivatives can now be constructed as

$$\mathcal{D}_\mu = \partial_\mu + ig\mathcal{A}_\mu , \quad (21)$$

with g as an arbitrary real constant. They satisfy

$$\mathcal{D}_\mu e^{i\Theta(x)} \Psi(x) = e^{i\Theta(x)} \mathcal{D}_\mu \Psi(x) , \quad (22)$$

provided the gauge fields transform as

$$\mathcal{A}_\mu(x) \rightarrow e^{i\Theta(x)} \mathcal{A}_\mu(x) e^{-i\Theta(x)} + \frac{i}{g} \left(\partial_\mu e^{i\Theta(x)} \right) e^{-i\Theta(x)} . \quad (23)$$

The Lagrangian density $\mathcal{L}(\Psi, \mathcal{D}\Psi)$ is invariant under the gauge transformations of Eqs. (17) and (23) with an x -dependent Θ , if $\mathcal{L}(\Psi, \partial\Psi)$ is invariant under the corresponding global ones of Eqs. (16) or (17). As with the electromagnetic field, we can include the degrees of freedom of the new gauge fields by adding to the Lagrangian density a gauge invariant kinetic term. It turns out that it is slightly more complicated than $F_{\mu\nu}$ of the Abelian case. Yang and Mills computed it for SU(2) but it is uniquely determined by geometry plus some obvious requirements, such as absence of higher-order derivatives. The result is given by

$$\mathcal{G}_{\mu\nu} = \partial_\mu \mathcal{A}_\nu - \partial_\nu \mathcal{A}_\mu - ig [\mathcal{A}_\mu, \mathcal{A}_\nu] . \quad (24)$$

The full gauge-invariant Lagrangian can now be written as

$$\mathcal{L}_{\text{inv}} = -\frac{1}{2} \text{Tr} \mathcal{G}_{\mu\nu} \mathcal{G}^{\mu\nu} + \mathcal{L}(\Psi, D\Psi). \quad (25)$$

By convention, in Eq. (24) the matrix \mathcal{A} is taken to be

$$\mathcal{A}_\mu = A_\mu^a t^a, \quad (26)$$

where we recall that the t^a are the matrices representing the generators in the fundamental representation. It is only with this convention that the kinetic term in Eq. (25) is correctly normalized. In terms of the component fields A_μ^a , $\mathcal{G}_{\mu\nu}$ reads

$$\mathcal{G}_{\mu\nu} = G_{\mu\nu}^a t^a, \quad G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c. \quad (27)$$

Under a gauge transformation $\mathcal{G}_{\mu\nu}$ transforms like a member of the adjoint representation:

$$\mathcal{G}_{\mu\nu}(x) \rightarrow e^{i\theta^a(x)t^a} \mathcal{G}_{\mu\nu}(x) e^{-i\theta^a(x)t^a}. \quad (28)$$

This completes the construction of the gauge invariant Lagrangian. We add some remarks below.

- As was the case with the electromagnetic field, the Lagrangian of Eq. (25) does not contain terms proportional to $A_\mu A^\mu$. This means that, under the usual quantization rules, the gauge fields describe massless particles.
- Since $\mathcal{G}_{\mu\nu}$ is not linear in the fields A_μ , the \mathcal{G}^2 term in Eq. (25), besides the usual kinetic term which is bilinear in the fields, contains tri-linear and quadri-linear terms. In perturbation theory, they will be treated as coupling terms whose strength is given by the coupling constant g . In other words, the non-Abelian gauge fields are self-coupled while the Abelian (photon) field is not. A Yang–Mills theory, containing only gauge fields, is still a dynamically rich quantum field theory, whereas a theory with the electromagnetic field alone is a trivial free theory.
- The same coupling constant g appears in the covariant derivative of the fields Ψ in Eq. (21). This simple consequence of gauge invariance has an important physical application: if we add another field Ψ' , its coupling strength with the gauge fields will still be given by the same constant g . Contrary to the Abelian case studied before, if electromagnetism is part of a non-Abelian simple group, gauge invariance implies charge quantization.
- The above analysis can be extended in a straightforward way to the case where the group G is the product of simple groups $G = G_1 \times \dots \times G_n$. The only difference is that one should introduce n coupling constants g_1, \dots, g_n , one for each simple factor. Charge quantization is still true inside each subgroup, but charges belonging to different factors are no longer related.
- The situation changes if one considers non semi-simple groups, where one or more of the factors G_i is Abelian. In this case, the associated coupling constants can be chosen different for each field and the corresponding Abelian charges are not quantized.

As we alluded to above, gauge theories have a deep geometrical meaning. In order to get a better understanding of this property without entering into complicated issues of differential geometry, it is instructive to consider a reformulation of the theory replacing the continuum of space–time with a four-dimensional Euclidean lattice. We can do that very easily. Let us consider, for simplicity, a lattice with hypercubic symmetry. The space–time point x_μ is replaced by

$$x_\mu \rightarrow n_\mu a, \quad (29)$$

where a is a constant length (the lattice spacing) and n_μ is a d -dimensional vector with components $n_\mu = (n_1, n_2, \dots, n_d)$ which take integer values $0 \leq n_\mu \leq N_\mu$. N_μ is the number of points of our lattice

in the direction μ . The total number of points, i.e. the volume of the system, is given by $V \sim \prod_{\mu=1}^d N_\mu$. The presence of a introduces an ultraviolet, or short distance, cut-off because all momenta are bounded from above by $2\pi/a$. The presence of N_μ introduces an infrared or large distance cut-off because the momenta are also bounded from below by $2\pi/Na$, where N is the maximum of N_μ . The infinite volume–continuum space is recovered at the double limit $a \rightarrow 0$ and $N_\mu \rightarrow \infty$.

The dictionary between quantities defined in the continuum and the corresponding ones on the lattice is easy to establish (we take the lattice spacing a equal to one):

- a field $\Psi(x) \Rightarrow \Psi_n$,
where the field Ψ is an r -component column vector as in Eq. (16);
- a local term such as $\bar{\Psi}(x)\Psi(x) \Rightarrow \bar{\Psi}_n\Psi_n$;
- a derivative $\partial_\mu\Psi(x) \Rightarrow (\Psi_n - \Psi_{n+\mu})$,
where $n + \mu$ should be understood as a unit vector joining the point n with its nearest neighbour in the direction μ ;
- the kinetic energy term⁴ $\bar{\Psi}(x)\partial_\mu\Psi(x) \Rightarrow \bar{\Psi}_n\Psi_n - \bar{\Psi}_n\Psi_{n+\mu}$.

We may be tempted to write similar expressions for the gauge fields, but we must be careful with the way gauge transformations act on the lattice. Let us repeat the steps we followed in the continuum. Under gauge transformations a field transforms as:

- gauge transformations $\Psi(x) \rightarrow e^{i\Theta(x)}\Psi(x) \Rightarrow \Psi_n \rightarrow e^{i\Theta_n}\Psi_n$,
so all local terms of the form $\bar{\Psi}_n\Psi_n$ remain invariant but the part of the kinetic energy which couples fields at neighbouring points does not;
- the kinetic energy $\bar{\Psi}_n\Psi_{n+\mu} \rightarrow \bar{\Psi}_n e^{-i\Theta_n} e^{i\Theta_{n+\mu}} \Psi_{n+\mu}$,
which shows that we recover the problem we had with the derivative operator in the continuum.

In order to restore invariance we must introduce a new field, which is an $r \times r$ matrix, and which has indices n and $n + \mu$. We denote it by $U_{n,n+\mu}$ and we shall impose on it the constraint $U_{n,n+\mu} = U_{n+\mu,n}^{-1}$. Under a gauge transformation, U transforms as

$$U_{n,n+\mu} \rightarrow e^{i\Theta_n} U_{n,n+\mu} e^{-i\Theta_{n+\mu}}. \quad (30)$$

With the help of this gauge field we write the kinetic-energy term with the covariant derivative on the lattice as:

$$\bar{\Psi}_n U_{n,n+\mu} \Psi_{n+\mu}, \quad (31)$$

which is invariant under gauge transformations.

U is an element of the gauge group but we can show that, at the continuum limit and for an infinitesimal transformation, it correctly reproduces A_μ , which belongs to the Lie algebra of the group. Notice that, contrary to the field Ψ , U does not live on a single lattice point, but it has two indices, n and $n + \mu$, in other words it lives on the oriented link joining the two neighbouring points. We see here that the mathematicians are right when they do not call the gauge field ‘a field’ but ‘a connection’.

In order to finish the story we want to obtain an expression for the kinetic energy of the gauge field, the analogue of $Tr\mathcal{G}_{\mu\nu}(x)\mathcal{G}^{\mu\nu}(x)$, on the lattice. As for the continuum, the guiding principle is gauge invariance. Let us consider two points on the lattice n and m . We shall call a path $p_{n,m}$ on the lattice a sequence of oriented links which continuously join the two points. Next, consider the product of the gauge fields U along all the links of the path $p_{n,m}$:

⁴We write here the expression for spinor fields which contain only first-order derivatives in the kinetic energy. The extension to scalar fields with second-order derivatives is obvious.

$$P^{(p)}(n, m) = \prod_p U_{n, n+\mu} \cdots U_{m-\nu, m} . \quad (32)$$

Using the transformation rule in Eq. (30), we see that $P^{(p)}(n, m)$ transforms as

$$P^{(p)}(n, m) \rightarrow e^{i\Theta_n} P^{(p)}(n, m) e^{-i\Theta_m} . \quad (33)$$

It follows that if we consider a closed path $c = p_{n,n}$, the quantity $\text{Tr } P^{(c)}$ is gauge invariant. The simplest closed path for a hypercubic lattice has four links and it is called a *plaquette*. The correct form of the Yang–Mills action on the lattice can be written in terms of the sum of $\text{Tr } P^{(c)}$ over all plaquettes.

4 Spontaneous symmetry breaking

Since gauge theories appear to predict the existence of massless gauge bosons, when they were first proposed they did not seem to have any direct application to particle physics outside electromagnetism. It is this handicap which plagued gauge theories for many years. In this section, we shall present a seemingly unrelated phenomenon that will turn out to provide the answer.

An infinite system may exhibit the phenomenon of phase transitions. It often implies a reduction in the symmetry of the ground state. A field theory is a system with an infinite number of degrees of freedom, so it is not surprising that field theories may also show the phenomenon of phase transitions. Let us consider the example of a field theory invariant under a set of transformations forming a group G . In many cases, we encounter at least two phases.

- *The unbroken or the Wigner phase*: the symmetry is manifest in the spectrum of the theory whose excitations form irreducible representations of the symmetry group. For a gauge theory, the vector gauge bosons are massless and belong to the adjoint representation. But we have good reason to believe that, for non-Abelian gauge theories, a strange phenomenon occurs in this phase: all physical states are singlets of the group. All non-singlet states, such as those corresponding to the gauge fields, are supposed to be *confined*, in the sense that they do not appear as physically realizable asymptotic states.
- *The spontaneously broken phase*: part of the symmetry is hidden from the spectrum. For a gauge theory, some of the gauge bosons become massive and appear as physical states.

It is this kind of phase transition that we want to study in this section.

4.1 An example from classical mechanics

A very simple example is provided by the problem of the bent rod. Let a cylindrical rod be charged as in Fig. 3. The problem is obviously symmetric under rotations around the z -axis. Let z measure the distance from the basis of the rod, and $X(z)$ and $Y(z)$ give the deviations, along the x and y directions respectively, of the axis of the rod at the point z from the symmetric position. For small deflections the equations of elasticity can be linearized and take the form

$$IE \frac{d^4 X}{dz^4} + F \frac{d^2 X}{dz^2} = 0 ; \quad IE \frac{d^4 Y}{dz^4} + F \frac{d^2 Y}{dz^2} = 0 . \quad (34)$$

where $I = \pi R^4/4$ is the moment of inertia of the rod and E is the Young modulus. It is obvious that the system shown in Eq. (34) always possesses a symmetric solution $X = Y = 0$. However, we can also look for asymmetric solutions of the general form $X = A + Bz + C \sin kz + D \cos kz$ with $k^2 = F/EI$, which satisfy the boundary conditions $X = X'' = 0$ at $z = 0$ and $z = l$. We find that such solutions

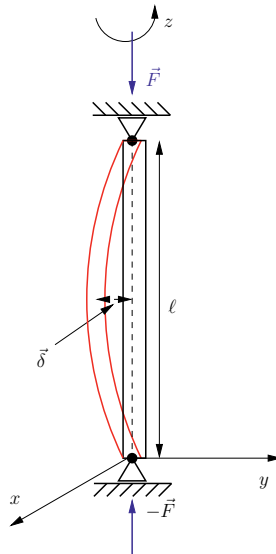


Fig. 3: A cylindrical rod bent under a force F along its symmetry axis

exist, $X = C \sin kz$, provided $kl = n\pi; n = 1, \dots$. The first such solution appears when F reaches a critical value F_{cr} given by

$$F_{cr} = \frac{\pi^2 EI}{l^2}. \quad (35)$$

The appearance of these solutions is already an indication of instability and a careful study of the stability problem proves that the non-symmetric solutions correspond to lower energy. From that point Eq. (34) is no longer valid because they only apply to small deflections, and we must use the general equations of elasticity. The result is that this instability of the symmetric solution occurs for all values of F larger than F_{cr} .

What has happened to the original symmetry of the equations? It is still hidden in the sense that we cannot predict in which direction the rod is going to bend in the x - y plane. They all correspond to solutions with precisely the same energy. In other words, if we apply a symmetry transformation (in this case a rotation around the z -axis) to an asymmetric solution, we obtain another asymmetric solution which is degenerate with the first one.

We call such a symmetry ‘spontaneously broken’, and in this simple example we see all its characteristics:

- there exists a critical point, i.e., a critical value of some external quantity which we can vary freely (in this case the external force F ; in several physical systems it is the temperature) which determines whether spontaneous symmetry breaking will take place or not. Beyond this critical point:
 - the symmetric solution becomes unstable;
 - the ground state becomes degenerate.

The complete mathematical analysis of this system requires the study of the exact equations of elasticity which are non-linear, but we can look at a simplified version. A quantity, which plays an important role in every phenomenon of phase transition, is *the order parameter*, whose value determines in which phase the system is. In our example, we choose it to be the two-component vector $\vec{\delta}$ shown in Fig. 3, which we write as a complex number $\delta = \delta_x + i\delta_y$ with $\delta = \rho e^{i\theta}$. The symmetric phase

corresponds to $\rho = 0$. It is instructive to express the energy of the system E as a function of the order parameter. Rotational invariance implies that E depends only on $\vec{\delta} \times \vec{\delta} = \rho^2$. At the vicinity of the critical point ρ^2 is small and we can expand E as

$$E = C_0 + C_1\rho^2 + C_2\rho^4 + \dots . \quad (36)$$

The C are constants which depend on the characteristics of the rod and the force F . Stability is obtained by

$$\frac{dE}{d\rho}(\rho = v) = 0 \Rightarrow v(C_1 + 2C_2v^2) = 0 . \quad (37)$$

We thus find the two solutions we mentioned above, namely $v = 0$ for the symmetric case and $v^2 = -C_1/2C_2$ for the spontaneously broken phase. Since ρ is real, this second solution is acceptable if C_1/C_2 is negative. C_2 must be positive for the energy to be bounded from below in the approximate Eq. (36). Therefore, C_1 must vanish at the critical point and change sign with $F - F_{\text{cr}}$. As a result, we can write $C_1 = \hat{C}_1(F_{\text{cr}} - F)$ with $\hat{C}_1 > 0$. For $F > F_{\text{cr}}$, C_1 is negative and we can write the energy as

$$E = C_0 + \hat{C}_1(F_{\text{cr}} - F)\vec{\delta} \times \vec{\delta} + C_2(\vec{\delta} \times \vec{\delta})^2 = \hat{C}_1(F - F_{\text{cr}})\frac{(\rho^2 - v^2)^2}{2v^2} , \quad (38)$$

with v given by the non-zero solution of Eq. (37). With the energy defined up to an arbitrary additive constant, we have fixed C_0 by the condition that the energy of the ground state $\rho = v$ vanishes. In the phase with spontaneous symmetry breaking, the energy of the symmetric $\rho = 0$ solution is positive and given by

$$E_0 = \hat{C}_1(F - F_{\text{cr}})\frac{v^2}{2} . \quad (39)$$

The expression for the energy given by Eq. (38) has the well-known form of Fig. 4 with a single minimum $v = 0$ for $F < F_{\text{cr}}$ and the Mexican hat form for $F > F_{\text{cr}}$.

There are a great variety of physical systems, both in classical and quantum physics, exhibiting spontaneous symmetry breaking, but we will not describe any others here. The Heisenberg ferro-magnet is a good example to keep in mind, because we shall often use it as a guide, but no essentially new phenomenon appears outside the ones already described. Therefore, we shall go directly to some field theory models.

4.2 A simple field theory model

Let $\phi(x)$ be a complex scalar field whose dynamics is described by the Lagrangian density

$$\mathcal{L}_1 = (\partial_\mu\phi)(\partial^\mu\phi^*) - M^2\phi\phi^* - \lambda(\phi\phi^*)^2 , \quad (40)$$

where \mathcal{L}_1 is a classical Lagrangian density and $\phi(x)$ is a classical field. No quantization is considered for the moment. Eq. (40) is invariant under the group U(1) of global transformations:

$$\phi(x) \rightarrow e^{i\theta}\phi(x) . \quad (41)$$

The current, $j_\mu \sim \phi\partial_\mu\phi^* - \phi^*\partial_\mu\phi$, whose conservation can be verified using the equations of motion, corresponds to this invariance.

We are interested in the classical field configuration which minimizes the energy of the system. We thus compute the Hamiltonian density given by

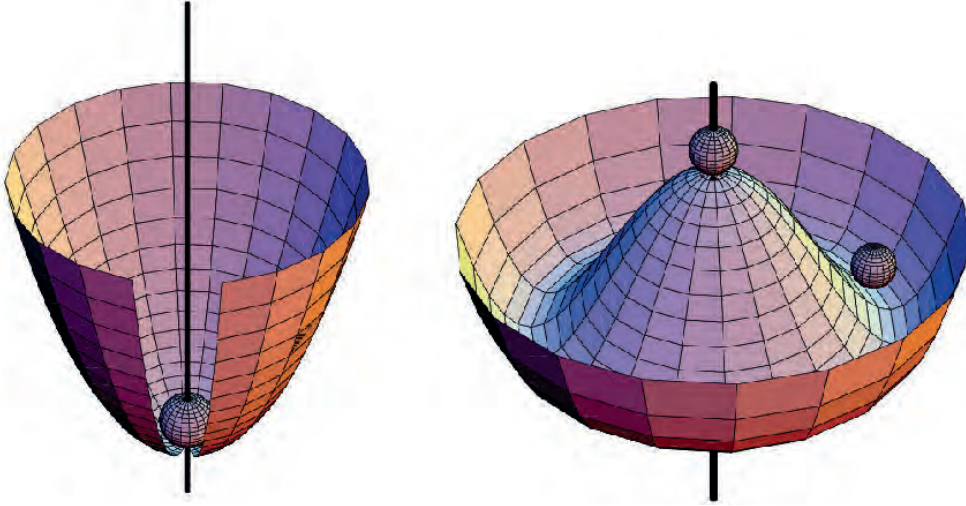


Fig. 4: The potential $V(\phi)$ with $M^2 \geq 0$ (left) and $M^2 < 0$ (right)

$$\mathcal{H}_1 = (\partial_0\phi)(\partial_0\phi^*) + (\partial_i\phi)(\partial_i\phi^*) + V(\phi) , \quad (42)$$

$$V(\phi) = M^2\phi\phi^* + \lambda(\phi\phi^*)^2 . \quad (43)$$

The first two terms of \mathcal{H}_1 are positive definite. They can only vanish for $\phi = \text{constant}$. Therefore, the ground state of the system corresponds to $\phi = \text{constant} = \text{minimum of } V(\phi)$. V has a minimum only if $\lambda > 0$. In this case, the position of the minimum depends on the sign of M^2 . (Notice that we are still studying a classical field theory and M^2 is just a parameter. One should not be misled by the notation into thinking that M is a ‘mass’ and M^2 is necessarily positive.)

For $M^2 > 0$, the minimum is at $\phi = 0$ (symmetric solution, shown in the left-hand side of Fig. 4), but for $M^2 < 0$ there is a whole circle of minima at the complex ϕ -plane with radius $v = (-M^2/2\lambda)^{1/2}$ (Fig. 4, right-hand side). Any point on the circle corresponds to a spontaneous breaking of Eq. (41).

We see that:

- the critical point is $M^2 = 0$;
- for $M^2 > 0$ the symmetric solution is stable;
- for $M^2 < 0$ spontaneous symmetry breaking occurs.

Let us choose $M^2 < 0$. In order to reach the stable solution we translate the field ϕ . It is clear that there is no loss of generality by choosing a particular point on the circle, since they are all obtained from any given one by applying the transformations from Eq. (41). Let us, for convenience, choose the point on the real axis in the ϕ -plane. We thus write

$$\phi(x) = \frac{1}{\sqrt{2}} [v + \psi(x) + i\chi(x)] . \quad (44)$$

Bringing (44) in (40) we find

$$\begin{aligned} \mathcal{L}_1(\phi) \rightarrow \mathcal{L}_2(\psi, \chi) &= \frac{1}{2}(\partial_\mu\psi)^2 + \frac{1}{2}(\partial_\mu\chi)^2 - \frac{1}{2}(2\lambda v^2)\psi^2 \\ &\quad - \lambda v\psi(\psi^2 + \chi^2) - \frac{\lambda}{4}(\psi^2 + \chi^2)^2 . \end{aligned} \quad (45)$$

Notice that \mathcal{L}_2 does not contain any term proportional to χ^2 , which is expected since V is locally flat in the χ direction. A second remark concerns the arbitrary parameters of the theory. \mathcal{L}_1 contains two such parameters: M , which has the dimensions of a mass, and λ , a dimensionless coupling constant. In \mathcal{L}_2 we again have the coupling constant λ and a new mass parameter v which is a function of M and λ . It is important to notice that, although \mathcal{L}_2 also contains trilinear terms, its coupling strength is not a new parameter but is proportional to $v\lambda$. \mathcal{L}_2 is still invariant under the transformations with infinitesimal parameter θ :

$$\delta\psi = -\theta\chi; \quad \delta\chi = \theta\psi + \theta v, \quad (46)$$

to which corresponds a conserved current

$$j_\mu \sim \psi\partial_\mu\chi - \chi\partial_\mu\psi + v\partial_\mu\chi. \quad (47)$$

The last term, which is linear in the derivative of χ , is characteristic of the phenomenon of spontaneous symmetry breaking.

It should be emphasized here that \mathcal{L}_1 and \mathcal{L}_2 are completely equivalent Lagrangians. They both describe the dynamics of the same physical system and a change of variables, as in Eq. (44), cannot change the physics. However, this equivalence is only true if we can solve the problem exactly. In this case, we shall find the same solution using either of them. However, we do not have exact solutions and we intend to apply perturbation theory, which is an approximation scheme. The equivalence is then no longer guaranteed and perturbation theory has much better chances to give sensible results using one language rather than the other. In particular, if we use \mathcal{L}_1 as a quantum field theory and we decide to apply perturbation theory, using the quadratic terms of \mathcal{L}_1 as the unperturbed part, we immediately see that we shall get nonsense. The spectrum of the unperturbed Hamiltonian would consist of particles with negative square mass, and no perturbation corrections at any finite order could change that. This is essentially because we are trying to calculate the quantum fluctuations around an unstable solution and perturbation theory is just not designed to do that. On the contrary, we see that the quadratic part of \mathcal{L}_2 gives a reasonable spectrum; thus we hope that perturbation theory will also give reasonable results. Therefore, we conclude that our physical system, considered now as a quantum system, consists of two interacting scalar particles, one with mass $m_\psi^2 = 2\lambda v^2$ and the other with $m_\chi = 0$. We believe that this is the spectrum we would have also found starting from \mathcal{L}_1 , if we could solve the dynamics exactly.

The appearance of a zero-mass particle in the quantum version of the model is an example of a general theorem attributable to J. Goldstone: for every generator of a spontaneously broken symmetry there corresponds a massless particle, called the Goldstone particle. This theorem is just the translation of the statement about the degeneracy of the ground state into quantum-field-theory language. The ground state of a system described by a quantum field theory is the vacuum state, and you need massless excitations in the spectrum of states in order to allow for the degeneracy of the vacuum.

4.3 Gauge symmetries

In this section, we want to study the consequences of spontaneous symmetry breaking in the presence of a gauge symmetry. We shall find a very surprising result. When combined together, the two problems, namely the massless gauge bosons on the one hand and the massless Goldstone bosons on the other, will solve each other. It is this miracle that we want to present here⁵. We start with the Abelian case.

We look at the model of the previous section in which the U(1) symmetry of Eq. (41) has been promoted to a local symmetry with $\theta \rightarrow \theta(x)$. As we explained already, this implies the introduction of

⁵In relativistic physics this mechanism was invented and developed by François Englert and Robert Brout, Peter Higgs, as well as Gerald Guralnik, Carl Richard Hagen and Thomas Walter Bannerman Kibble.

a massless vector field, which we can call the ‘photon’ and the interactions are obtained by replacing the derivative operator ∂_μ by the covariant derivative D_μ and adding the photon kinetic energy term:

$$\mathcal{L}_1 = -\frac{1}{4}F_{\mu\nu}^2 + |(\partial_\mu + ieA_\mu)\phi|^2 - M^2\phi\phi^* - \lambda(\phi\phi^*)^2. \quad (48)$$

\mathcal{L}_1 is invariant under the gauge transformation:

$$\phi(x) \rightarrow e^{i\theta(x)}\phi(x); \quad A_\mu \rightarrow A_\mu - \frac{1}{e}\partial_\mu\theta(x). \quad (49)$$

The same analysis as before shows that for $\lambda > 0$ and $M^2 < 0$ there is a spontaneous breaking of the U(1) symmetry. Replacing Eq. (44) for (48) we obtain

$$\begin{aligned} \mathcal{L}_1 \rightarrow \mathcal{L}_2 = & -\frac{1}{4}F_{\mu\nu}^2 + \frac{e^2v^2}{2}A_\mu^2 + evA_\mu\partial^\mu\chi \\ & + \frac{1}{2}(\partial_\mu\psi)^2 + \frac{1}{2}(\partial_\mu\chi)^2 - \frac{1}{2}(2\lambda v^2)\psi^2 + \dots, \end{aligned} \quad (50)$$

where the dots stand for coupling terms which are at least trilinear in the fields.

The surprising term is the second one, which is proportional to A_μ^2 . It looks as though the photon has become massive. Notice that Eq. (50) is still gauge invariant since it is equivalent to Eq. (48). The gauge transformation is now obtained by replacing Eq. (44) with Eq. (49):

$$\begin{aligned} \psi(x) & \rightarrow \cos\theta(x)[\psi(x) + v] - \sin\theta(x)\chi(x) - v \\ \chi(x) & \rightarrow \cos\theta(x)\chi(x) + \sin\theta(x)[\psi(x) + v] \\ A_\mu & \rightarrow A_\mu - \frac{1}{e}\partial_\mu\theta(x). \end{aligned} \quad (51)$$

This means that our previous conclusion, that gauge invariance forbids the presence of an A_μ^2 term, was simply wrong. Such a term can be present, but the gauge transformation is slightly more complicated; it must be accompanied by a translation of the field.

The Lagrangian of Eq. (50), if taken as a quantum field theory, seems to describe the interaction of a massive vector particle (A_μ) and two scalars, one massive (ψ) and one massless (χ). However, we can immediately see that something is wrong with this counting. A warning is already contained in the non-diagonal term between A_μ and $\partial^\mu\chi$. Indeed, the perturbative particle spectrum can be read from the Lagrangian only after we have diagonalized the quadratic part. A more direct way to see the trouble is to count the apparent degrees of freedom⁶ before and after the translation:

- Lagrangian of Eq. (48):
 - (i) one massless vector field: 2 degrees;
 - (ii) one complex scalar field: 2 degrees;
 - total: 4 degrees.

⁶The terminology here is misleading. As we pointed out earlier, any field theory, considered as a dynamical system, is a system with an infinite number of degrees of freedom. For example, the quantum theory of a free neutral scalar field is described by an infinite number of harmonic oscillators, one for every value of the three-dimensional momentum. Here, we use the same term ‘degrees of freedom’ to denote the independent one-particle states. We know that for a massive spin- s particle we have $2s + 1$ one-particle states, and for a massless particle with spin other than zero we only have two. In fact, it would have been more appropriate to talk about a $(2s + 1)$ -infinity and 2-infinity degrees of freedom, respectively.

- Lagrangian of Eq. (50):
 - (i) one massive vector field: 3 degrees;
 - (ii) two real scalar fields: 2 degrees;
 - total: 5 degrees.

Since physical degrees of freedom cannot be created by a simple change of variables, we conclude that the Lagrangian of Eq. (50) must contain fields which do not create physical particles. This is indeed the case, and we can show a transformation which makes the unphysical fields disappear. Instead of parametrizing the complex field ϕ by its real and imaginary parts, let us choose its modulus and its phase. The choice is dictated by the fact that it is a change of phase that describes the motion along the circle of the minima of the potential $V(\phi)$. We thus write

$$\phi(x) = \frac{1}{\sqrt{2}}[v + \rho(x)]e^{i\zeta(x)/v}; \quad A_\mu(x) = B_\mu(x) - \frac{1}{ev}\partial_\mu\zeta(x). \quad (52)$$

In this notation, the gauge transformation Eq. (49) or Eq. (51) is simply a translation of the field ζ : $\zeta(x) \rightarrow \zeta(x) + v\theta(x)$. Replacing Eq. (52) with Eq. (48) we obtain

$$\begin{aligned} \mathcal{L}_1 \rightarrow \mathcal{L}_3 = & -\frac{1}{4}B_{\mu\nu}^2 + \frac{e^2v^2}{2}B_\mu^2 + \frac{1}{2}(\partial_\mu\rho)^2 - \frac{1}{2}(2\lambda v^2)\rho^2 \\ & - \frac{\lambda}{4}\rho^4 + \frac{1}{2}e^2B_\mu^2(2v\rho + \rho^2) \\ & B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \end{aligned} \quad (53)$$

The $\zeta(x)$ field has disappeared. Equation (53) describes two massive particles, a vector (B_μ) and a scalar (ρ). It exhibits no gauge invariance, since the original symmetry $\zeta(x) \rightarrow \zeta(x) + v\theta(x)$ is now trivial.

We see that there are three different Lagrangians describing the same physical system. \mathcal{L}_1 is invariant under the usual gauge transformation, but it contains a negative square mass and it is therefore unsuitable for quantization. \mathcal{L}_2 is still gauge invariant, but the transformation law from Eq. (51) is more complicated. It can be quantized in a space containing unphysical degrees of freedom. This by itself is not a great obstacle and it occurs frequently. For example, ordinary QED is usually quantized in a space involving unphysical (longitudinal and scalar) photons. In fact, it is \mathcal{L}_2 , in a suitable gauge, which is used for general proofs of renormalizability as well as for practical calculations. Finally, \mathcal{L}_3 is no longer invariant under any kind of gauge transformation, but clearly exhibits the particle spectrum of the theory. It contains only physical particles and they are all massive. This is the miracle that was announced earlier. Although we start from a gauge theory, the final spectrum contains massive particles only. Actually, \mathcal{L}_3 can be obtained from \mathcal{L}_2 by an appropriate choice of gauge. The conclusion so far can be stated as follows.

In a spontaneously broken gauge theory, the gauge vector bosons acquire a mass and the would-be massless Goldstone bosons decouple and disappear. Their degrees of freedom are used to make the transition from massless to massive vector bosons possible.

The extension to the non-Abelian case is straightforward. Let us consider a gauge group G with m generators and, thus, m massless gauge bosons. The claim is that we can break part of the symmetry spontaneously, leaving a subgroup H with h generators unbroken. The h gauge bosons associated with H remain massless while the $m - h$ others acquire a mass. In order to achieve this result we need $m - h$ scalar degrees of freedom with the same quantum numbers as the broken generators. They will disappear from the physical spectrum and will re-appear as zero-helicity states of the massive vector bosons. As previously, we shall see that one needs at least one more scalar state which remains physical.

In the remaining part of this section, we show these results for a general gauge group. The reader who is not interested in technical details may skip this part.

We introduce a multiplet of scalar fields ϕ_i which transform according to some representation, not necessarily irreducible, of G of dimension n . According to the rules we explained in the last section, the Lagrangian of the system is given by

$$\mathcal{L} = -\frac{1}{4}\text{Tr}(G_{\mu\nu}G^{\mu\nu}) + (D_\mu\Phi)^\dagger D^\mu\Phi - V(\Phi). \quad (54)$$

In component notation, the covariant derivative is, as usual, $D_\mu\phi_i = \partial_\mu\phi_i - ig^{(a)}T_{ij}^a A_\mu^a\phi_j$ where we have allowed for the possibility of having arbitrary coupling constants $g^{(a)}$ for the various generators of G because we do not assume that G is simple or semi-simple. $V(\Phi)$ is a polynomial in the Φ invariant under G of degree equal to four. As before, we assume that we can choose the parameters in V such that the minimum is not at $\Phi = 0$ but rather at $\Phi = v$ where v is a constant vector in the representation space of Φ . v is not unique. The m generators of G can be separated into two classes: h generators which annihilate v and form the Lie algebra of the unbroken subgroup H ; and $m - h$ generators, shown in the representation of Φ by matrices T^a , such that $T^a v \neq 0$ and all vectors $T^a v$ are independent and can be chosen to be orthogonal. Any vector in the orbit of v , i.e. of the form $e^{iw^a T^a} v$, is an equivalent minimum of the potential. As before, we should translate the scalar fields Φ by $\Phi \rightarrow \Phi + v$. It is convenient to decompose Φ into components along the orbit of v and orthogonal to it, the analogue of the χ and ψ fields of the previous section. We can write

$$\Phi = i \sum_{a=1}^{m-h} \frac{\chi^a T^a v}{|T^a v|} + \sum_{b=1}^{n-m+h} \psi^b u^b + v, \quad (55)$$

where the vectors u^b form an orthonormal basis in the space orthogonal to all $T^a v$. The corresponding generators span the coset space G/H . As before, we shall show that the fields χ^a will be absorbed by the Brout–Englert–Higgs mechanism and the fields ψ^b will remain physical. Note that the set of vectors u^b contains at least one element since, for all a , we have

$$v \times T^a v = 0 \quad (56)$$

because the generators in a real unitary representation are anti-symmetric. This shows that the dimension n of the representation of Φ must be larger than $m - h$ and, therefore, there will remain at least one physical scalar field which, in the quantum theory, will give a physical scalar particle⁷.

Let us now bring in the Lagrangian from Eq. (54) the expression of Φ from Eq. (55). We obtain

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \sum_{a=1}^{m-h} (\partial_\mu \chi^a)^2 + \frac{1}{2} \sum_{b=1}^{n-m+h} (\partial_\mu \psi^b)^2 - \frac{1}{4} \text{Tr}(F_{\mu\nu} F^{\mu\nu}) \\ & + \frac{1}{2} \sum_{a=1}^{m-h} g^{(a)2} |T^a v|^2 A_\mu^a A^{\mu a} - \sum_{a=1}^{m-h} g^{(a)} T^a v \partial^\mu \chi^a A_\mu^a - V(\Phi) + \dots, \quad (57) \end{aligned}$$

where the dots stand for coupling terms between the scalars and the gauge fields. In writing Eq. (57) we took into account that $T^b v = 0$ for $b > m - h$ and that the vectors $T^a v$ are orthogonal.

⁷Obviously, the argument assumes the existence of scalar fields which induce the phenomenon of spontaneous symmetry breaking. We can construct models in which the role of the latter is played by some kind of fermion–anti-fermion bound states and they come under the name of models with a *dynamical symmetry breaking*. In such models the existence of a physical spin-zero state, the analogue of the σ -particle of the chiral symmetry breaking of quantum chromodynamics (QCD), is a dynamical question and in general hard to answer.

The analysis that gave us Goldstone's theorem shows that

$$\frac{\partial^2 V}{\partial \phi_k \partial \phi_l} \Big|_{\Phi=v} (T^a v)_l = 0, \quad (58)$$

which shows that the χ -fields would correspond to the Goldstone modes. As a result, the only mass terms which appear in V in Eq. (57) are of the form $\psi^k M^{kl} \psi^l$ and do not involve the χ -fields.

As far as the bilinear terms in the fields are concerned, the Lagrangian from Eq. (57) is the sum of terms of the form found in the Abelian case. All gauge bosons which do not correspond to H generators acquire a mass equal to $m_a = g^{(a)} |T^a v|$ and, through their mixing with the would-be Goldstone fields χ , develop a zero-helicity state. All other gauge bosons remain massless. The ψ represent the remaining physical Higgs fields.

5 Building the Standard Model: a five-step programme

In this section we shall construct the Standard Model of electro-weak interactions as a spontaneously broken gauge theory. We shall follow the hints given by experiment following a five-step programme.

- Step 1: Choose a gauge group G .
- Step 2: Choose the fields of the 'elementary' particles and assign them to representations of G . Include scalar fields to allow for the Brout–Englert–Higgs mechanism.
- Step 3: Write the most general renormalizable Lagrangian invariant under G . At this stage, gauge invariance is still exact and all gauge vector bosons are massless.
- Step 4: Choose the parameters of the scalar potential so that spontaneous symmetry breaking occurs.
- Step 5: Translate the scalars and rewrite the Lagrangian in terms of the translated fields. Choose a suitable gauge and quantize the theory.

Note that gauge theories provide only the general framework, not a detailed model. The latter will depend on the particular choices made in Steps 1 and 2.

5.1 The lepton world

We start with the leptons and, in order to simplify the presentation, we shall assume that neutrinos are massless. We follow the five steps.

Step 1: Looking at the table of elementary particles we see that, for the combined electromagnetic and weak interactions, we have four gauge bosons, namely W^\pm , Z^0 and the photon. As we explained earlier, each one of them corresponds to a generator of the group G , more precisely its Lie algebra. The only non-trivial algebra with four generators is that of $U(2) \approx SU(2) \times U(1)$.

Following the notation which was inspired by the hadronic physics, we call T_i , $i = 1, 2, 3$, the three generators of $SU(2)$ and Y that of $U(1)$. Then, the electric charge operator Q will be a linear combination of T_3 and Y . By convention, we write

$$Q = T_3 + \frac{1}{2} Y. \quad (59)$$

The coefficient in front of Y is arbitrary and only fixes the normalization of the $U(1)$ generator relatively to those of $SU(2)$ ⁸. This ends our discussion of the first step.

⁸The normalization of the generators for non-Abelian groups is fixed by their commutation relations. That of the Abelian generator is arbitrary. The relation of Eq. (59) is one choice which has only a historical value. It is not the most natural one from the group theory point of view, as you will see in the discussion concerning Grand-Unified theories.

Step 2: The number and the interaction properties of the gauge bosons are fixed by the gauge group. This is no longer the case with the fields describing the other particles. In principle, we can choose any number and assign them to any representation. It follows that the choice here will be dictated by the phenomenology.

Leptons have always been considered as elementary particles. We have six leptons but, as we noted already, a striking feature of the data is the phenomenon of family repetition. We do not understand why nature chooses to repeat itself three times, but the simplest way to incorporate this observation into the model is to use the same representations three times, one for each family. This leaves SU(2) doublets and/or singlets as the only possible choices. A further experimental input we shall use is the fact that the charged W couple only to the left-handed components of the lepton fields, in contrast to the photon which couples with equal strength to both right and left. These considerations lead us to assign the left-handed components of the lepton fields to doublets of SU(2):

$$\Psi_L^i(x) = \frac{1}{2}(1 + \gamma_5) \begin{pmatrix} \nu_i(x) \\ \ell_i^-(x) \end{pmatrix}; \quad i = 1, 2, 3, \quad (60)$$

where we have used the same symbol for the particle and the associated Dirac field.

The right-handed components are assigned to singlets of SU(2):

$$\nu_{iR}(x) = \frac{1}{2}(1 - \gamma_5)\nu_i(x) \quad (?); \quad \ell_{iR}^-(x) = \frac{1}{2}(1 - \gamma_5)\ell_i^-(x). \quad (61)$$

The question mark next to the right-handed neutrinos means that the presence of these fields is not confirmed by the data. We shall drop them in this lecture, but we may come back to this point later. We shall also simplify the notation and put $\ell_{iR}^-(x) = R_i(x)$. The resulting transformation properties under local SU(2) transformations are

$$\Psi_L^i(x) \rightarrow e^{i\vec{\tau}\vec{\theta}(x)}\Psi_L^i(x); \quad R_i(x) \rightarrow R_i(x), \quad (62)$$

with $\vec{\tau}$ the three Pauli matrices. This assignment and the Y normalization given by Eq. (59), also fix the U(1) charge and, therefore, the transformation properties of the lepton fields. For all i we find

$$Y(\Psi_L^i) = -1; \quad Y(R_i) = -2. \quad (63)$$

If a right-handed neutrino exists, it has $Y(\nu_{iR}) = 0$, which shows that it is not coupled to any gauge boson.

We are left with the choice of the Higgs scalar fields and we shall choose the solution with the minimal number of fields. We must give masses to three vector gauge bosons and keep the fourth one massless. The latter will be identified with the photon. We recall that, for every vector boson acquiring mass, a scalar with the same quantum numbers decouples. At the end we shall remain with at least one physical, neutral, scalar field. It follows that the minimal number to start with is four, two charged and two neutral. We choose to put them, under SU(2), into a complex doublet:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}; \quad \Phi(x) \rightarrow e^{i\vec{\tau}\vec{\theta}(x)}\Phi(x), \quad (64)$$

with the conjugate fields ϕ^- and ϕ^{0*} forming Φ^\dagger . The U(1) charge of Φ is $Y(\Phi) = 1$.

This ends our choices for the second step. At this point the model is complete. All further steps are purely technical and uniquely defined.

Step 3: What follows is straightforward algebra. We write the most general, renormalizable, Lagrangian, involving the fields of Eqs. (60), (61) and (64) invariant under gauge transformations of

SU(2) \times U(1). We shall also assume the separate conservation of the three lepton numbers, leaving the discussion on the neutrino mixing to a specialized lecture. The requirement of renormalizability implies that all terms in the Lagrangian are monomials in the fields and their derivatives, and their canonical dimension is less than or equal to four. The result is

$$\begin{aligned} \mathcal{L} = & -\frac{1}{4}\vec{W}_{\mu\nu} \times \vec{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + |D_\mu\Phi|^2 - V(\Phi) \\ & + \sum_{i=1}^3 [\bar{\Psi}_L^i i\not{D}\Psi_L^i + \bar{R}_i i\not{D}R_i - G_i(\bar{\Psi}_L^i R_i \Phi + \text{h.c.})] . \end{aligned} \quad (65)$$

If we call \vec{W} and B the gauge fields associated with SU(2) and U(1) respectively, the corresponding field strengths $\vec{W}_{\mu\nu}$ and $B_{\mu\nu}$ appearing in Eq. (65) are given by Eqs. (24) and (15).

Similarly, the covariant derivatives in Eq. (65) are determined by the assumed transformation properties of the fields, as shown in Eq. (21):

$$\begin{aligned} D_\mu\Psi_L^i &= \left(\partial_\mu - ig\frac{\vec{\tau}}{2} \times \vec{W}_\mu + ig'B_\mu\right)\Psi_L^i ; \quad D_\mu R_i = (\partial_\mu + ig'B_\mu)R_i, \\ D_\mu\Phi &= \left(\partial_\mu - ig\frac{\vec{\tau}}{2} \times \vec{W}_\mu - ig'B_\mu\right)\Phi . \end{aligned} \quad (66)$$

The two coupling constants g and g' correspond to the groups SU(2) and U(1), respectively. The most general potential $V(\Phi)$ compatible with the transformation properties of the field Φ is

$$V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2 . \quad (67)$$

The last term in Eq. (65) is a Yukawa coupling term between the scalar Φ and the fermions. In the absence of right-handed neutrinos, this is the most general term which is invariant under SU(2) \times U(1). As usual, h.c. stands for ‘hermitian conjugate’. G_i are three arbitrary coupling constants. If right-handed neutrinos exist there is a second Yukawa term with R_i replaced by ν_{iR} and Φ by the corresponding doublet proportional to $\tau_2\Phi^*$, where * means ‘complex conjugation’. We see that the Standard Model can perfectly well accommodate a right-handed neutrino, but it couples only to the Higgs field.

A final remark: as expected, the gauge bosons \vec{W}_μ and B_μ appear to be massless. The same is true for all fermions. This is not surprising because the assumed different transformation properties of the right- and left-handed components forbid the appearance of a Dirac mass term in the Lagrangian. On the other hand, the Standard Model quantum numbers also forbid the appearance of a Majorana mass term for the neutrinos. In fact, the only dimensionful parameter in (65) is μ^2 , the parameter in the Higgs potential in Eq. (67). Therefore, the mass of every particle in the model is expected to be proportional to $|\mu|$.

Step 4: The next step of our programme consists of choosing the parameter μ^2 negative to trigger the phenomenon of spontaneous symmetry breaking and the Brout–Englert–Higgs mechanism. The minimum potential occurs at a point $v^2 = -\mu^2/\lambda$. As we have explained earlier, we can choose the direction of the breaking to be along the real part of ϕ^0 .

Step 5: Translating the scalar field by a real constant,

$$\Phi \rightarrow \Phi + \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v^2 = -\frac{\mu^2}{\lambda}, \quad (68)$$

transforms the Lagrangian and generates new terms, as it was explained in the previous section. Let us look at some of them.

(i) *Fermion mass terms.* Replacing ϕ^0 by v in the Yukawa term in Eq. (65) creates a mass term for the charged leptons, leaving the neutrinos massless:

$$m_e = \frac{1}{\sqrt{2}}G_e v, \quad m_\mu = \frac{1}{\sqrt{2}}G_\mu v, \quad m_\tau = \frac{1}{\sqrt{2}}G_\tau v. \quad (69)$$

Since we have three arbitrary constants G_i , we can fit the three observed lepton masses. If we introduce right-handed neutrinos we can also apply whichever Dirac neutrino masses we wish.

(ii) *Gauge-boson mass terms.* They come from the $|D_\mu\Phi|^2$ term in the Lagrangian. A straight substitution produces the following quadratic terms among the gauge boson fields:

$$\frac{1}{8}v^2[g^2(W_\mu^1W^{1\mu} + W_\mu^2W^{2\mu}) + (g'B_\mu - gW_\mu^3)^2]. \quad (70)$$

Defining the charged vector bosons as

$$W_\mu^\pm = \frac{W_\mu^1 \mp iW_\mu^2}{\sqrt{2}}, \quad (71)$$

we obtain their masses,

$$m_W = \frac{vg}{2}. \quad (72)$$

The neutral gauge bosons B_μ and W_μ^3 have a 2×2 non-diagonal mass matrix. After diagonalization, we define the mass eigenstates as

$$\begin{aligned} Z_\mu &= \cos\theta_W B_\mu - \sin\theta_W W_\mu^3 \\ A_\mu &= \cos\theta_W W_\mu^3 + \sin\theta_W B_\mu, \end{aligned} \quad (73)$$

with $\tan\theta_W = g'/g$. They correspond to the mass eigenvalues

$$\begin{aligned} m_Z &= \frac{v(g^2 + g'^2)^{1/2}}{2} = \frac{m_W}{\cos\theta_W} \\ m_A &= 0. \end{aligned} \quad (74)$$

As expected, one of the neutral gauge bosons is massless and will be identified with the photon. The Brout–Englert–Higgs mechanism breaks the original symmetry according to $SU(2) \times U(1) \rightarrow U(1)_{em}$ and θ_W is the angle between the original $U(1)$ and the one left unbroken. It is the parameter first introduced by S.L. Glashow, although it is often referred to as the ‘Weinberg angle’.

(iii) *Physical Higgs mass.* Three out of the four real fields of the Φ doublet will be absorbed in order to allow for the three gauge bosons W^\pm and Z^0 to acquire a mass. The fourth one, which corresponds to $(|\phi^0\phi^{0\dagger}|)^{1/2}$, remains physical. Its mass is given by the coefficient of the quadratic part of $V(\Phi)$ after the translation of Eq. (68) and is equal to

$$m_h = \sqrt{-2\mu^2} = \sqrt{2\lambda v^2}. \quad (75)$$

In addition, we produce various coupling terms which we shall present, together with the hadronic ones, in the next section.

5.2 Extension to hadrons

Introducing the hadrons into the model presents some novel features largely because the individual quark quantum numbers are not separately conserved. With regard to the second step, there is currently a consensus regarding the choice of the ‘elementary’ constituents of matter: besides the six leptons, there are six quarks. They are fractionally charged and come each in three ‘colours’. The observed lepton–hadron universality property tells us to also use doublets and singlets for the quarks. The first novel feature we mentioned above is that all quarks appear to have non-vanishing Dirac masses, so we must introduce both right-handed singlets for each family. A naïve assignment would be to write the analogue of Eqs. (60) and (61) as

$$Q_L^i(x) = \frac{1}{2}(1 + \gamma_5) \begin{pmatrix} U^i(x) \\ D^i(x) \end{pmatrix}; \quad U_R^i(x); \quad D_R^i(x), \quad (76)$$

with the index i running over the three families as $U^i = \text{u,c,t}$ and $D^i = \text{d,s,b}$ for $i = 1, 2, 3$, respectively⁹. This assignment determines the SU(2) transformation properties of the quark fields. It also fixes their Y charges and, hence their U(1) properties. Using Eq. (59), we find

$$Y(Q_L^i) = \frac{1}{3}; \quad Y(U_R^i) = \frac{4}{3}; \quad Y(D_R^i) = -\frac{2}{3}. \quad (77)$$

The presence of the two right-handed singlets has an important consequence. Even if we had only one family, we would have two distinct Yukawa terms between the quarks and the scalar field of the form

$$\mathcal{L}_{\text{Yuk}} = G_d(\bar{Q}_L D_R \Phi + \text{h.c.}) + G_u(\bar{Q}_L U_R \tilde{\Phi} + \text{h.c.}). \quad (78)$$

$\tilde{\Phi}$ is the doublet proportional to $\tau_2 \Phi^*$. It has the same transformation properties under SU(2) as Φ , but the opposite Y charge.

If there were only one family, this would have been the end of the story. The hadron Lagrangian $\mathcal{L}_h^{(1)}$ is the same as Eq. (65) with quark fields replacing leptons and the extra term of Eq. (78). The complication we alluded to before comes with the addition of more families. In this case the total Lagrangian is not just the sum over the family index. The physical reason is the non-conservation of the individual quark quantum numbers we mentioned previously. In writing Eq. (76), we implicitly assumed a particular pairing of the quarks in each family; u with d, c with s and t with b. In general, we could choose any basis in family space and, since we have two Yukawa terms, we will not be able to diagonalize both of them simultaneously. It follows that the most general Lagrangian will contain a matrix with non-diagonal terms which mix the families. By convention, we attribute it to a different choice of basis in the d–s–b space. It follows that the correct generalization of the Yukawa Lagrangian of Eq. (78) to many families is given by

$$\mathcal{L}_{\text{Yuk}} = \sum_{i,j} \left[(\bar{Q}_L^i G_d^{ij} D_R^j \Phi + \text{h.c.}) \right] + \sum_i \left[G_u^i (\bar{Q}_L^i U_R^i \tilde{\Phi} + \text{h.c.}) \right], \quad (79)$$

where the Yukawa coupling constant G_d has become a matrix in family space. After translation of the scalar field, we shall produce masses for the up quarks given by $m_u = G_u^1 v$, $m_c = G_u^2 v$ and $m_t = G_u^3 v$, as well as a 3×3 mass matrix for the down quarks given by $G_d^{ij} v$. As usual, we want to work in a field space where the masses are diagonal, so we change our initial d–s–b basis to bring G_d^{ij} into a diagonal form. This can be done through a 3×3 unitary matrix $\tilde{D}^i = U^{ij} D^j$ such that $U^\dagger G_d U = \text{diag}(m_d, m_s, m_b)$. In the simplest example of only two families, it is easy to show that the most general such matrix, after using all freedom for field redefinitions and phase choices, is a real rotation:

⁹An additional index a , also running through 1, 2 and 3 and denoting the colour, is understood.

$$C = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \quad (80)$$

with θ being our familiar Cabibbo angle. For three families, an easy counting shows that the matrix has three angles, the three Euler angles, and an arbitrary phase. It is traditionally written in the form

$$KM = \begin{pmatrix} c_1 & s_1 c_3 & s_1 s_3 \\ -s_1 c_3 & c_1 c_2 c_3 - s_2 s_3 e^{i\delta} & c_1 c_2 s_3 + s_2 c_3 e^{i\delta} \\ -s_1 s_2 & c_1 s_2 c_3 + c_2 s_3 e^{i\delta} & c_1 s_2 s_3 - c_2 c_3 e^{i\delta} \end{pmatrix}, \quad (81)$$

with the notation $c_k = \cos \theta_k$ and $s_k = \sin \theta_k$, $k = 1, 2, 3$. The novel feature is the possibility of introducing the phase δ . This means that a six-quark model has a natural source of CP or T violation, whereas a four-quark model does not.

The total Lagrangian density, before the translation of the field Φ , is now

$$\begin{aligned} \mathcal{L} = & -\frac{1}{4} \vec{W}_{\mu\nu} \times \vec{W}^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + |D_\mu \Phi|^2 - V(\Phi) \\ & + \sum_{i=1}^3 [\bar{\Psi}_L^i i \not{D} \Psi_L^i + \bar{R}_i i \not{D} R_i - G_i (\bar{\Psi}_L^i R_i \Phi + \text{h.c.})] \\ & + [\bar{Q}_L^i i \not{D} Q_L^i + \bar{U}_R^i i \not{D} U_R^i + \bar{D}_R^i i \not{D} D_R^i + G_u^i (\bar{Q}_L^i U_R^i \tilde{\Phi} + \text{h.c.})] \\ & + \sum_{i,j=1}^3 [(\bar{Q}_L^i G_d^{ij} D_R^j \Phi + \text{h.c.})]. \end{aligned} \quad (82)$$

The covariant derivatives on the quark fields are given by

$$\begin{aligned} D_\mu Q_L^i &= \left(\partial_\mu - ig \frac{\vec{\tau}}{2} \times \vec{W}_\mu - i \frac{g'}{6} B_\mu \right) Q_L^i \\ D_\mu U_R^i &= \left(\partial_\mu - i \frac{2g'}{3} B_\mu \right) U_R^i \\ D_\mu D_R^i &= \left(\partial_\mu + i \frac{g'}{3} B_\mu \right) D_R^i. \end{aligned} \quad (83)$$

The classical Lagrangian in Eq. (82) contains 17 arbitrary real parameters. They are:

- the two gauge coupling constants g and g' ;
- the two parameters of the scalar potential λ and μ^2 ;
- three Yukawa coupling constants for the three lepton families, $G_{e,\mu,\tau}$;
- six Yukawa coupling constants for the three quark families, $G_u^{u,c,t}$; and $G_d^{d,s,b}$.
- four parameters of the KM matrix, the three angles and the phase δ .

A final remark: 15 out of these 17 parameters are directly connected with the Higgs sector.

Translating the scalar field by Eq. (68) and diagonalizing the resulting down-quark mass matrix produces the mass terms for fermions and bosons as well as several coupling terms. We shall write here the ones which involve the physical fields¹⁰.

¹⁰We know from QED that, in order to determine the Feynman rules of a gauge theory, one must first decide on a choice of gauge. For Yang–Mills theories, this step introduces new fields called *Faddeev–Popov ghosts*. This point is explained in every standard text book on quantum field theory, but we have not discussed it in these lectures.

(i) *The gauge boson–fermion couplings.* They are the ones which generate the known weak and electromagnetic interactions. A_μ is coupled to the charged fermions through the usual electromagnetic current:

$$\frac{gg'}{(g^2 + g'^2)^{1/2}} \left[\bar{e}\gamma^\mu e + \sum_{a=1}^3 \left(\frac{2}{3} \bar{u}^a \gamma^\mu u^a - \frac{1}{3} \bar{d}^a \gamma^\mu d^a \right) + \dots \right] A_\mu, \quad (84)$$

where the dots stand for the contribution of the other two families $e \rightarrow \mu, \tau$, $u \rightarrow c, t$ and $d \rightarrow s, b$ and the summation over a extends over the three colours. Equation (84) shows that the electric charge e is given in terms of g and g' by

$$e = \frac{gg'}{(g^2 + g'^2)^{1/2}} = g \sin \theta_W = g' \cos \theta_W. \quad (85)$$

Similarly, the couplings of the charged W to the weak current are

$$\frac{g}{2\sqrt{2}} \left(\bar{\nu}_e \gamma^\mu (1 + \gamma_5) e + \sum_{a=1}^3 \bar{u}^a \gamma^\mu (1 + \gamma_5) d_{KM}^a + \dots \right) W_\mu^+ + \text{h.c.} \quad (86)$$

Combining all these relations, we can determine the experimental value of the parameter v , the vacuum expectation value of the Higgs field. We find $v \sim 246$ GeV.

As expected, only left-handed fermions participate. d_{KM} is the linear combination of d – s – b given by the KM matrix in Eq. (81). By diagonalizing the down-quark mass matrix, we introduced the off-diagonal terms into the hadron current. When considering processes, like nuclear β -decay or μ -decay, where the momentum transfer is very small compared to the W mass, the W propagator can be approximated by m_W^{-2} and the effective Fermi coupling constant is given by

$$\frac{G}{\sqrt{2}} = \frac{g^2}{8m_W^2} = \frac{1}{2v^2}. \quad (87)$$

In contrast to the charged weak current shown in Eq. (86), the Z^0 -fermion couplings involve both left- and right-handed fermions:

$$-\frac{e}{2 \sin \theta_W \cos \theta_W} \left[\bar{\nu}_L \gamma^\mu \nu_L + (\sin^2 \theta_W - \cos^2 \theta_W) \bar{e}_L \gamma^\mu e_L + 2 \sin^2 \theta_W \bar{e}_R \gamma^\mu e_R + \dots \right] Z_\mu, \quad (88)$$

$$\frac{e}{2} \sum_{a=1}^3 \left[\left(\frac{1}{3} \tan \theta_W - \cot \theta_W \right) \bar{u}_L^a \gamma^\mu u_L^a + \left(\frac{1}{3} \tan \theta_W + \cot \theta_W \right) \bar{d}_L^a \gamma^\mu d_L^a + \frac{2}{3} \tan \theta_W (2 \bar{u}_R^a \gamma^\mu u_R^a - \bar{d}_R^a \gamma^\mu d_R^a) + \dots \right] Z_\mu. \quad (89)$$

Again, the summation is over the colour indices and the dots stand for the contribution of the other two families. In this formula we verify the property of the weak neutral current to be diagonal in the quark-flavour space. Another interesting property is that the axial part of the neutral current is proportional to $[\bar{u}\gamma_\mu\gamma_5 u - \bar{d}\gamma_\mu\gamma_5 d]$. This particular form of the coupling is important for phenomenological applications, such as the induced parity violating effects in atoms and nuclei.

(ii) *The gauge boson self-couplings.* One of the characteristic features of Yang–Mills theories is the particular form of the self-couplings among the gauge bosons. They come from the square of the

non-Abelian curvature in the Lagrangian, which, in our case, is the term $-\frac{1}{4}\vec{W}_{\mu\nu} \times \vec{W}^{\mu\nu}$. Expressed in terms of the physical fields, this term gives

$$\begin{aligned}
 & -ig(\sin\theta_W A^\mu - \cos\theta_W Z^\mu)(W^{\nu-}W_{\mu\nu}^+ - W^{\nu+}W_{\mu\nu}^-) \\
 & -ig(\sin\theta_W F^{\mu\nu} - \cos\theta_W Z^{\mu\nu})W_\mu^- W_\nu^+ \\
 & -g^2(\sin\theta_W A^\mu - \cos\theta_W Z^\mu)^2 W_\nu^+ W^{\nu-} \\
 & +g^2(\sin\theta_W A^\mu - \cos\theta_W Z^\mu)(\sin\theta_W A^\nu - \cos\theta_W Z^\nu)W_\mu^+ W_\nu^- \\
 & -\frac{g^2}{2}(W_\mu^+ W^{\mu-})^2 + \frac{g^2}{2}(W_\mu^- W^{\mu+})^2,
 \end{aligned} \tag{90}$$

where we have used the following notation: $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $W_{\mu\nu}^\pm = \partial_\mu W_\nu^\pm - \partial_\nu W_\mu^\pm$ and $Z_{\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu$ with $g \sin\theta_W = e$. Let us concentrate on the photon- W^+W^- couplings. If we forget, for the moment, about the SU(2) gauge invariance, we can use different coupling constants for the two trilinear couplings in Eq. (90), say e for the first and $e\kappa$ for the second. For a charged, massive W, the magnetic moment μ and the quadrupole moment Q are given by

$$\mu = \frac{(1+\kappa)e}{2m_W} \quad Q = -\frac{e\kappa}{m_W^2}. \tag{91}$$

Looking at Eq. (90), we see that $\kappa = 1$. Therefore, SU(2) gauge invariance gives very specific predictions concerning the electromagnetic parameters of the charged vector bosons. The gyromagnetic ratio equals two and the quadrupole moment equals $-em_W^{-2}$.

(iii) *The scalar fermion couplings.* They are given by the Yukawa terms in Eq. (65). The same couplings generate the fermion masses through spontaneous symmetry breaking. It follows that the physical Higgs scalar couples to quarks and leptons with strength proportional to the fermion mass. Therefore, the prediction is that it will decay predominantly to the heaviest possible fermion compatible with phase space. This property provides a typical signature for its identification.

(iv) *The scalar gauge boson couplings.* They come from the covariant derivative term $|D_\mu\Phi|^2$ in the Lagrangian. If we call ϕ the field of the physical neutral Higgs, we find

$$\frac{1}{4}(v+\phi)^2 [g^2 W_\mu^+ W^{\mu-} + (g^2 + g'^2) Z_\mu Z^\mu]. \tag{92}$$

This gives a direct coupling ϕ - W^+W^- , as well as ϕ - Z - Z , which has been very useful in the Higgs searches.

(v) *The scalar self-couplings.* They are proportional to $\lambda(v+\phi)^4$. Equations (75) and (87) show that $\lambda = Gm_h^2/\sqrt{2}$, so, in the tree approximation, this coupling is related to the Higgs mass. It could provide a test of the Standard Model Higgs, but it will not be easy to measure. On the other hand, this relation shows that, were the physical Higgs very heavy, it would also have been strongly interacting, and this sector of the model would become non-perturbative.

The five-step programme is now complete for both leptons and quarks. The 17 parameters of the model have all been determined by experiment. Although the number of arbitrary parameters seems very large, we should not forget that they are all mass and coupling parameters, like the electron mass and the fine structure constant of QED. The reason we have more of them is that the Standard Model describes a much larger number of particles and interactions in a unified framework.

6 The Standard Model and experiment

Our confidence in this model is amply justified on the basis of its ability to accurately describe the bulk of our present-day data and, especially, of its enormous success in predicting new phenomena. Let us mention a few of them. We shall follow the historical order.

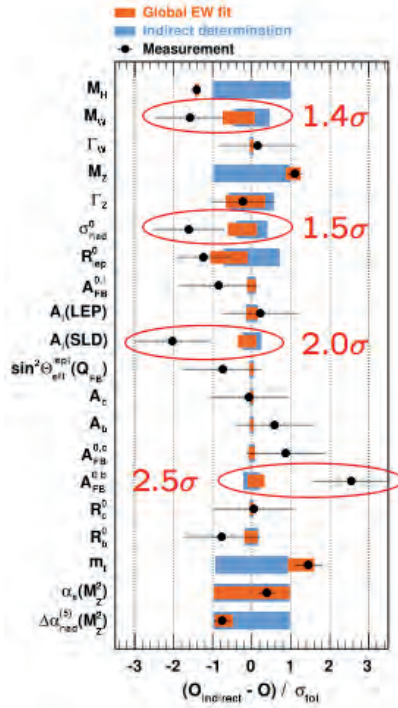


Fig. 5: A comparison between measured and computed values for various physical quantities

- The discovery of weak neutral currents by Gargamelle in 1972:

$$\nu_\mu + e^- \rightarrow \nu_\mu + e^- ; \quad \nu_\mu + N \rightarrow \nu_\mu + X .$$

Both their strength and their properties were predicted by the Standard Model.

- The discovery of charmed particles at SLAC in 1974. Their presence was essential to ensure the absence of strangeness changing neutral currents, for example $K^0 \rightarrow \mu^+ + \mu^-$. Their characteristic property is to decay predominantly into strange particles.
- A necessary condition for the consistency of the Model is that $\sum_i Q_i = 0$ inside each family. When the τ lepton was discovered this implied a prediction for the existence of the b and t quarks with the right electric charges.
- The observed CP violation could be naturally incorporated into a model with three families. The b and t quarks were indeed discovered.
- The discovery of the W and Z bosons at CERN in 1983 with the masses predicted by the theory. The characteristic relation of the Standard Model with an isodoublet Brout–Englert–Higgs mechanism $m_Z = m_W / \cos \theta_W$ has been checked with very high accuracy (including radiative corrections).
- The t-quark was *seen* at LEP through its effects in radiative corrections before its actual discovery at Fermilab.
- The vector boson self-couplings, $\gamma-W^+-W^-$ and $Z^0-W^+-W^-$ have been measured at LEP and confirm the Yang–Mills predictions given in Eq. (91).
- The recent discovery of a new boson which can be identified with the Higgs particle of the Standard Model is the last of this impressive series of successes.

All these discoveries should not make us forget that the Standard Model has been equally successful in fitting a large number of experimental results. You have all seen the global fit given in Fig. 5. The conclusion is obvious: *the Standard Model has been enormously successful.*

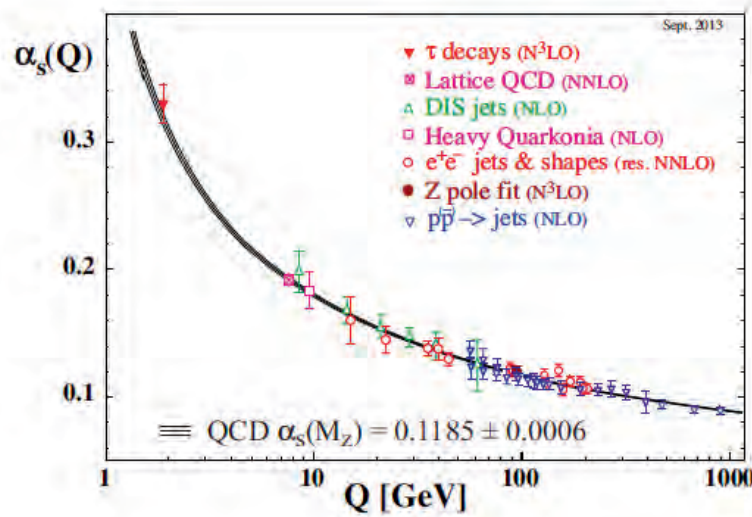


Fig. 6: The effective coupling constant for strong interactions as a function of the energy scale

Although in these lectures we did not discuss QCD, the gauge theory of strong interactions, the computations whose results are presented in Fig. 5, take into account the radiative corrections induced by virtual gluon exchanges. The fundamental property of QCD, the one which allows for perturbation theory calculations, is the property of asymptotic freedom, which is the particular dependence of the effective coupling strength on the energy scale. This is presented in Fig. 6 which shows the theoretical prediction based on QCD calculations, including the theoretical uncertainties. We see that the agreement with the experimentally measured values of the effective strong interaction coupling constant α_s is truly remarkable. Notice that this agreement extends to rather low values of Q of the order of 1–2 GeV, where α_s equals approximately $1/3$.

This brings us to our next point, namely that the success presented so far is in fact *a success of renormalized perturbation theory*. The extreme accuracy of the experimental measurements, mainly at LEP but also at FermiLab and elsewhere, allow a detailed comparison between theory and experiment to be made for the first time *including the purely weak interaction radiative corrections*.

In Fig. 7 we show the comparison between theory and experiment for two quantities, ϵ_1 and ϵ_3 , defined in Eqs. (93) and (94), respectively:

$$\epsilon_1 = \frac{3G_F m_t^2}{8\sqrt{2}\pi^2} - \frac{3G_F m_W^2}{4\sqrt{2}\pi^2} \tan^2 \theta_W \ln \frac{m_H}{m_Z} + \dots, \quad (93)$$

$$\epsilon_3 = \frac{G_F m_W^2}{12\sqrt{2}\pi^2} \ln \frac{m_H}{m_Z} - \frac{G_F m_W^2}{6\sqrt{2}\pi^2} \ln \frac{m_t}{m_Z} + \dots. \quad (94)$$

They are defined with the following properties: (i) they include the strong and electromagnetic radiative corrections; and (ii) they vanish in the Born approximation for the weak interactions. So, they measure the weak interaction radiative corrections. The figure shows that, in order to obtain agreement with the data, one must include these corrections. Weak interactions are no longer a simple phenomenological model, but have become a precision theory.

The moral of the story is that the perturbation expansion of the Standard Model is reliable as long as all coupling constants remain small. The only coupling which does become large in some kinematical regions is α_s , which grows at small energy scales, as shown in Fig. 6. In this region, we know that a hadronization process occurs and perturbation theory breaks down. New techniques are necessary in

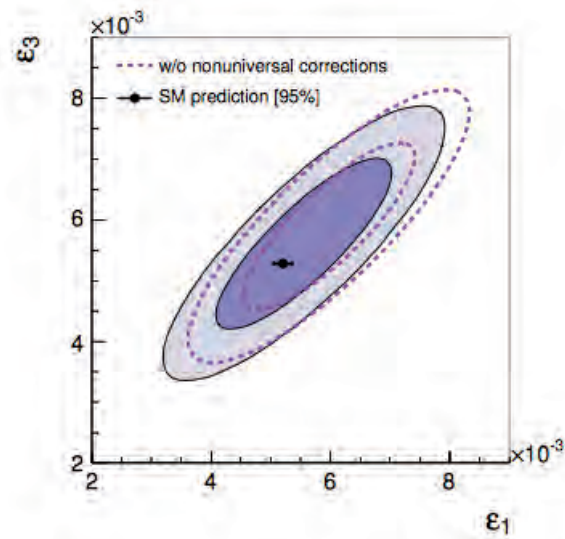


Fig. 7: Comparison between theory and experiment for two quantities sensitive to weak interaction radiative corrections.

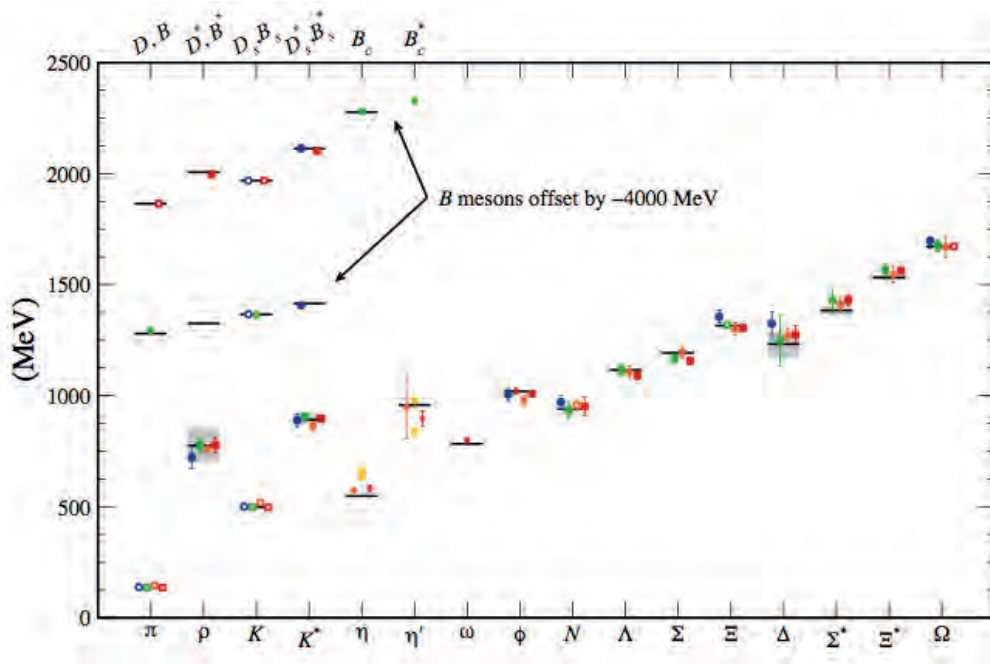


Fig. 8: The hadron spectrum obtained by numerical simulations of QCD on a space–time lattice

order to compare theoretical predictions with experimental data. In recent years, considerable effort has been devoted to this question with extensive numerical studies of QCD in the approximation in which the four-dimensional space–time has been replaced by a finite lattice. In Fig. 8 we show the computed spectrum of low-lying hadron states and the comparison with the data. The agreement makes us believe that we control the theory at both the weak- and strong-coupling regime. We should no longer talk about the Standard Model, but rather about the Standard Theory of the interactions among elementary particles. As a by-product of this analysis, we feel confident to say that at high energies perturbation theory is expected to be reliable *unless there are new strong interactions*.

This brings us to our last point that this very success shows that the Standard Model cannot be a complete theory. In other words there must be new physics beyond the Standard Model. The argument is simple and it is based on a straightforward application of perturbation theory with an additional assumption which we shall explain presently.

We assume that the Standard Model is correct up to a certain scale Λ . The precise value of Λ does not matter, provided it is larger than any energy scale reached so far¹¹.

A quantum field theory is defined through a functional integral over all classical field configurations, the Feynman path integral. By a Fourier transformation we can express it as an integral over the fields defined in momentum space. Following K. Wilson, let us split this integral in two parts: the high-energy part with modes above Λ and the low-energy part with the modes below Λ . Let us imagine that we perform the high-energy part. The result will be an effective theory expressed in terms of the low-energy modes of the fields. We do not know how to perform this integration explicitly, so we cannot write down the correct low-energy theory, but the most general form will be a series of operators made out of powers of the fields and their derivatives. Since integrating over the heavy modes does not break any of the symmetries of the initial Lagrangian, only operators allowed by the symmetries will appear. Wilson remarked that, when Λ is large compared to the mass parameters of the theory, we can determine the leading contributions by simple dimensional analysis¹². We distinguish three kinds of operators, according to their canonical dimension.

- Those with dimension larger than four. Dimensional analysis shows that they will come with a coefficient proportional to inverse powers of Λ , so, by choosing a scale large enough, we can make their contribution arbitrarily small. We shall call them *irrelevant operators*.
- Those with dimension equal to four. They are the ones which appeared already in the original Lagrangian. Their coefficient will be independent of Λ , up to logarithmic corrections which we ignore. We shall call them *marginal operators*.
- Finally, we have the operators with dimension smaller than four. In the Standard Model there is only one such operator, the square of the scalar field Φ^2 which has dimension equal to two¹³. This operator will appear with a coefficient proportional to Λ^2 , which means that its contribution will grow quadratically with Λ . We shall call it the *relevant operator*. It will give an effective mass to the scalar field proportional to the square of whichever scale we can think of. This problem was first identified in the framework of Grand Unified Theories and is known since as *the hierarchy problem*. Let me emphasize here that this does not mean that the mass of the scalar particle will be necessarily equal to Λ . The Standard Model is a renormalizable theory and the mass is fixed by a renormalization condition to its physical value. It only means that this condition should be adjusted to arbitrary precision order by order in perturbation theory. It is this extreme sensitivity to high scales, known as *the fine tuning problem*, which is considered unacceptable for a fundamental theory.

Let us summarize: the great success of the Standard Model tells us that renormalized perturbation theory is reliable in the absence of strong interactions. The same perturbation theory shows the need of a fine tuning for the mass of the scalar particle. If we do not accept the latter, we have the following two options.

¹¹The scale Λ should not be confused with a cut-off that is often introduced when computing Feynman diagrams. This cut-off disappears after renormalization is performed. Here Λ is a physical scale which indicates how far the theory can be trusted.

¹²There are some additional technical assumptions concerning the dimensions of the fields, but they are satisfied in perturbation theory.

¹³There exists also the unit operator with dimension equal to zero which induces an effective cosmological constant. Its effects are not observable in a theory which ignores the gravitational interactions, so we shall not discuss it here. One could think of the square of a fermion operator $\bar{\Psi}\Psi$, whose dimension is equal to three, but it is not allowed by the chiral symmetry of the model.

- Perturbation theory breaks down at some scale Λ . We can imagine several reasons for a such a breakdown to occur. The simplest is the appearance of new strong interactions. The so-called *technicolor* models, in which the role of the Higgs field is played by a bound state of new strongly coupled fermions, were in this class. More exotic possibilities include the appearance of new, compact space dimensions with compactification length $\sim \Lambda^{-1}$.
- Perturbation theory is still valid but the numerical coefficient of the Λ^2 term which multiplies the Φ^2 operator vanishes to all orders of perturbation theory. For this to happen we must modify the Standard Model introducing appropriate new particles. Supersymmetry is the only systematic way we know to achieve this goal.

7 Conclusions

In these lectures we saw the fundamental role of geometry in the dynamics of the forces among the elementary particles. It was the understanding of this role which revolutionized our way of thinking and led to the construction of the Standard Model. It incorporates the ideas of gauge theories, as well as those of spontaneous symmetry breaking. Its agreement with experiment is spectacular. It fits all data known today. However, unless one is willing to accept a fine tuning with arbitrary precision, one should conclude that new physics will appear beyond a scale Λ . The precise value of Λ cannot be computed, but the amount of fine tuning grows quadratically with it, so it cannot be too large. Hopefully, it will be within reach of the LHC.

Appendix A: The principles of renormalization

In this appendix I want to recall and summarize the basic principles of perturbative renormalization theory. Since renormalization has a well-deserved reputation of complexity, this will be done by omitting all technical details. My purpose is to dissipate a widely spread belief according to which renormalization is a mathematically murky procedure: adding and subtracting infinities. On the contrary, I want to explain that it offers the only known mathematically consistent way to define the perturbation expansion of a quantum field theory.

A.1 The need for renormalization

Everyone who has attempted to compute a one-loop Feynman diagram knows that divergent expressions are often encountered. For example, in the ϕ^4 theory we find the diagram of Fig. A.1 involving the integral

$$I = \int \frac{d^4k}{(k^2 - m^2 + i\epsilon)[(k-p)^2 - m^2 + i\epsilon]}, \quad (\text{A1})$$

which diverges logarithmically at large k . Similar divergences can be found in any theory, such as QED, Yang–Mills, etc. They have no place in a well-defined mathematical theory. So, if we find them, it means that we have made a mathematical mistake somewhere. Where is it? Let us first notice that the divergence in Eq. (A1) occurs at large values of the internal momentum, which, by Fourier transform, implies short distances. Did we make a mistake at short distances? Yes we did! We wrote the Lagrangian density as

$$\mathcal{L} = \frac{1}{2} (\partial_\mu \phi(x)) (\partial^\mu \phi(x)) - \frac{1}{2} m^2 (\phi(x))^2 - \frac{\lambda}{4!} (\phi(x))^4. \quad (\text{A2})$$

On the other hand, the canonical commutation relations for a scalar quantum field are given by

$$\left[\phi(\vec{x}, t), \dot{\phi}(\vec{y}, t) \right] = i\hbar \delta^3(\vec{x} - \vec{y}). \quad (\text{A3})$$

We know that the Dirac δ -function is not really a ‘function’ but a special form of what we call ‘a distribution’. Many properties of well-behaved functions do not apply to it. In particular, the multiplication is not always a well-defined operation. $(\delta(x))^2$ is meaningless. The presence of the δ -function on the right-hand side of Eq. (A3) implies that the field $\phi(x)$ is also a distribution¹⁴, so the product ϕ^2 is ill defined. Yet, it is precisely expressions of this kind that we wrote in every single term of our Lagrangian Eq. (A2). Since our initial Lagrangian is not well defined, it is not surprising that our calculations yield divergent results.

Now that we have identified the origin of the problem, we can figure out ways to solve it. A conceptually simple one would be to replace the field products in Eq. (A2) by splitting the points:

$$\phi(x)\phi(x) \rightarrow \lim_{a \rightarrow 0} \phi\left(x + \frac{a}{2}\right) \phi\left(x - \frac{a}{2}\right). \quad (\text{A4})$$

This expression is perfectly well defined for all values of the parameter a , except $a = 0$. In terms of distributions this means that the product is defined up to an arbitrary distribution $\mathcal{F}(a)$ which has support (i.e. it is non-zero), only at $a = 0$. Such a distribution is a superposition of the δ -function and its derivatives,

$$\mathcal{F}(a) = \sum_i C_i \delta^{(i)}(a) \quad (\text{A5})$$

¹⁴The precise term is ‘operator valued distribution’.

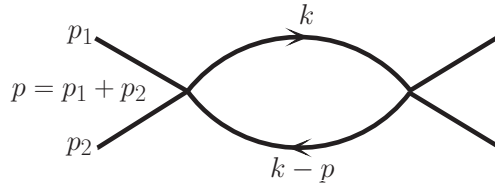


Fig. A.1: An one-loop divergent diagram in the ϕ^4 theory

with the C_i arbitrary real constants. The moral of the story is that the quantization rules for a local-field theory imply that every term in the Lagrangian contains a set of arbitrary constants which must be determined by experiment. Renormalization is the mathematical procedure which allows us to do it. A final remark: how many parameters are needed in order to define a given field theory? The answer involves the distinction between *renormalizable* and *non-renormalizable* theories. For the first, a finite number suffices. For the second, we need an infinite number, which means that non-renormalizable theories have no predictive power.

A.2 The theory of renormalization

In this section, I want to give some more information concerning the renormalization prescription. The process we outlined above was formulated in x -space. It is intuitively easier to understand, but not very convenient for practical calculations, which are usually performed in momentum space. The connection is by Fourier transform. The derivatives of the δ -function in Eq. (A5) become polynomials in the external momenta.

The renormalization programme follows three steps:

- *the power counting* which determines how many constants C we shall need for a given field theory;
- *the regularization* which is a prescription to make every Feynman diagram finite with the price of introducing a new parameter in the theory, the analogue of the point-splitting parameter a we used in Eq. (A4);
- *the renormalization* which is the mathematical procedure to eliminate the regularization parameter and determine the values of the necessary constants C .

A.2.1 The power counting

As the term indicates, it is the counting which determines whether a given diagram is divergent or not. We shall need to introduce some terminology. First, we have the obvious notions of *disconnected* and *connected* diagrams. A further specification is the *one-particle irreducible* (1PI) diagrams. A diagram is 1PI if it cannot be separated into two disconnected pieces by cutting a single internal line. A general connected diagram is constructed by joining together 1PI pieces, see Fig. A.2. It is obvious that a connected diagram is divergent if, and only if, one or more of its 1PI pieces is divergent, because the momenta of the internal connecting lines are fixed by energy-momentum conservation in terms of the external momenta and bring no new integrations.

This brings us to the power-counting argument. A single loop integral will be ultravioletly divergent if and only if the numerator is of equal or higher degree in the loop momentum than the denominator. For multiloop diagrams this may not be the case, since the divergence may be entirely due to a particular sub-diagram. However, in the spirit of perturbation theory, the divergent sub-diagram must be treated first. We thus arrive at the notion of *superficial degree of divergence* d of a given 1PI diagram, defined as the difference between the degree of integration momenta of the numerator minus that of the denominator. The diagram will be called *primitively divergent* if $d \geq 0$. Let us compute, as an example, d for

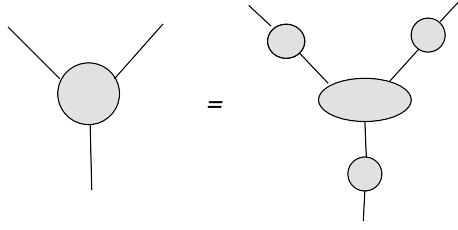


Fig. A.2: The 1PI decomposition of the three point function

the diagrams of the scalar field theory described in Eq. (A2), in the generalization in which we replace the interaction term ϕ^4 by ϕ^m with m integer, $m \geq 3$. Let us consider an 1PI diagram of n th order in perturbation with I internal and E external lines. Every internal line brings four powers of k to the numerator through the d^4k factor and two powers in the denominator through the propagator. Every vertex brings a δ^4 -function of the energy-momentum conservation. All but one of them can be used to eliminate one integration each, the last reflecting the overall conservation which involves only external momenta. Therefore, we obtain

$$d = 2I - 4n + 4. \quad (\text{A6})$$

This expression can be made more transparent by expressing I in terms of E and m . A simple counting gives $2I + E = mn$ and Eq. (A6) becomes

$$d = (m - 4)n - E + 4. \quad (\text{A7})$$

This is the main result. Although it is shown here as a plausibility argument, it is in fact a rigorous result. We see that $m = 4$ is a critical value and we can distinguish three cases.

1. $m = 3$, $d = 4 - n - E$. d is a decreasing function of n , the order of perturbation theory. Only a limited number of diagrams are primitively divergent. Above a certain order they are all convergent. For reasons that will be clear soon, we shall call such theories *super-renormalizable*.
2. $m = 4$, $d = 4 - E$. d is independent of the order of perturbation theory. If a Green function is divergent at some order, it will be divergent at all orders. For the ϕ^4 theory we see that the primitively divergent diagrams are those with $E = 2$, which have $d = 2$ and are quadratically divergent and those with $E = 4$ which have $d = 0$ and are logarithmically divergent. (Notice that, for this theory, all Green functions with odd E vanish identically because of the symmetry $\phi \rightarrow -\phi$). We shall call such theories *renormalizable*.
3. $m > 4$, d is an increasing function of n . Every Green function, irrespective of the number of external lines, will be divergent above some order of perturbation. We call such theories *non-renormalizable*.

This power-counting analysis can be repeated for any quantum field theory. As a second example, we can look at QED. We should now distinguish between photon and electron lines, which we shall denote by I_γ , I_e , E_γ and E_e for internal and external lines, respectively. Taking into account the fact that the fermion propagator behaves like k^{-1} at large momenta, for the superficial degree of divergence of an 1PI diagram we obtain

$$d = 2I_\gamma + 3I_e - 4n + 4 = 4 - E_\gamma - \frac{3}{2}E_e. \quad (\text{A8})$$

We see that d is independent of the order of perturbation theory and, therefore, the theory is renormalizable.

We leave it as an exercise to the reader to establish the renormalization properties of other field theories. In four dimensions of space–time, the result is:

1. there exists only one super-renormalizable field-theory with interaction of the form ϕ^3 ;
2. there exist five renormalizable ones:
 - (a) ϕ^4 ;
 - (b) Yukawa $\bar{\psi}\psi\phi$;
 - (c) QED $\bar{\psi}\gamma_\mu A^\mu\psi$;
 - (d) scalar electrodynamics, it contains two terms $[\phi^\dagger\partial_\mu\phi - (\partial_\mu\phi^\dagger)\phi]A^\mu$ and $A^\mu A_\mu\phi^\dagger\phi$;
 - (e) Yang–Mills $\text{Tr } G_{\mu\nu}G^{\mu\nu}$;
3. all other theories are non-renormalizable.

For ϕ^3 , the energy will turn out to be unbounded from below, so this theory alone cannot be a fundamental theory for a physical system. A most remarkable fact is that, as we shall see later, nature uses *all* renormalizable theories to describe the interactions among elementary particles.

Before closing this section we want to make a remark which is based on ordinary dimensional analysis. In four dimensions, a boson field has dimensions of a mass (remember, we are using units such that the speed of light c and Planck’s constant h are dimensionless) and a fermion field with a mass to the power $3/2$. Since all terms in a Lagrangian density must have dimensions equal to four, we conclude that the coupling constant of a super-renormalizable theory must have the dimensions of a mass, a renormalizable theory must be dimensionless, and a non-renormalizable theory must have the dimensions of an inverse power of mass. In fact we can rephrase the power-counting argument for the superficial degree of divergence of an 1PI diagram as an argument based on dimensional analysis. The result will be this connection between the dimensions of the coupling constant and the renormalization properties of the theory. However, there is a fine point: for this argument to work we must assume that all boson propagators behave like k^{-2} at large momenta and all fermion ones like k^{-1} . So, the argument will fail if this behaviour is not true. The most important example of such a failure is a theory containing massive vector fields whose propagator is like a constant at large k . As a result, such theories, although they may have dimensionless coupling constants, are in fact non-renormalizable.

A.2.2 Regularization

The point splitting we presented in Eq. (A4) is an example of a procedure we shall call *regularization*. It consists of introducing an extra parameter in the theory (in the case considered, it was the splitting distance a), to which we do not necessarily attach a physical meaning, with the following properties: (i) the initial theory is recovered for a particular value of the parameter, in our example $a = 0$; (ii) the theory is finite for all values of the parameter in a region which contains the ‘physical’ one $a = 0$; and (iii) at this value we get back the divergences of the initial theory. We shall call this parameter *a cut-off*.

If our purpose is to perform computations of Feynman diagrams, we may choose any cut-off procedure that renders these diagrams finite. There is a plethora of such methods and there is no need to give a complete list. A direct method would be to cut all integrations of loop momenta at a scale Λ . The initial theory is recovered at the limit $\Lambda \rightarrow \infty$. For practical calculations it is clear that we must choose a cut-off procedure that renders these computations as simple as possible. By trial and error, the simplest regularization scheme turned out to be a quite counter-intuitive one. We start by illustrating it in the simple example of the divergent integral of Eq. (A1). Since we are interested only in the divergent part, we can simplify the discussion by considering the value of I at $p = 0$. We thus obtain

$$I = \int \frac{d^4k}{(2\pi)^4} \frac{1}{(k^2 - m^2 + i\epsilon)^2}. \quad (\text{A9})$$

Ignoring the divergence for the moment, we notice that the integrand depends only on k^2 , so we choose spherical coordinates and write $d^4k = k^3 dk d\Omega^{(3)}$, where $d\Omega^{(3)}$ is the surface element on the three-dimensional unit sphere. Further, we notice that I would have been convergent if we were working in a space-time of three, two or one dimensions. The crucial observation is that in all three cases we can write the result in a compact form as follows¹⁵:

$$I^{(d)} = \int \frac{d^d k}{(2\pi)^d} \frac{1}{(k^2 + m^2)^2} = \frac{1}{(4\pi)^{d/2}} \frac{\Gamma(2 - d/2)}{(m^2)^{(2-d/2)}}; \quad d = 1, 2, 3, \quad (\text{A10})$$

where $\Gamma(z)$ is the well-known special function which generalizes the concept of the factorial for a complex z . The important values for Eq. (A10) are given by

$$\Gamma(n) = (n - 1)!; \quad \Gamma(n + 1/2) = \frac{(\pi)^{1/2}}{2^n} (2n - 1)!!; \quad n = 1, 2, \dots \quad (\text{A11})$$

And now comes the big step. Nothing on the right-hand side of Eq. (A10) forces us to consider this expression only for $d = 1, 2$ or 3 . In fact, Γ is a meromorphic function in the entire complex plane with poles whenever its argument becomes equal to an integer $n \leq 0$. For the integral $I^{(d)}$, using the identity $n\Gamma(n) = \Gamma(n + 1)$, we see that, when $d \rightarrow 4$, the Γ function behaves as $\Gamma(2 - d/2) \sim 2/(4 - d)$. So we can argue that, at least for this integral, we have introduced a regularization, i.e. a new parameter, namely $\epsilon = 4 - d$, such that the expression is well defined for all values in a region of ϵ and diverges when $\epsilon \rightarrow 0$.

Before showing how to generalize this approach to all other integrals we may encounter in the calculation of Feynman diagrams, let us try to make the logic clear by emphasizing what this regularization does not claim to be. First, it does not claim to be the result one would have obtained by quantizing the theory in a complex number of dimensions. In fact we do not know how to consistently perform such an operation. In this sense, dimensional regularization does not offer a non-perturbative definition of the field theory. The prescription applies directly to the integrals obtained order by order in the perturbation expansion. Second, it cannot even be viewed as the analytic continuation to the complex d plane of the results we obtain in performing the integral for $d = 1, 2, 3$. Indeed, the knowledge of the values of a function on a finite number of points on the real axis does not allow for a unique analytic continuation. Instead, the claim is that Eq. (A10), appropriately generalized, offers an unambiguous prescription to obtain a well-defined answer for any Feynman diagram as long as ϵ stays away from zero.

The observation which allows for such a generalization is that Feynman rules always yield a special class of integrals. In purely bosonic theories, whether renormalizable or not, they are of the form

$$I(p_1, p_2, \dots, p_n) = \int \prod_i \left(\frac{d^d k_i}{(2\pi)^d} \right) \frac{N(k_1, k_2, \dots)}{D(k_1, k_2, \dots)} \prod_r \left((2\pi)^d \delta^d(k, p) \right), \quad (\text{A12})$$

where the k and the p are the momenta of the internal and external lines respectively, the product over i runs over all internal lines, that of r over all vertices, the δ functions denote the energy and momentum conservation on every vertex, and N and D are polynomials of the form

$$N(k_1, k_2, \dots) = k_1^{\mu_1} k_1^{\mu_2} \dots k_2^{\nu_1} k_2^{\nu_2} \dots, \quad (\text{A13})$$

$$D(k_1, k_2, \dots) = \prod_i (k_i^2 + m_i^2). \quad (\text{A14})$$

D is just the product of all propagators and m_i is the mass of the i th line. N appears through derivative couplings and/or the $k^\mu k^\nu$ parts of the propagators of higher-spin bosonic fields. It equals

¹⁵We write the result after a Wick rotation in Euclidean space

one in theories with only scalar fields and non-derivative couplings, such as ϕ^4 . All scalar products are written in terms of the d -dimensional Euclidean metric $\delta_{\mu\nu}$ which satisfies

$$\delta_{\mu}^{\mu} = \text{Tr } \mathbf{1} = d. \quad (\text{A15})$$

The dimensional regularization consists of giving a precise expression for $I(p_1, p_2, \dots, p_n)$ as a function of d which coincides with the usual value whenever the latter exists and is well defined for every value of d in the complex d plane except for those positive integer values for which the original integral is divergent.

At one loop the integral Eq. (A12) reduces to

$$I(p_1, p_2, \dots, p_n) = \int \frac{d^d k}{(2\pi)^d} \frac{N(k)}{D(k, p_1, p_2, \dots)}, \quad (\text{A16})$$

with k being the loop momentum. The denominator D is of the form

$$D(k, p_1, p_2, \dots) = \prod_i [(k - \Sigma_{(i)} p)^2 + m_i^2], \quad (\text{A17})$$

where $\Sigma_{(i)} p$ denotes the combination of external momenta which goes through the i th internal line. This product of propagators can be cast in a more convenient form by using a formula first introduced by Feynman:

$$\frac{1}{P_1 P_2 \dots P_\eta} = (\eta - 1)! \int_0^1 \frac{dz_1 dz_2 \dots dz_\eta \delta(1 - \sum_i z_i)}{[z_1 P_1 + z_2 P_2 + \dots + z_\eta P_\eta]^\eta}. \quad (\text{A18})$$

With the help of Eq. (A18) and an appropriate change of variables, all one-loop integrals become of the general form

$$\hat{I}(p_1, p_2, \dots, p_n) = \int \frac{d^d k}{(2\pi)^d} \frac{k_{\mu_1} k_{\mu_2} \dots k_{\mu_l}}{[k^2 + F^2(p, m, z)]^\eta}, \quad (\text{A19})$$

with F some scalar function of the external momenta, the masses and the Feynman parameters. F has the dimensions of a mass. $I(p_1, p_2, \dots, p_n)$ is obtained from $\hat{I}(p_1, p_2, \dots, p_n)$ after integration with respect to the Feynman parameters z_i of Eq. (A18). For odd values of l , \hat{I} vanishes by symmetric integration. For l even it can be easily computed using spherical coordinates. Some simple cases are as follows:

$$\int \frac{d^d k}{(2\pi)^d} \frac{1}{[k^2 + F^2(p, m, z)]^\eta} = \frac{1}{(4\pi)^{d/2}} \frac{\Gamma(\eta - d/2)}{\Gamma(\eta)} [F^2]^{(d/2 - \eta)}. \quad (\text{A20})$$

$$\int \frac{d^d k}{(2\pi)^d} \frac{k_\mu k_\nu}{[k^2 + F^2(p, m, z)]^\eta} = \frac{1}{(4\pi)^{d/2}} \frac{\delta_{\mu\nu}}{2} \frac{\Gamma(\eta - 1 - d/2)}{\Gamma(\eta)} [F^2]^{(d/2 + 1 - \eta)}. \quad (\text{A21})$$

At the end, we are interested in the limit $d \rightarrow 4$. The first integral Eq. (A20) diverges for $\eta \leq 2$ and the second Eq. (A21) for $\eta \leq 3$. For $\eta = 2$ and $d = 4$, Eq. (A20) is logarithmically divergent and our regularized expression is regular for $\text{Re } d < 4$ and presents a simple pole $\sim 1/(d - 4)$. For $\eta = 1$, it is quadratically divergent but our expression still has a simple pole at $d = 4$. The difference is that now the first pole from the left is at $d = 2$. We arrive at the same conclusions looking at the integral of Eq. (A21): by dimensionally regularizing a one-loop integral corresponding to a Feynman diagram which, by power counting, diverges as Λ^{2n} , we obtain a meromorphic function of d with simple poles starting at $d = 4 - 2n$. By convention, $n = 0$ denotes a logarithmic divergence.

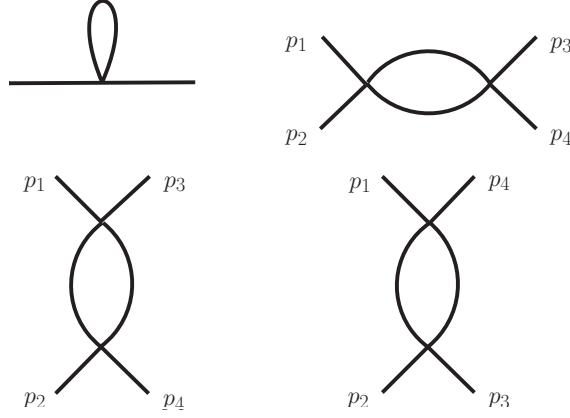


Fig. A.3: The one loop primitively divergent diagrams of the ϕ^4 theory

A.2.3 Renormalization

In this section, we want to address the physical question of under which circumstances can a meaningful four-dimensional theory be recovered from the regularized ϵ -dependent expressions. As one could have anticipated, the answer will turn out to be that this is only possible for the renormalizable (and super-renormalizable) theories we introduced before. The procedure to do so is called *renormalization*. In this section, we shall present some simple examples.

Let us start with the simplest four-dimensional renormalizable theory given by our already familiar Lagrangian density from Eq. (A2). In $d = 4$, the field ϕ has the dimensions of a mass and the coupling constant λ is dimensionless. Since we intend to use dimensional regularization, we introduce a mass parameter μ and write the coefficient of the interaction term $\lambda \rightarrow \mu^\epsilon \lambda$, so that the coupling constant λ remains dimensionless at all values of ϵ . We shall present the renormalization programme for this theory at the lowest non-trivial order, that which includes all diagrams up to and including those with one closed loop.

The power-counting argument presented previously shows that, at one loop, the only divergent 1PI diagrams are the ones of Fig. A.3.

The two-point diagram is quadratically divergent and the four-point diagram is logarithmically divergent¹⁶. We choose to work entirely with dimensional regularization and for these diagrams in Minkowski space-time, using (A20) at the limit $d \rightarrow 4$, we obtain

$$\Gamma_1^{(2)} = \frac{\lambda \mu^\epsilon}{2} \int \frac{d^d k}{(2\pi)^d} \frac{1}{k^2 - m^2} = \frac{i\lambda m^2}{16\pi^2} \frac{1}{\epsilon} \quad (\text{A22})$$

$$\begin{aligned} \Gamma_1^{(4)}(p_1, \dots, p_4) &= \frac{1}{2} \lambda^2 \mu^{2\epsilon} \int \frac{d^d k}{(2\pi)^d} \frac{1}{(k^2 - m^2)[(k - P)^2 - m^2]} + \text{crossed} \\ &= \frac{1}{2} \lambda^2 \mu^{2\epsilon} \int_0^1 dz \int \frac{d^d k}{(2\pi)^d} \frac{1}{[k^2 - m^2 + P^2 z(1 - z)]^2} + \text{crossed} \\ &= \frac{3i\lambda^2}{16\pi^2} \frac{1}{\epsilon} + \text{finite terms}, \end{aligned} \quad (\text{A23})$$

¹⁶We could prevent the appearance of the first diagram by ‘normal ordering’ the ϕ^4 term in the interaction Lagrangian, but, for pedagogical purposes, we prefer not to do so. Normal ordering is just a particular prescription to avoid certain divergences, but it is not always the most convenient one. First, it is not general. For example, it will not prevent the appearance of divergence in the two-point function at higher orders and second, its use may complicate the discussion of possible gauge symmetries of \mathcal{L} .

where $P = p_1 + p_2$, ‘crossed’ stands for the contribution of the two crossed diagrams in Fig. A.3 and ‘finite terms’ represent the contributions which are regular when $d = 4$. We can make the following remarks.

1. The divergent contributions are constants, independent of the external momenta. We shall see shortly, in the example of QED, that this is a particular feature of the ϕ^4 theory. In fact, even for ϕ^4 , it is no longer true when higher loops are considered. However, we can prove the following general property: all divergent terms are proportional to monomials in the external momenta. We have already introduced this result. For one-loop diagrams the proof is straightforward. We start from the general expression of Eq. (A19) and notice that we can expand the integrand in powers of the external momenta p taken around some fixed point. Every term in this expansion increases the value of η , so after a finite number of terms, the integral becomes convergent. It takes some more work to generalize the proof to multi-loop diagrams, but it can be done,
2. The dependence of the divergent terms on m^2 could be guessed from dimensional analysis. This is one of the attractive features of dimensional regularization,
3. The finite terms in Eq. (A23) depend on the parameter μ . The Laurent expansion in ϵ brings terms of the form $\ln\{[m^2 - P^2 z(1 - z)]/\mu^2\}$.

The particular form of the divergent terms suggests the prescription to remove them. Let us start with the two-point function. In the loop expansion we write

$$\Gamma^{(2)}(p^2) = \sum_{l=0}^{\infty} \Gamma_l^{(2)}(p^2) = \Gamma_0^{(2)}(p^2) + \Gamma_1^{(2)}(p^2) + \dots, \quad (\text{A24})$$

where the index l denotes the contribution of the diagrams with l loops. In the tree approximation we have

$$\Gamma_0^{(2)}(p^2) = -i(p^2 - m^2). \quad (\text{A25})$$

The one-loop diagram adds the term given by Eq. (A22). Since it is a constant, it can be interpreted as a correction to the value of the mass in Eq. (A25). Therefore, we can introduce a *renormalized* mass m_R^2 , which is a function of m , λ and ϵ . Of course, this function can only be computed as a formal power series in λ . Up to and including one-loop diagrams we write

$$m_R^2(m, \lambda, \epsilon) = m^2 \left(1 + \frac{\lambda}{16\pi^2} \frac{1}{\epsilon} \right) + \mathcal{O}(\lambda^2). \quad (\text{A26})$$

A formal power series whose zero-order term is non-vanishing is invertible in terms of another formal power series. So, we can write m as a function of m_R , λ and ϵ :

$$m^2(m_R, \lambda, \epsilon) = m_R^2 \left(1 - \frac{\lambda}{16\pi^2} \frac{1}{\epsilon} \right) + \mathcal{O}(\lambda^2) \equiv m_R^2 Z_m + \mathcal{O}(\lambda^2), \quad (\text{A27})$$

where we have defined the function $Z_m(\lambda, \epsilon)$ as a formal power series in λ with ϵ -dependent coefficients.

The parameter m is often called the *bare* mass. In the Lagrangian Eq. (A2), replacing the bare mass m with the help of Eq. (A27) results in: (i) changing the Feynman rules m by m_R and (ii) introducing a new term in \mathcal{L} of the form

$$\delta\mathcal{L}_m = m_R^2 \frac{\lambda}{32\pi^2} \frac{1}{\epsilon} \phi^2(x). \quad (\text{A28})$$



Fig. A.4: The new diagram resulting from $\delta\mathcal{L}_m$ of Eq. (A28)

Since $\delta\mathcal{L}_m$ is proportional to the coupling constant λ , we can view it as a new vertex in the perturbation expansion which, to first order, gives the diagram of Fig. A.4. In this case the complete two-point function to first order in λ is given by

$$\begin{aligned}\Gamma^{(2)}(p^2) &= -i(p^2 - m_{\text{R}}^2) + \frac{i\lambda m_{\text{R}}^2}{16\pi^2} \frac{1}{\epsilon} - \frac{i\lambda m_{\text{R}}^2}{16\pi^2} \frac{1}{\epsilon} + \mathcal{O}(\lambda^2) \\ &= -i(p^2 - m_{\text{R}}^2) + \mathcal{O}(\lambda^2),\end{aligned}\tag{A29}$$

which means that, if we keep fixed m_{R} and λ instead of m and λ , we can take the limit $d \rightarrow 4$ and find no divergences up to and including one-loop diagrams for the two-point function.

Now that we have understood the principle, it is straightforward to apply it to the four-point function. In the same spirit we write

$$\Gamma^{(4)}(p_1, \dots, p_4) = \sum_{l=0}^{\infty} \Gamma_l^{(4)}(p_1, \dots, p_4) = \Gamma_0^{(4)}(p_1, \dots, p_4) + \Gamma_1^{(4)}(p_1, \dots, p_4) + \dots\tag{A30}$$

In the tree approximation, $\Gamma_0^{(4)}(p_1, \dots, p_4) = -i\lambda$. Including the one-loop diagrams we obtain

$$\Gamma^{(4)}(p_1, \dots, p_4) = -i\lambda \left(1 - \frac{3\lambda}{16\pi^2} \frac{1}{\epsilon} + \text{finite terms} \right) + \mathcal{O}(\lambda^3).\tag{A31}$$

We change from the *bare* coupling constant λ to the *renormalized* one λ_{R} by writing

$$\lambda_{\text{R}}(\lambda, \epsilon) = \lambda \left(1 - \frac{3\lambda}{16\pi^2} \frac{1}{\epsilon} + \mathcal{O}(\lambda^2) \right),\tag{A32}$$

or, equivalently,

$$\lambda(\lambda_{\text{R}}, \epsilon) = \lambda_{\text{R}} \left(1 + \frac{3\lambda_{\text{R}}}{16\pi^2} \frac{1}{\epsilon} + \mathcal{O}(\lambda_{\text{R}}^2) \right) \equiv \lambda_{\text{R}} Z_{\lambda}.\tag{A33}$$

Again, replacing λ with λ_{R} in \mathcal{L} produces a new four-point vertex which cancels the divergent part of the one-loop diagrams of Fig. A.3. Let us also notice that we can replace λ with λ_{R} in Eq. (A27) since the difference will appear only at the higher order.

Until now we have succeeded in building a new, *renormalized* Lagrangian, and the resulting theory is free from divergences up to and including one-loop diagrams. It involves two new terms which change the coefficients of the ϕ^2 and ϕ^4 terms of the original Lagrangian. These terms are usually called *counter-terms*. They are the expression, in terms of the dimensional regularization cut-off parameter ϵ , of the process we outlined in Eqs. (A4) and (A5). They provide the correct definition, up to this order of perturbation, of the Lagrangian density, by removing the short-distance ambiguities inherent in the local expressions ϕ^2 and ϕ^4 .

Before looking at higher orders, let us see the price we had to pay for this achievement. It can be better seen at the four-point function. Looking back at the Eq. (A23), we make the following two observations. First, as we noticed already, the finite part seems to depend on a new arbitrary parameter with the dimensions of a mass μ . Second, the definition of Z_{λ} in Eq. (A33) also seems arbitrary. We

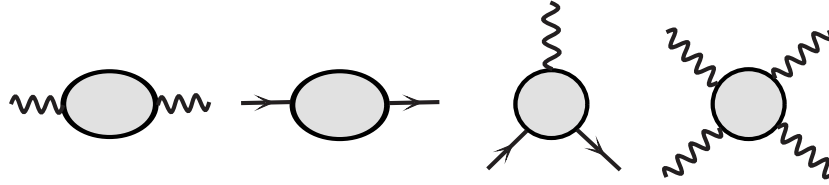


Fig. A.5: The primitively divergent 1PI Green's functions of QED. The last one, the light-by-light scattering, is convergent as a consequence of gauge invariance.

could add to it any term of the form $C\lambda_R$ with C any arbitrary constant independent of ϵ . Such an addition would change the value of the coupling constant at the one-loop order. The two observations are not unrelated. Indeed, changing the parameter μ from μ_1 to μ_2 in Eq. (A23) adds a constant term proportional to $\lambda \ln(\mu_1/\mu_2)$ which, as we just saw, can be absorbed in a redefinition of Z_λ and thus of the value of the coupling constant. This μ dependence can be studied systematically and gives rise to the renormalization group equation which I will not present here. We conclude that, at the one-loop level, all arbitrariness of the renormalization programme consists of assigning prescribed values to two parameters of the theory, which can be chosen to be the mass and the coupling constant. A convenient choice is given by two conditions of the form

$$\Gamma^{(2)}(p^2 = m_R^2) = 0 \quad (\text{A34})$$

and

$$\Gamma^{(4)}(p_1, \dots, p_4)|_{\text{point } M} = i\lambda_R^{(M)}. \quad (\text{A35})$$

The first one, Eq. (A34), defines the physical mass as the pole of the complete propagator. Although this choice is the most natural for physics, from a purely technical point of view, we could use any condition assigning a prescribed value to $\Gamma^{(2)}(p^2)$ at a fixed point $p^2 = M^2$, provided it is a point in which $\Gamma^{(2)}(p^2)$ is regular. Similarly, in the second condition Eq. (A35), by ‘point M ’ we mean some point in the space of the four momenta $p_i, i = 1, \dots, 4$, provided it is a point in which $\Gamma^{(4)}$ is regular. For a massive theory the point $p_i = 0$ is an example. Once these conditions are imposed, all Green functions at one loop are well defined and calculable. A final remark: at one loop no counter-term corresponding to the kinetic energy term $(\partial_\mu\phi)^2$ is needed. This is an accident of the one-loop for the ϕ^4 theory. It appears only at higher orders.

This process of removing the ambiguities by introducing counter-terms in the original Lagrangian can be extended to all orders of perturbation. The proof is rather complicated but essentially elementary. No new ideas are necessary. We must prove that, at any order, the terms appear with the correct combinatoric factor, even in the cases in which sub-diagrams are divergent to which counter-terms have already been assigned. At the end, all Green functions of a renormalizable theory, or any combination of renormalizable theories, are well defined and calculable.

As a second example, we shall present the renormalization for the one-loop diagrams of QED. The method is exactly the same and yields ‘renormalized’ values of the various terms which appear in the QED Lagrangian. Looking at the power-counting Eq. (A8), we see that the only possibly divergent 1PI diagrams with one loop are those of Fig. A.5. A simple calculation gives:

- the photon self-energy

$$\Gamma_{\mu\nu}^{(2,0)}(q) = \frac{2i\alpha}{3\pi} \frac{1}{\epsilon} (q_\mu q_\nu - q^2 g_{\mu\nu}) + \dots, \quad (\text{A36})$$

where $\alpha = e^2/4\pi$ is the fine-structure constant and the dots stand for finite terms;

– the electron self-energy

$$\Gamma^{(0,2)}(p) = \frac{i\alpha}{2\pi} \frac{1}{\epsilon} \not{p} - \frac{2i\alpha}{\pi} \frac{1}{\epsilon} m + \dots, \quad (\text{A37})$$

where we have suppressed spinor indices and, again, the dots stand for finite terms—we can see that in Eqs. (A36) and (A37) the divergent terms are monomials in the external momenta;

– the vertex function

$$\Gamma_\mu^{(1,2)}(p, p') = \frac{i\alpha}{2\pi} \frac{1}{\epsilon} e\gamma_\mu + \dots. \quad (\text{A38})$$

As before, all these divergences can be absorbed in the definition of renormalized quantities as

$$A^\mu(x) = Z_3^{1/2} A_R^\mu(x) = \left(1 - \frac{\alpha}{3\pi} \frac{1}{\epsilon} + \mathcal{O}(\alpha^2)\right) A_R^\mu(x), \quad (\text{A39})$$

$$\psi(x) = Z_2^{1/2} \psi_R(x) = \left(1 - \frac{\alpha}{4\pi} \frac{1}{\epsilon} + \mathcal{O}(\alpha^2)\right) \psi_R(x), \quad (\text{A40})$$

$$m = Z_m m_R = \left(1 - \frac{2\alpha}{\pi} \frac{1}{\epsilon} + \mathcal{O}(\alpha^2)\right) m_R, \quad (\text{A41})$$

$$\Gamma_\mu^{(1,2)}(p, p') = -ieZ_1\gamma_\mu + \dots = -ie\gamma_\mu \left(1 - \frac{\alpha}{2\pi} \frac{1}{\epsilon} + \mathcal{O}(\alpha^2)\right) + \dots. \quad (\text{A42})$$

As we noticed already, in QED the counter-terms corresponding to the kinetic energies of the electron and the photon appear already at the one-loop order. Putting all counter-terms together, the interaction Lagrangian becomes:

$$-e\bar{\psi}\gamma_\mu\psi A^\mu = -Z_e Z_2 Z_3^{1/2} e_R \bar{\psi}_R \gamma_\mu \psi_R A_R^\mu. \quad (\text{A43})$$

It follows that the condition which determines the charge renormalization constant Z_e is

$$Z_e Z_2 Z_3 = Z_1. \quad (\text{A44})$$

By comparing Eqs. (A42) and (A40), we see that, at least at this order, $Z_1 = Z_2$. Therefore, the entire charge renormalization is determined by the photon self-energy diagram. We can show that this property is valid to all orders of perturbation theory and is a consequence of gauge invariance. It is the same property of gauge invariance which guarantees that the last diagram of Fig. A.5, when computed using dimensional regularization which respects gauge invariance, is in fact finite.

This completes a very sketchy discussion of renormalization theory. Only straightforward calculations are needed to adapt it to any renormalizable theory and to any order in the perturbation expansion.

Lectures on Flavor Physics and CP Violation

B. Grinstein

University of California, San Diego, USA

Abstract

These lectures on flavor physics are an introduction to the subject. First lecture: We discuss the meaning of flavor and the importance of flavor physics in restricting extensions of the Standard Model (SM) of Electroweak interactions. We explain the origin of the KM matrix and how its elements are determined. We discuss FCNC and the GIM mechanism, followed by how a principle of Minimal Flavor Violation leads to SM extensions that are safe as far as FCNC are concerned even if the new physics comes in at low, TeVish scales. This is illustrated by the example of B radiative decays ($b \rightarrow s\gamma$). Second lecture: We then turn our attention to CP-violation. We start by presenting neutral meson mixing. Then we consider various CP-asymmetries, culminating in the theoretically clean interference between mixing and decay into CP eigenstates.

Keywords

Lectures; flavor; CP violation; CKM matrix; flavor changing neutral currents; GIM mechanism.

Preface

I created this document in preparation for lectures I am to present at the 8th CERN Latin American School of High Energy Physics (CLASHEP) during the Winter (Summer?) of 2015. These lectures are intended for graduate students of experimental particle physics. I aim at pedagogy, so don't look here for a complete list of topics, nor a complete set of references. Plainly, this document is not intended as a reference work. It is not complete, but rather introductory. My hope is that a physics student who has been exposed to the Standard Model of electroweak interactions will come out with an idea of why flavor physics remains one of the most vibrant areas in particle physics, both in theory and particularly in experiment. She or he will hopefully have an appreciation of the main aspects of the field and the crucial interconnections between theory and experiment that characterize it.

I started preparing this course as an adaptation of lectures I presented at TASI in 2013 and at Schladming in 2014. But because of the difference in scope and in audience I had to make major adjustments, definite choices on what to retain and what to omit. While some old hats may disagree with my choices, I am satisfied with the outcome and reasonably confident that the product will satisfy my customers. Of course, the jury is out. If you, the reader, happens to be one of those customers, I would really appreciate some feedback: email me, text me, call me, whatever (but beware, I don't Tweet). I hope to get invited to lecture somewhere again in the future, and your valuable opinion can help me improve as a lecturer.

Particle Physics has just entered an era of great excitement. You may not appreciate this if you live and work in the US, as government funding of the discipline erodes there, but its palpable in Physics departments of universities and laboratories around the world. This bodes well for the future of the field. I need not explain why it is that much of the excitement is coming from CERN. But CERN has not only become the leading laboratory of high energy physics in the world, it has also taken a leadership role in education, at least in areas that pertain the lab's disciplines. This makes sense. It is the youngsters of today that will be the researchers of that tomorrow. And these youngsters need training. The CLASHEP is but one of CERN's contribution to this effort. It gives students in Latin America a rare opportunity to study topics that are unlikely found in the curriculum at their institutions and to meet with other students

from Latin America and researcher-instructors from around the world. I feel privileged and honored that I have been given the opportunity to present these lectures on Flavor Physics and CP Violation and hope that the writeup of these lectures can be of use to many current and future students that may not have the good fortune of attending a CLASHEP.

Being lectures, there are lots of exercises that go with these. The exercises are interspersed in the material rather than collected at the end of chapters. The problems tend to expand or check on one point and I think it's best for a student to solve the exercises in context. I have many ideas for additional exercises, but only limited time. I hope to add some more in time and keep an update accesible on the web. Some day I will publish the solutions. Some are already typed into the TeX source and I hope to keep adding to it.

No one is perfect and I am certainly far from it. I would appreciate alert readers to send me any typos, errors or any other needed corrections they may find. Suggestions for any kind of improvement are welcome. I will be indebted if you'd send them to me at bgrinstein@ucsd.edu

Benjamín Grinstein
San Diego, February 2015

1 Flavor Theory

1.1 Introduction: What/Why/How?

WHAT:

There are six different types of quarks: u (“up”), d (“down”), s (“strange”), c (“charm”), b (“bottom”) and t (“top”). Flavor physics is the study of different types of quarks, or “flavors,” their spectrum and the transmutations among them. More generally different types of leptons, “lepton flavors,” can also be included in this topic, but in this lectures we concentrate on quarks and the hadrons that contain them.

WHY:

Flavor physics is very rich. You should have a copy of the PDG, or at least a bookmark to pdg.lbl.gov on your computer. A quick inspection of the PDG reveals that a great majority of content gives transition rates among hadrons with different quark content, mostly decay rates. This is the realm of flavor physics. We aim at understanding this wealth of information in terms of some simple basic principles. That we may be able to do this is striking endorsement of the validity of our theoretical model of nature, and gives stringent constraints on any new model of nature you may invent. Indeed, many models you may have heard about, in fact many of the most popular models, like gauge mediated SUSY breaking and walking technicolor, were invented to address the strong constraints imposed by flavor physics. Moreover, all observed CP violation (CPV) in nature is tied to flavor changing interactions, so understanding of this fundamental phenomenon is the domain of flavor physics.

HOW:

The richness of flavor physics comes at a price: while flavor transitions occur intrinsically at the quark level, we only observe transitions among hadrons. Since quarks are bound in hadrons by the strong interactions we face the problem of confronting theory with experiment in the context of mathematical models that are not immediately amenable to simple analysis, like perturbation theory. Moreover, the physics of flavor more often than not involves several disparate time (or energy) scales, making even dimensional analysis somewhere between difficult and worthless. Many tools have been developed to address these issues, and these lectures will touch on several of them. Among these:

- Symmetries allow us to relate different processes and sometimes even to predict the absolute rate of a transition.

- Effective Field Theory (EFT) allows to systematically disentangle the effects of disparate scales. Fermi theory is an EFT for electroweak interactions at low energies. Chiral Lagrangians encapsulate the information of symmetry relations of transitions among pseudo-Goldstone bosons. Heavy Quark Effective Theory (HQET) disentangles the scales associated with the masses of heavy quarks from the scale associated with hadron dynamics and makes explicit spin and heavy-flavor symmetries. And so on.
- Monte-Carlo simulations of strongly interacting quantum field theories on the lattice can be used to compute some quantities of basic interest that cannot be computed using perturbation theory.

1.2 Flavor in the Standard Model

Since the Standard Model of Strong and Electroweak interactions (SM) works so well, we will adopt it as our standard (no pun intended) paradigm. All alternative theories that are presently studied build on the SM; we refer to them collectively as Beyond the SM (BSM). Basing our discussion on the SM is very useful:

- It will allow us to introduce concretely the methods used to think about and quantitatively analyze Flavor physics. It should be straightforward to extend the techniques introduced in the context of the SM to specific BSM models.
- Only to the extent that we can make precise calculations in the SM and confront them with comparably precise experimental results can we meaningfully study effects of other (BSM) models.

So let's review the SM. At the very least, this allows us to agree on notation. The SM is a gauge theory, with gauge group $SU(3) \times SU(2) \times U(1)$. The $SU(3)$ factor models the strong interactions of “colored” quarks and gluons, $SU(2) \times U(1)$ is the famous Glashow-Weinberg-Salam model of the electroweak interactions. Sometimes we will refer to these as $SU(3)_c$ and $SU(2)_W \times U(1)_Y$ to distinguish them from other physical transformations characterized by the same mathematical groups. The matter content of the model consists of color triplet quarks: left handed spinor doublets q_L^i with $U(1)$ “hypercharge” $Y = 1/6$ and right handed spinor singlets u_R^i and d_R^i with $Y = 2/3$ and $Y = -1/3$. The color ($SU(3)$), weak ($SU(2)$), and Lorentz-transformation indices are implicit. The “ i ” index runs over $i = 1, 2, 3$ accounting for three copies, or “generations.” A more concise description is $q_L^i = (3, 2)_{1/6}$, meaning that q_L^i transforms as a **3** under $SU(3)$, a **2** under $SU(2)$ and has $Y = 1/6$ (the $U(1)$ charge). Similarly, $u_R^i = (3, 1)_{2/3}$ and $d_R^i = (3, 1)_{-1/3}$. The leptons are color singlets: $\ell_L^i = (1, 2)_{-1/2}$ and $e_R^i = (1, 1)_{-1}$.

We give names to the quarks in different generations:

$$q_L^i = \left(\begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \right), \quad u_R^i = (u_R, c_R, t_R), \quad d_R^i = (d_R, s_R, b_R). \quad (1)$$

Note that we have used the same symbols, “ u ” and “ d ,” to denote the collection of quarks in a generation and the individual elements in the first generation. When the superscript i is explicit this should give rise to no confusion. But soon we will want to drop the superscript to denote collectively the generations as vectors q_L , u_R and d_R , and then we will have to rely on the context to figure out whether it is the collection or the individual first element that we are referring to. For this reason some authors use the capital letters U_R and D_R to denote the vectors in generation space. But I want to reserve U for unitary transformations, and I think you should have no problem figuring out what we are talking about from context.

Similarly, for leptons we have

$$\ell_L^i = \left(\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \right), \quad e_R^i = (e_R, \mu_R, \tau_R). \quad (2)$$

The last ingredient of the SM is the Brout-Englert-Higgs (BEH) field, H , a collection of complex scalars transforming as $(1, 2)_{1/2}$. The BEH field has an expectation value, which we take to be

$$\langle H \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (3)$$

The hermitian conjugate field $\tilde{H} = i\sigma^2 H^*$ transforms as $(1, 2)_{-1/2}$ and is useful in constructing Yukawa interactions invariant under the electroweak group. The covariant derivative is

$$D_\mu = \partial_\mu + ig_s T^a A_\mu^a + ig_2 \frac{\sigma^j}{2} W_\mu^j + ig_1 Y B_\mu. \quad (4)$$

Here we have used already the Pauli σ^i matrices as generators of $SU(2)$, since the only fields which are non-singlets under this group are all doublets (and, of course, one should replace zero for σ^j above in the case of singlets). It should also be clear that we are using the generalized Einstein convention: the repeated index a is summed over $a = 1, \dots, N_c^2 - 1$, where $N_c = 3$ is the number of colors, and j is summed over $j = 1, 2, 3$. The generators T^a of $SU(3)$ are normalized so that in the fundamental representation $\text{Tr}(T^a T^b) = \frac{1}{2} \delta^{ab}$. With this we see that $\langle H \rangle$ is invariant under $Q = \frac{1}{2} \sigma^3 + Y$, which we identify as the generator of an unbroken $U(1)$ gauge group, the electromagnetic charge. The field strength tensors for A_μ^a , W_μ^j and B_μ are denoted as $G_{\mu\nu}^a$, $W_{\mu\nu}^j$, and $B_{\mu\nu}$, respectively, and that of electromagnetism by $F_{\mu\nu}$.

The Lagrangian of the SM is the most general combination of monomials (terms) constructed out of these fields constrained by (i) giving a hermitian Hamiltonian, (ii) Lorentz invariance, (iii) Gauge invariance, and (iv) renormalizability. This last one implies that these monomials, or ‘‘operators,’’ are of dimension no larger than four.¹ Field redefinitions by linear transformations that preserve Lorentz and gauge invariance bring the kinetic terms to canonical form. The remaining terms are potential energy terms, either Yukawa interactions or BEH-field self-couplings. The former are central to our story:

$$- \mathcal{L}_{\text{Yuk}} = \sum_{i,j} \left[\lambda_{U_j^i} \tilde{H} \bar{q}_{Li} u_R^j + \lambda_{D_j^i} H \bar{q}_{Li} d_R^j + \lambda_{E_j^i} H \bar{\ell}_{Li} e_R^j + \text{h.c.} \right] \quad (5)$$

We will mostly avoid explicit index notation from here on. The reason for upper and lower indices will become clear below. The above equation can be written more compactly as

$$- \mathcal{L}_{\text{Yuk}} = \tilde{H} \bar{q}_L \lambda_U u_R + H \bar{q}_L \lambda_D d_R + H \bar{\ell}_L \lambda_E e_R + \text{h.c.} \quad (6)$$

1.2.0.1 Flavor ‘‘symmetry.’’

In the absence of Yukawa interactions (*i.e.*, setting $\lambda_U = \lambda_D = \lambda_E = 0$ above) the SM Lagrangian has a large global symmetry. This is because the Lagrangian is just the sum of covariantized kinetic energy terms, $\sum_n \bar{\psi}_n i \not{D} \psi_n$, with the sum running over all the fields in irreducible representations of the the SM gauge group, and one can make linear unitary transformations among the fields in a given SM-representation without altering the Lagrangian:

$$q_L \rightarrow U_q q_L, \quad u_R \rightarrow U_u u_R, \quad \dots \quad e_R \rightarrow U_e e_R,$$

where $U_q^\dagger U_q = \dots = U_e^\dagger U_e = 1$. Since there are $N_f = 3$ copies of each SM-representation this means these are $N_f \times N_f$ matrices, so that for each SM-representation the redefinition freedom is by elements of the group $U(N_f)$. Since there are five distinct SM-representations (3 for quarks and 2 for leptons), the full

¹The action integral $S = \int d^4x \mathcal{L}$ has units of \hbar , and since we take $\hbar = 1$, the engineering dimensions of the Lagrangian density \mathcal{L} must be -4 .

symmetry group is $U(N_f)^5 = U(3)^{5,2}$. In the quantum theory each of the $U(1)$ factors (corresponding to a redefinition of the N_f fields in a given SM-representation by multiplication by a common phase) is anomalous, so the full symmetry group is smaller. One can make non-anomalous combinations of these $U(1)$'s, most famously $B - L$, a symmetry that rotates quarks and leptons simultaneously, quarks by $-1/3$ the phase of leptons. For our purposes it is the non-abelian factors that are most relevant, so we will be happy to restrict our attention to the symmetry group $SU(N_f)^5$.

The flavor symmetry is broken explicitly by the Yukawa interactions. We can keep track of the pattern of symmetry breaking by treating the Yukawa couplings as ‘‘spurions,’’ that is, as constant fields. For example, under $SU(N_f)_q \times SU(N_f)_u$ the first term in (6) is invariant if we declare that λ_U transforms as a bi-fundamental, $\lambda_U \rightarrow U_q \lambda_U U_u^\dagger$; check:

$$\bar{q}_L \lambda_U u_R \rightarrow \bar{q}_L U_q^\dagger (U_q \lambda_U U_u^\dagger) U_u u_R = \bar{q}_L \lambda_U u_R.$$

So this, together with $\lambda_D \rightarrow U_q \lambda_D U_d^\dagger$ and $\lambda_E \rightarrow U_\ell \lambda_E U_e^\dagger$ renders the whole Lagrangian invariant.

Why do we care? As we will see, absent tuning or large parametric suppression, *new interactions that break this ‘‘symmetry’’ tend to produce rates of flavor transformations that are inconsistent with observation*. This is not an absolute truth, rather a statement about the generic case.

In these lectures we will be mostly concerned with hadronic flavor, so from here on we focus on the $G_F \equiv SU(3)^3$ that acts on quarks.

1.3 The KM matrix and the KM model of CP-violation

Replacing the BEH field by its VEV, Eq. (3), in the Yukawa terms in (6) we obtain mass terms for quarks and leptons:

$$-\mathcal{L}_m = \frac{v}{\sqrt{2}} \bar{u}_L \lambda_U u_R + \frac{v}{\sqrt{2}} \bar{d}_L \lambda_D d_R + \frac{v}{\sqrt{2}} \bar{e}_L \lambda_E e_R + \text{h.c.} \quad (7)$$

For simpler computation and interpretation of the model it is best to make further field redefinitions that render the mass terms diagonal while maintaining the canonical form of the kinetic terms (diagonal, with unit normalization). The field redefinition must be linear (to maintain explicit renormalizability of the model) and commute with the Lorentz group and the part of the gauge group that is unbroken by the electroweak VEV (that is, the $U(1) \times SU(3)$ of electromagnetism and color). This means the linear transformation can act to mix only quarks with the same handedness and electric charge (and the same goes for leptons):

$$u_R \rightarrow V_{u_R} u_R, \quad u_L \rightarrow V_{u_L} u_L, \quad d_R \rightarrow V_{d_R} d_R, \quad d_L \rightarrow V_{d_L} d_L. \quad (8)$$

Finally, the linear transformation will preserve the form of the kinetic terms, say, $\bar{u}_L i \not{\partial} u_L \rightarrow (\bar{u}_L V_{u_L}^\dagger) i \not{\partial} (V_{u_L} u_L)$, if $V_{u_L}^\dagger V_{u_L} = 1$, that is, if they are unitary.

Now, choose to make these field redefinitions by matrices that diagonalize the mass terms,

$$V_{u_L}^\dagger \lambda_U V_{u_R} = \lambda'_U, \quad V_{d_L}^\dagger \lambda_D V_{d_R} = \lambda'_D. \quad (9)$$

Here the matrices with a prime, λ'_U and λ'_D , are diagonal, real and positive.

²Had we kept indices explicitly we would have written $q_L^i \rightarrow U_q^{i_j} q_L^j$, $u_R^i \rightarrow U_u^{i_j} u_R^j$, \dots , $e_R^i \rightarrow U_e^{i_j} e_R^j$. The fields transform in the fundamental representation of $SU(N_f)$. We use upper indices for this. Objects, like the hermitian conjugate of the fields, that transform in the anti-fundamental representation, carry lower indices. The transformation matrices have one upper and one lower indices, of course.

Exercises

Exercise 1.3-1: Show that this can always be done. That is, that an arbitrary matrix M can be transformed into a real, positive diagonal matrix $M' = P^\dagger M Q$ by a pair of unitary matrices, P and Q .

Then from

$$-\mathcal{L}_m = \frac{v}{\sqrt{2}} \left(\bar{u}_L \lambda'_U u_R + \bar{d}_L \lambda'_D d_R + \bar{e}_L \lambda_E e_R + \text{h.c.} \right) = \frac{v}{\sqrt{2}} \left(\bar{u} \lambda'_U u + \bar{d} \lambda'_D d + \bar{e} \lambda_E e \right) \quad (10)$$

we read off the diagonal mass matrices, $m_U = v \lambda'_U / \sqrt{2}$, $m_D = v \lambda'_D / \sqrt{2}$ and $m_E = v \lambda_E / \sqrt{2}$.

Since the field redefinitions in (8) are not symmetries of the Lagrangian (they fail to commute with the electroweak group), it is not guaranteed that the Lagrangian is independent of the matrices V_{u_L}, \dots, V_{d_R} . We did choose the transformations to leave the kinetic terms in canonical form. We now check the effect of (8) on the gauge interactions. Consider first the singlet fields u_R . Under the field redefinition we have

$$\bar{u}_R (g_s \mathbb{A}^a T^a + \frac{2}{3} g_1 \mathbb{B}) u_R \rightarrow \bar{u}_R V_{u_R}^\dagger (g_s \mathbb{A}^a T^a + \frac{2}{3} g_1 \mathbb{B}) V_{u_R} u_R = \bar{u}_R (g_s \mathbb{A}^a T^a + \frac{2}{3} g_1 \mathbb{B}) u_R .$$

It remains unchanged (you can see this by making explicit the so-far-implicit indices for color and for spinor components). Clearly the same happens with the d_R fields. The story gets more interesting with the left handed fields, since they form doublets. First let's look at the terms that are diagonal in the doublet space:

$$\begin{aligned} \bar{q}_L (g_s \mathbb{A}^a T^a + \frac{1}{2} g_2 \mathbb{W}^3 \sigma^3 + \frac{1}{6} g_1 \mathbb{B}) q_L \\ = \bar{u}_L (g_s \mathbb{A}^a T^a + \frac{1}{2} g_2 \mathbb{W}^3 + \frac{1}{6} g_1 \mathbb{B}) u_L + \bar{d}_L (g_s \mathbb{A}^a T^a - \frac{1}{2} g_2 \mathbb{W}^3 + \frac{1}{6} g_1 \mathbb{B}) d_L \end{aligned}$$

where in going to the second line we have expanded out the doublets in their components. The result is invariant under (8) very much the same way that the u_R and d_R terms are. Finally we have the off-diagonal terms. For these let us introduce

$$\sigma^\pm = \frac{\sigma^1 \pm i\sigma^2}{\sqrt{2}}, \quad \text{and} \quad W^\pm = \frac{W^1 \mp iW^2}{\sqrt{2}}$$

so that $\sigma^1 W^1 + \sigma^2 W^2 = \sigma^+ W^+ + \sigma^- W^-$ and $(\sigma^+)_{12} = \sqrt{2}$, $(\sigma^-)_{21} = \sqrt{2}$, and all other elements vanish. It is now easy to expand:

$$\begin{aligned} \bar{q}_L \frac{1}{2} g_2 (\sigma^1 W^1 + \sigma^2 W^2) q_L = \frac{1}{\sqrt{2}} g_2 \bar{u}_L W^+ d_L + \frac{1}{\sqrt{2}} g_2 \bar{d}_L W^- u_L \\ \rightarrow \frac{1}{\sqrt{2}} g_2 \bar{u}_L (V_{u_L}^\dagger V_{d_L}) W^+ d_L + \frac{1}{\sqrt{2}} g_2 \bar{d}_L (V_{d_L}^\dagger V_{u_L}) W^- u_L \quad (11) \end{aligned}$$

A relic of our field redefinitions has remained in the form of the unitary matrix $V = V_{u_L}^\dagger V_{d_L}$. We call this the Kobayashi-Maskawa (KM) matrix. You will also find this as the Cabibbo-Kobayashi-Maskawa, or CKM, matrix in the literature. Cabibbo figured out the 2×2 case, in which the matrix is orthogonal and given in terms of a single angle, the *Cabibbo angle*. Because Kobayashi and Maskawa were first to introduce the 3×3 version with an eye to incorporate CP violation in the model (as we will study in detail below), in these notes we refer to it as the KM matrix.

A general unitary 3×3 matrix has 3^2 complex entries, constrained by 3 complex plus 3 real conditions. So the KM matrix is in general parametrized by 9 real entries. But not all are of physical consequence. We can perform further transformations of the form of (8) that leave the mass matrices in (9) diagonal and non-negative if the unitary matrices are diagonal with $V_{u_L} = V_{u_R} = \text{diag}(e^{i\alpha_1}, e^{i\alpha_2}, e^{i\alpha_3})$

and $V_{dL} = V_{dR} = \text{diag}(e^{i\beta_1}, e^{i\beta_2}, e^{i\beta_3})$. Then V is redefined by $V_{ij} \rightarrow e^{i(\beta_j - \alpha_i)} V_{ij}$. These five independent phase differences reduce the number of independent parameters in V to $9 - 5 = 4$. It can be shown that this can in general be taken to be 3 rotation angles and one complex phase. It will be useful to label the matrix elements by the quarks they connect:

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}.$$

Observations:

1. That there is one irremovable phase in V implies that CP is not a symmetry of the SM Lagrangian. It is broken by the terms $\bar{u}_L V \not{W}^+ d_L + \bar{d}_L V^\dagger \not{W}^- u_L$. To see this, recall that under CP $\bar{u}_L \gamma^\mu d_L \rightarrow -\bar{d}_L \gamma^\mu u_L$ and $W^{+\mu} \rightarrow -W_\mu^-$. Hence CP invariance requires $V^\dagger = V^T$.

Exercises

Exercise 1.3-2: In QED, charge conjugation is $\bar{e} \gamma^\mu e \rightarrow -\bar{e} \gamma^\mu e$ and $A^\mu \rightarrow -A^\mu$. So $\bar{e} \not{A} e$ is invariant under C .



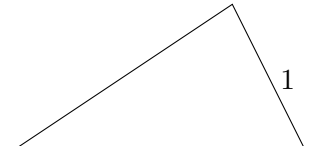
So what about QCD? Under charge conjugation $\bar{q} T^a \gamma^\mu q \rightarrow \bar{q} (-T^a)^T \gamma^\mu q$, but $(-T^a)^T = (-T^a)^*$ does not equal $-T^a$ (nor T^a). So what does charge conjugation mean in QCD? How does the gluon field, A_μ^a , transform?

Exercise 1.3-3: If two entries in m_U (or in m_D) are equal show that V can be brought into a real matrix and hence is an orthogonal transformation (an element of $O(3)$).

2. Precise knowledge of the elements of V is necessary to constrain new physics (or to test the validity of the SM/CKM theory). We will describe below how well we know them and how. But for now it is useful to have a sketch that gives a rough order of magnitude of the magnitude of the elements in V :

$$V \sim \begin{pmatrix} \epsilon^0 & \epsilon^1 & \epsilon^3 \\ \epsilon^1 & \epsilon^0 & \epsilon^2 \\ \epsilon^3 & \epsilon^2 & \epsilon^0 \end{pmatrix}, \quad \text{with } \epsilon \sim 10^{-1}. \quad (12)$$

3. Since $V V^\dagger = V^\dagger V = 1$ the rows as well as the columns of V are orthonormal vectors. In particular, $\sum_k V_{ik} V_{jk}^* = 0$ for $j \neq i$. Three complex numbers that sum to zero are represented on the complex plane as a triangle. As the following table shows, the resulting triangles are very different in shape. Two of them are very squashed, with one side much smaller than the other two, while the third one has all sides of comparable size. As we shall see, this will play a role in understanding when CP asymmetries in decay rates can be sizable.

ij	$\sum V_{ik} V_{jk}^* = 0$	$\sim \epsilon^n$	shape (normalized to unit base)
12	$V_{ud} V_{cd}^* + V_{us} V_{cs}^* + V_{ub} V_{cb}^* = 0$	$\epsilon + \epsilon + \epsilon^5 = 0$	 ϵ^4
23	$V_{cd} V_{td}^* + V_{cs} V_{ts}^* + V_{cb} V_{tb}^* = 0$	$\epsilon^4 + \epsilon^2 + \epsilon^2 = 0$	 ϵ^2
13	$V_{ud} V_{td}^* + V_{us} V_{ts}^* + V_{ub} V_{tb}^* = 0$	$\epsilon^3 + \epsilon^3 + \epsilon^3 = 0$	 1

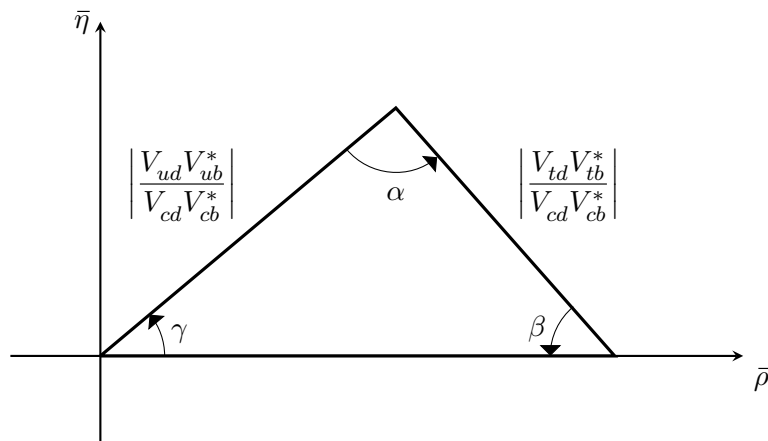


Fig. 1: Unitarity triangle in the $\bar{\rho}$ - $i\bar{\eta}$ plane. The base is of unit length. The sense of the angles is indicated by arrows.

These are called “unitarity triangles.” The most commonly discussed is in the 1-3 *columns*,

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \quad \Rightarrow \quad \begin{array}{c} \sim 1 \\ \triangle \\ \sim 1 \\ 1 \end{array}$$

Dividing by the middle term we can be more explicit as to what we mean by the unit base unitarity triangle:

$$\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} + 1 + \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} = 0$$

We draw this on the complex plane and introduced some additional notation: the complex plane is $z = \bar{\rho} + i\bar{\eta}$ and the internal angles of the triangle are³ α , β and γ ; see Fig. 1.

The angles of the unitarity triangle, of course, are completely determined by the KM matrix, as you will now explicitly show:

Exercises

Exercise 1.3-4: Show that

(i) $\beta = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right)$, $\alpha = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right)$ and $\gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right)$.

(ii) These are invariant under phase redefinitions of quark fields (that is, under the remaining arbitrariness). Hence these are candidates for observable quantities.

(iii) The area of the triangle is $-\frac{1}{2} \text{Im} \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} = -\frac{1}{2} \frac{1}{|V_{cd}V_{cb}^*|^2} \text{Im}(V_{ud}V_{cd}^*V_{cb}V_{ub}^*)$.

(iv) The product $J = \text{Im}(V_{ud}V_{cd}^*V_{cb}V_{ub}^*)$ (a “Jarlskog invariant”) is also invariant under phase redefinitions of quark fields.

Note that $\text{Im}(V_{ij}V_{kl}V_{il}^*V_{kj}^*) = J(\delta_{ij}\delta_{kl} - \delta_{il}\delta_{kj})$ is the common area of all the un-normalized triangles. The area of a normalized triangle is J divided by the square of the magnitude of the side that is normalized to unity.

³This convention is popular in the US, while in Japan a different convention is more common: $\phi_1 = \beta$, $\phi_2 = \alpha$ and $\phi_3 = \gamma$.

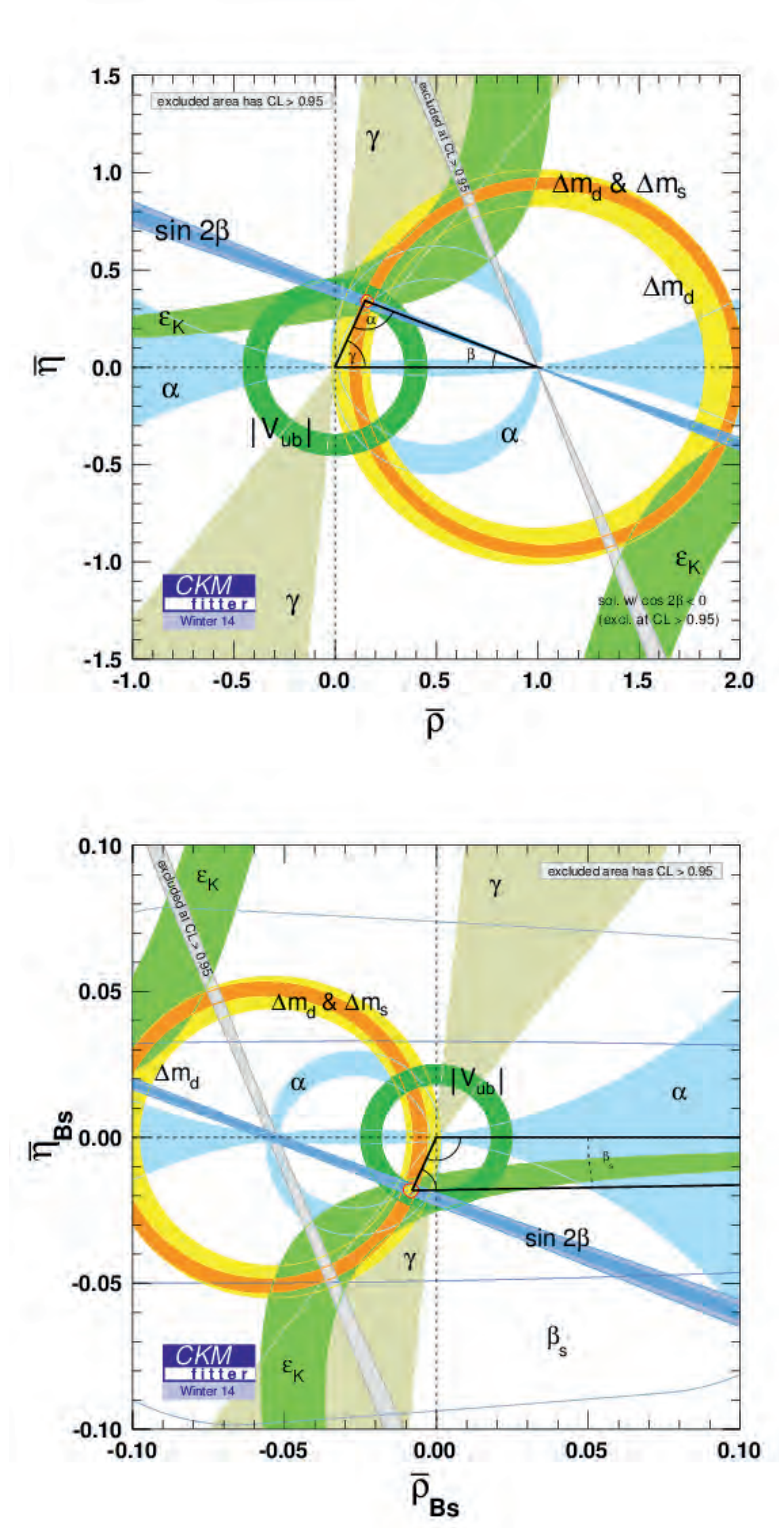


Fig. 2: Experimentally determined unitarity triangles [1]. Upper pane: "fat" 1-3 columns triangle. Lower pane: "skinny" 2-3 columns triangle.

4. *Parametrization of V* : Since there are only four independent parameters in the matrix that contains 3×3 complex entries, it is useful to have a completely general parametrization in terms of four parameters. The standard parametrization can be understood as a sequence of rotations about the three axes, with the middle rotation incorporating also a phase transformation:

$$V = CBA,$$

where

$$A = \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix}.$$

Here we have used the shorthand, $c_{ij} = \cos \theta_{ij}$, $s_{ij} = \sin \theta_{ij}$, where the angles θ_{ij} all lie on the first quadrant. From the phenomenologically observed rough order of magnitude of elements in V in (12) we see that the angles θ_{ij} are all small. But the phase δ is large, else all triangles would be squashed.

An alternative and popular parametrization is due to Wolfenstein. It follows from the above by introducing parameters A , λ , ρ and η according to

$$s_{12} = \lambda, \quad s_{23} = A\lambda^2, \quad s_{13}e^{i\delta} = A\lambda^3(\rho + i\eta) \quad (13)$$

The advantage of this parametrization is that if λ is of the order of ϵ , while the other parameters are of order one, then the KM matrix elements have the rough order in (12). It is easy to see that ρ and η are very close to, but not quite, the coordinates of the apex of the unitarity triangle in Fig. 1. One can adopt the alternative, but tightly related parametrization in terms of A , λ , $\bar{\rho}$ and $\bar{\eta}$:

$$s_{12} = \lambda, \quad s_{23} = A\lambda^2, \quad s_{13}e^{i\delta} = A\lambda^3(\bar{\rho} + i\bar{\eta}) \frac{\sqrt{1 - A^2\lambda^4}}{\sqrt{1 - \lambda^2[1 - A^2\lambda^4(\bar{\rho} + i\bar{\eta})]}}.$$

Exercises

Exercise 1.3-5: (i) Show that

$$\bar{\rho} + i\bar{\eta} = -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*},$$

hence $\bar{\rho}$ and $\bar{\eta}$ are indeed the coordinates of the apex of the unitarity triangle and are invariant under quark phase redefinitions.

(ii) Expand in $\lambda \ll 1$ to show

$$V = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4)$$

1.4 Determination of KM Elements

Figure 2 shows the state of the art in our knowledge of the angles of the unitarity triangles for the 1-3 and 2-3 columns of the KM matrix. How are these determined? More generally, how are KM elements measured? Here we give a tremendously compressed description.

The relative phase between elements of the KM matrix is associated with possible CP violation. So measurement of rates for processes that are dominated by one entry in the KM are insensitive to the relative phases. Conversely, CP asymmetries directly probe relative phases.

1.4.1 Magnitudes

The magnitudes of elements of the KM matrix are measured as follows:

- (i) $|V_{ud}|$ is measured through allowed nuclear transitions. The theory is fairly well understood (even if it is nuclear physics) because the transition matrix elements are constrained by symmetry considerations.
- (ii) $|V_{us}|, |V_{cd}|, |V_{cs}|, |V_{ub}|, |V_{cb}|$, are primarily probed through semi-leptonic decays of mesons, $M \rightarrow M' \ell \nu$ (e.g., $K^+ \rightarrow \pi^0 e^+ \nu$).
- (iii) $|V_{tq}|$, ($q = d, s, b$) are inferred from processes that proceed at 1-loop through a virtual top-quark. It is also possible to measure some of these directly from single top production (or decay).

The theoretical difficulty is to produce a reliable estimate of the rate, in terms of the KM matrix elements, in light of the quarks being strongly bound in hadrons. Moreover, theorists have to produce a good estimate for a quantity that experimentalists can measure. There is some tension between these. We will comment on this again below, but let me give one example. The inclusive rate for semileptonic decay of B mesons can be reliably calculated. By inclusive we mean B decays to a charged lepton, say μ , plus a neutrino, plus other stuff, and the rate is measured regardless of what the other stuff is. The decay rate is then the sum over the rates of decays into any particular type of whatever makes up the “stuff.” Sometimes the decay product is a D meson, sometimes a D^* meson and other times seven pions or whatever, always plus $\mu\nu$. Now these decays sometimes involve $b \rightarrow c\mu\nu$ which comes in the rate with a factor of $|V_{cb}|^2$ that we would like to determine, and sometimes involves $b \rightarrow u\mu\nu$ with a factor of $|V_{ub}|^2$ that we also want to determine. But the total semileptonic rate does not allow us to infer separately $|V_{cb}|^2$ and $|V_{ub}|^2$. Knowing that $|V_{cb}|^2 \gg |V_{ub}|^2$ means we can measure well $|V_{cb}|$ from the inclusive semileptonic rate. But then how do we get at $|V_{ub}|$? One possibility, and that was the first approach at this measurement, is to measure the rate of inclusive semileptonic B decays only for large μ energy. Since hadrons containing charm are far heavier than those containing up-quarks, there is a range of energies for the μ resulting from the decay that is not possible if B decayed into charm. These must go through $b \rightarrow u\mu\nu$ and therefore their rate is proportional to $|V_{ub}|^2$. But this is not an inclusive rate, because it does not sum over all possible decay products. It is difficult to get an accurate theoretical prediction for this.

The determination of magnitudes is usually done from semi-leptonic decays because the theory is more robust than for hadronic decays. Purely leptonic decays, as in $B^- \rightarrow \mu^- \bar{\nu}$ are also under good theoretical control, but their rates are very small because they are helicity suppressed in the SM (meaning that the “ $V - A$ ” nature of the weak interactions, $V =$ vector, $A =$ axial, gives a factor of m_μ/m_b in the decay amplitude). We lump them into the category of “rare” decays and use them, with an independent determination of the KM elements, to test the accuracy of the SM and put bounds on new physics. We distinguish exclusive from inclusive semileptonic decay measurements:

1.4.2 Exclusive semileptonic decays

By an “exclusive” decay we mean that the final state is fixed as in, for example, $B \rightarrow D\pi e\nu$. To appreciate the theoretical challenge consider the decay of a pseudoscalar meson to another pseudoscalar meson. The weak interaction couples to a $V - A$ hadronic current, $\bar{\psi}'(\gamma^\mu - \gamma^\mu\gamma_5)\psi$, and a corresponding leptonic current; see Eq. (11). The probability amplitude for the transition is given by

$$\mathcal{A} = \langle M' \ell \nu | \frac{g_2^2 V_{ij}}{M_W^2} \bar{u}_L^i \gamma^\mu d_L^j \bar{e}_L \gamma_\mu \nu_L | M \rangle.$$

The leptonic current, being excluded from the strong interactions, offers no difficulty and we can immediately compute its contribution to the amplitude. The contribution to the amplitude from the hadronic

side then involves

$$\langle \vec{p}' | V^\mu | \vec{p} \rangle = f_+(q^2)(p + p')^\mu + f_-(q^2)q^\mu, \quad (14)$$

where $V^\mu = \bar{u}^i \gamma^\mu d^j$ and $q = p - p'$. The bra and ket stand for the meson final and initial states, characterized only by their momentum and internal quantum numbers, which are implicit in the formula. The matrix element is to be computed non-perturbatively with regard to the strong interactions. Only the vector current (not the axial) contributes, by parity symmetry of the strong interactions. The expression on the right-hand-side of (14) is the most general function of p and p' that is co-variant under Lorentz transformations (*i.e.*, transforms as a four vector). It involves the coefficients f_\pm , or “form factors,” that are a function of q^2 only, since the other invariants are fixed ($p^2 = m_M^2$ and $p'^2 = m_{M'}^2$). In the 3-body decay, $p = p' + q$ so q is the sum of the momenta of the leptons. It is conventional to write the form factors as functions of q^2 . When the term $f_-(q^2)q^\mu$ is contracted with the leptonic current one gets a negligible contribution, $q \cdot (V - A) \sim m_\ell$, when $\ell = e$ or μ . So the central problem is to determine f_+ . Symmetry considerations can produce good estimates of f_+ at specific kinematic points, which is sufficient for the determination of the magnitude of the KM matrix elements. Alternatively one may determine the form factor using Monte Carlo simulations of QCD on the lattice.

Exercises

Exercise 1.4.2-1: Show that $q \cdot (V - A) \sim m_\ell$ for the leptonic charged current. Be more precise than “ \sim .”

To see how this works, consider a simpler example first. We will show that the electromagnetic form factor for the pion is determined by the charge of the pion at $q^2 = 0$. Take J^μ to be the electromagnetic current of light quarks, $J^\mu(x) = \frac{2}{3}\bar{u}(x)\gamma^\mu u(x) - \frac{1}{3}\bar{d}(x)\gamma^\mu d(x)$. Charge conservation means $\partial_\mu J^\mu = 0$. Now, the matrix element of this between pion states is

$$\langle \pi(\vec{p}') | J^\mu(0) | \pi(\vec{p}) \rangle = f_+(q^2)(p + p')^\mu + f_-(q^2)q^\mu \quad (15)$$

Restoring the x dependence in J^μ is easy, $J^\mu(x) = e^{i\hat{P}\cdot x} J^\mu(0) e^{-i\hat{P}\cdot x}$ where \hat{P}^μ is the 4-momentum operator. This just gives the above times $\exp(-iq \cdot x)$. Hence the matrix element of the divergence of J^μ is just the above contracted with q^μ . But $\partial_\mu J^\mu = 0$ so we have

$$f_+(q^2)(p + p') \cdot q + f_-(q^2)q^2 = 0$$

The first term has $(p + p') \cdot q = (p + p') \cdot (p - p') = p^2 - p'^2 = m_\pi^2 - m_\pi^2 = 0$ so we have $f_-(q^2) = 0$. Moreover, the electric charge operator is

$$\hat{Q} = \int d^3x J^0(x)$$

and we should have

$$\langle \pi(\vec{p}') | \hat{Q} | \pi(\vec{p}) \rangle = Q_\pi \langle \pi(\vec{p}') | \pi(\vec{p}) \rangle = Q_\pi (2\pi)^3 2E \delta^{(3)}(\vec{p} - \vec{p}') \quad (16)$$

where Q_π is the charge of the π state (± 1 for a π^\pm and 0 for a π^0) and we have used the relativistic normalization of states. Integrating the time component of (15) to compute the matrix element of \hat{Q} is the same as inserting a factor of

$$\int d^3x e^{-iq\cdot x} = (2\pi)^3 \delta^{(3)}(\vec{p} - \vec{p}')$$

into the left hand side of (15) and comparing both sides we have

$$2EQ_\pi = f_+(q^2)(E + E')$$

or $f_+(0) = Q_\pi$ since the condition $\vec{p}' = \vec{p}$ for equal mass particles gives $E' = E$ and therefore $q^\mu = 0$. To recap, conservation of J^μ implies $f_-(q^2) = 0$ and $f_+(0) = \pm 1$ for charged pions, $f_+(0) = 0$ for neutral pions.

$K \rightarrow \pi \ell \nu$: One can repeat this for kaons and pions, where the symmetry now is Gell-Mann's flavor- $SU(3)$. Let me remind you of this, so you do not confuse this ‘‘flavor’’ symmetry with the ‘‘flavor’’ symmetry we introduced earlier. If we want to understand the behavior of matter at energies sufficiently high that kaons are produced but still too low to produce charmed states, we can use for the Lagrangian

$$\mathcal{L} = \bar{u}i\not{D}u + \bar{d}i\not{D}d + \bar{s}i\not{D}s$$

where the covariant derivative only contains the gluon field. Electromagnetic and weak interactions have to be added as perturbations. The Lagrangian is invariant under the $SU(3)$ group of transformations in which the u , d and s quarks form a triplet: if $q = (u, d, s)^T$, the symmetry is $q \rightarrow Uq$ with U a unitary 3×3 matrix. The pions and kaons, together with the η particle form an octet of $SU(3)$: the 3×3 traceless matrix

$$M = \begin{pmatrix} \frac{\pi^0}{\sqrt{2}} - \frac{\eta}{\sqrt{6}} & \pi^+ & K^+ \\ \pi^- & -\frac{\pi^0}{\sqrt{2}} - \frac{\eta}{\sqrt{6}} & K^0 \\ K^- & \bar{K}^0 & \frac{\eta}{\sqrt{3}} \end{pmatrix}.$$

The flavor quantum numbers of these are in 1-to-1 correspondance with the matrix $q \times \bar{q}^T$. In particular note that the 2-3 element, the K^0 , has content $q_2\bar{q}_3 = d\bar{s}$: kaons have strangeness -1 , while anti-kaons have strangeness $+1$. Symmetry means that the quantum mechanical probability amplitudes (a.k.a. matrix elements) have to be invariant under $M \rightarrow U M U^\dagger$. The symmetry implies $f_-(q^2) = 0$ and $f_+(0) = 1$ for the form factors of the conserved currents associated with the $SU(3)$ symmetry transformations. In reality, however, this symmetry does not hold as accurately as isospin. A better Lagrangian includes masses for the quarks, and masses vary among the quarks, breaking the symmetry:

$$\mathcal{L} = \bar{u}(i\not{D} - m_u)u + \bar{d}(i\not{D} - m_d)d + \bar{s}(i\not{D} - m_s)s$$

Since the largest source of symmetry breaking is the mass of the strange quark ($m_s \gg m_d \gtrsim m_u$), one expects corrections to $f_+(0) = 1$ of order m_s . But since f_+ is dimensionless the correction must be relative to some scale, $f_+(0) - 1 \propto m_s/\Lambda$, with Λ a hadronic scale, say, $\Lambda \sim 1$ GeV. This seems like bad news, an uncontrolled 10% correction. Fortunately, by a theorem of Ademollo and Gatto, the symmetry breaking parameter appears at second order, $f_+(0) - 1 \propto (m_s/\Lambda)^2 \sim 1\%$. Combining data for neutral and charged semi-leptonic K decays the PDG gives $|V_{us}|f_+(0) = 0.2163 \pm 0.0005$ [2] which to a few percent can be read off as the value of the magnitude of the KM matrix element. Monte-Carlo simulations of QCD on a lattice give a fairly accurate determination of the form factor; the same section of the PDG reports $f_+(0) = 0.960 \pm 0.005$ which it uses to give $|V_{us}| = 0.2253 \pm 0.0008$. Note that the theoretical calculation of f_+ is remarkably accurate, about at the half per-cent level. The reason this accuracy can be achieved is that one only needs to calculate the deviation of $f_+(0)$ from unity, an order $(m_s/\Lambda)^2$ effect, with moderate accuracy.

$B \rightarrow D \ell \nu$: We cannot extend this to the heavier quarks because then $m_c/\Lambda > 1$ is a bad expansion parameter. Remarkably, for transitions among heavy quarks there is another symmetry, dubbed ‘‘Heavy Quark Symmetry’’ (HQS), that allows similarly successful predictions; for a basic introduction see [3]. For transitions from a heavy meson (containing a heavy quark, like the B or D mesons) to a light meson (made exclusively of light quarks, like the π or K mesons) one requires other methods, like lattice QCD, to determine the remaining KM matrix elements.

A word about naming of mesons. Since K^0 by convention has strangeness -1 , we take by analogy B^0 to have bottomness (or beauty, in Europe) -1 . So the flavor quantum numbers of heavy mesons are $\bar{B}^0 = b\bar{d}$, $B^- = b\bar{u}$, $\bar{B}_s = b\bar{s}$, $D^0 = c\bar{u}$, $D^+ = c\bar{d}$, $D_s = c\bar{s}$.

Here is an elementary, mostly conceptual, explanation of how HQS works. The heavy mesons are composed of a quark that is very heavy compared to the binding energy of mesons, plus a light anti-quark making the whole thing neutral under color, plus a whole bunch of glue and quark-antiquark pairs. This “brown muck” surrounding and color-neutralizing the heavy quark is complicated and we lack good, let alone precise, mathematical models for it. The interactions of this brown muck have low energy compared to the mass of the heavy quark, so that they do not change the state of motion of the heavy quark: in the rest frame of the meson, the heavy quark is at rest. The central observation of HQS is that all the brown muck sees is a static source of color, regardless of the heavy quark mass. Hence there is a symmetry between B mesons and D mesons: they have the same brown muck, only different static color sources. A useful analogy to keep in mind is from atomic physics: the chemical properties of different isotopes of the same element are the same to high precision because the electronic cloud (the atomic brown muck) does not change even as the mass of the atomic nucleus (the atomic heavy quark) changes.

To put this into equations, we start by characterizing the heavy meson state by its velocity rather than its momentum, $v^\mu = p^\mu/m$. That is because we are considering the limit of infinite mass of the heavy quark, $m \rightarrow \infty$. Notice that infinite mass does not mean the meson is at rest. You can boost to a frame where it moves. More interestingly, even if both b and c quarks are infinitely heavy, the process $b \rightarrow c\ell\nu$ can produce a moving c quark in the rest-frame of the decaying b -quark. Another trivial complication is that the relativistic normalization of states, as in (16), includes a factor of energy, $E \rightarrow \infty$. So we take $|\vec{v}\rangle = (1/\sqrt{m})|\vec{p}\rangle$. For the application of the HQS it is more convenient (and natural) to parametrize the matrix element of the vector current in terms of the 4-velocities. Doing so, and using an argument analogous to that introduced previously to show $f_-(q^2) = 0$, we have

$$\langle \vec{v}' | V^\mu | \vec{v} \rangle = \xi(v \cdot v')(v + v')^\mu.$$

Comments: (i) the infinitely heavy states could be two same flavored mesons with a flavor diagonal current, e.g., $B^- \rightarrow B^-$ with $V^\mu = \bar{b}\gamma^\mu b$, or two different flavors with an off-diagonal current, e.g. $B^- \rightarrow D^0$ with $V^\mu = \bar{c}\gamma^\mu b$; (ii) the form factor, now labeled ξ and called an “Isgur-Wise” function, is in principle a function of the three Lorentz invariants we can make out of the 4-vectors v^μ and v'^μ , but since $v^2 = v'^2 = 1$ it only depends on $v \cdot v'$; (iii) rewriting this in terms of 4-momenta gives a relation between f_+ and f_- (but not $f_- = 0$); and, most importantly, (iv) the analogue to $f_+(0) = 1$ is

$$\xi(1) = 1.$$

Note that $v \cdot v' = 1$ corresponds to the resulting meson not moving relative to the decaying one (in other words, remaining at rest in the rest frame of the decaying meson), so that the invariant mass of the lepton pair, q^2 , is as large as it can be: $v \cdot v' = 1$ is $q^2 = q_{\max}^2 = (m_B - m_D)^2$.

The analogue of the theorem of Ademolo and Gato for HQS is Luke’s theorem [4]. It states that the corrections to the infinite mass predictions for form factors at $v \cdot v' = 1$ first appear at order $1/m^2$ rather than the naïvely expected $1/m$.

The prediction of the $B \rightarrow D$ form factors at one kinematic point ($q^2 = q_{\max}^2$) can be used to experimentally determine $|V_{cb}|$. Again a tension arises between theory and experiment: at the best theory point ($q^2 = q_{\max}^2$) the decay rate vanishes. In practice this problem is circumvented by extrapolating from $q^2 < q_{\max}^2$ and by including $B \rightarrow D^*\ell\nu$ in the analysis. The D^* is the spin-1 partner of the D meson. We have not explained this here, but HQS relates the D to the D^* mesons: they share a common brown muck. The reason is simple, the spin of the heavy quark interacts with the brown muck via a (chromo-)magnetic interaction, but magnetic moments are always of the form charge-over-mass, g/m , so they vanish at infinite mass. We can combine the spin- $\frac{1}{2}$ heavy quark with the spin- $\frac{1}{2}$ brown muck in a spin-0 or a spin-1 state, and since the spin does not couple, they have the same mass and the same matrix elements (form factors).

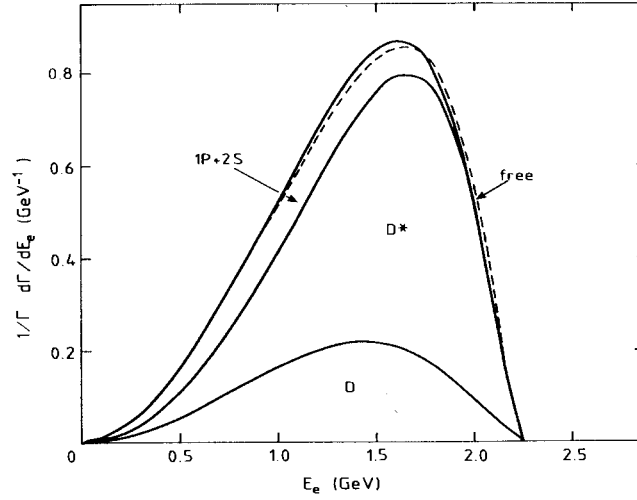


Fig. 3: Quark-hadron duality in $B \rightarrow X_c e \nu$ in a non-relativistic model of mesons. The figure, taken from [5], shows how the spectrum with respect to the electron energy normalized to the total semileptonic width, $\frac{1}{\Gamma} \frac{d\Gamma}{dE_e}$, is built up from exclusive decays. The lowest solid line is the contribution from $B \rightarrow D e \nu$, the next higher one includes the D^* final state and the highest one is the total contribution from all 1S, 1P and 2S states. The dashed line corresponds to the free quark $b \rightarrow c \ell \nu$ rate.

Exercises

Exercise 1.4.2-2: For $B \rightarrow D \ell \nu$ write the form factors $f_{\pm}(q^2)$ in terms of the Isgur-Wise function. What does $\xi(1) = 1$ imply for f_{\pm} ? Eliminate the Isgur-Wise function to obtain a relation between f_+ and f_- .

1.4.3 Inclusive semileptonic decays

As we have said, the inclusive semileptonic decay rate $\Gamma(\bar{B} \rightarrow X \ell \nu)$ means the rate of decay of a \bar{B} to $\ell \nu$ plus anything. We further distinguish $\Gamma(\bar{B} \rightarrow X_c \ell \nu)$ when the anything contains a charm quark and therefore the underlying process at the quark level is $b \rightarrow c \ell \nu$ and similarly $\Gamma(\bar{B} \rightarrow X_u \ell \nu)$ from $b \rightarrow u \ell \nu$.

There is good reason to believe that *quark-hadron duality* holds for these quantities. Quark-hadron duality means that instead of computing the rate for the transition between hadrons, in this case mesons, we can compute the rate for the transition between quarks and the answer is the same, $\Gamma(\bar{B} \rightarrow X_c \ell \nu) = \Gamma(b \rightarrow c \ell \nu)$. Figure 3 shows in solid curves how the spectrum with respect to the electron energy, $d\Gamma(B \rightarrow X e \nu)/dE_e$, builds up from exclusive modes, starting with $B \rightarrow D e \nu$ and adding to it $B \rightarrow D^* e \nu$ and then the sum of all 1S, 1P and 2S states. By comparison the $b \rightarrow c e \nu$ spectrum is shown as a dashed line. The agreement between the sum over exclusives and the free quark decay is apparent. By comparison Fig. 4 shows the $b \rightarrow u e \nu$ case. To reproduce the free quark rate many more states must be included.

Notice that the endpoint of the spectrum for $B \rightarrow X_u e \nu$ extends beyond that of $B \rightarrow X_c e \nu$. This was the basis for early determinations of $|V_{ub}|$, as mentioned above. The point is that $|V_{ub}| \ll |V_{cb}|$ so the $b \rightarrow u e \nu$ transition hides under $b \rightarrow c e \nu$ for most electron energies. But the theoretical determination of the spectrum constrained to the narrow region close to the end of the spectrum is not accurate. Modern determinations of $|V_{ub}|$ rely on summing over precise measurements of exclusive non-charm decay exclusive modes over the whole spectrum and using kinematic variables other than E_e .

Remarkably, quark-hadron duality for semileptonic heavy quark decays can be established from

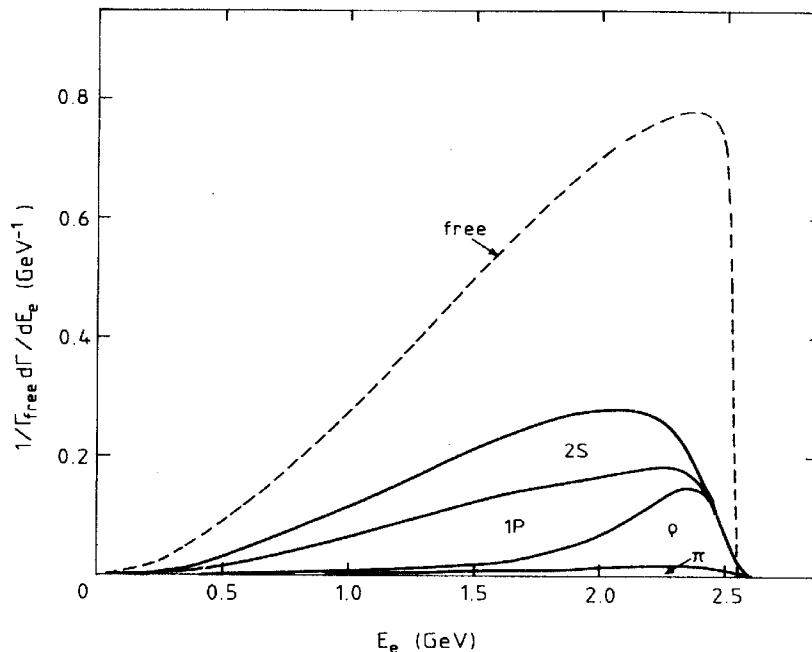


Fig. 4: As in Fig. 3 but for $b \rightarrow ue\nu$, from [5].

first principles using HQS [6]. Moreover, finite mass corrections can be systematically incorporated [7, 8]. Theory gives solid predictions for moments of the spectrum in terms of few unknown non-perturbative parameters that can be accurately fit to experiment [9], resulting in a determination at about 1% precision.

The green ring in Fig. 2 shows the region of the $\bar{\rho}-\bar{\eta}$ plane allowed by the determination of $|V_{ub}|$. More precisely, note that $\sqrt{\rho^2 + \eta^2} = |V_{ub}/V_{us}V_{cb}|$ so that the ring requires the determination of the three KM elements. It is labeled “ $|V_{ub}|$ ” because this is the least accurately determined of the three KM elements required.

1.4.4 Collecting results

While we have not presented a full account of the measurements and theory that are used in the determination of the KM magnitudes, by now you should have an idea of the variety of methods employed.

The PDG gives for the full fit of the magnitudes of the KM matrix elements

$$|V| = \begin{pmatrix} 0.97427 \pm 0.00014 & 0.22536 \pm 0.00061 & 0.00355 \pm 0.00015 \\ 0.22522 \pm 0.00061 & 0.97343 \pm 0.00015 & 0.0414 \pm 0.0012 \\ 0.00886^{+0.00033}_{-0.00032} & 0.0405^{+0.0011}_{-0.0012} & 0.99914 \pm 0.00005 \end{pmatrix},$$

or, in terms of the Wolfenstein parameters,

$$\begin{aligned} \lambda &= 0.22537 \pm 0.00061, & A &= 0.814^{+0.023}_{-0.024}, \\ \bar{\rho} &= 0.117 \pm 0.021, & \bar{\eta} &= 0.353 \pm 0.013. \end{aligned}$$

It also gives, for the Jarlskog determinant, $J = (3.06^{+0.21}_{-0.20}) \times 10^{-5}$.

1.4.5 Angles

The angles of the unitarity triangle are associated with CP violation. Next chapter is devoted to this. Here is a brief summary to two routes to their determination:

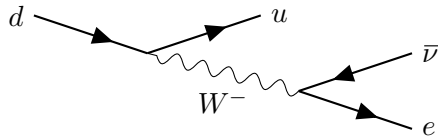
- (i) Neutral Meson Mixing. It gives, for example, $V_{tb}V_{td}^*$ in the case of B_d mixing and $V_{tb}V_{ts}^*$ for B_s mixing. The case of K^0 mixing is, as we will see, more complex. The yellow (“ Δm_d ”) and orange (“ Δm_d & Δm_s ”) circular rings centered at (1, 0) in Fig. 2 are determined by the rate of B_d mixing and by the ratio of rates of B_d and B_s mixing, respectively. The ratio is used because in it some uncertainties cancel, hence yielding a thinner ring. The bright green region labeled ε_K is determined by CP violation in K^0 - \bar{K}^0 mixing.
- (ii) CP asymmetries. Decay asymmetries, measuring the difference in rates of a process and the CP conjugate process, directly probe relative phases of KM elements, and in particular the unitarity triangle angles α , β and γ . We will also study these, with particular attention to the poster boy, the determination of $\sin(2\beta)$ from $B_d \rightarrow \psi K_S$, which is largely free from hadronic uncertainties. In Fig. 2 the blue and brown wedges labeled $\sin 2\beta$ and γ , respectively, and the peculiarly shaped light blue region labeled α are all obtained from various CP asymmetries in decays of B_d mesons.

1.5 FCNC

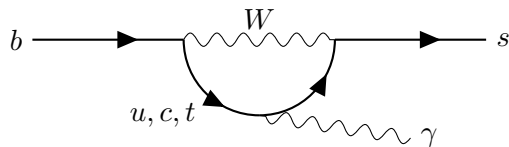
This stands for **F**lavor **C**hanging **N**eutral **C**urrents, but it is used more generally to mean Flavor Changing Neutral transitions, not necessarily “currents.” By this we mean an interaction that changes flavor but does not change electric charge. For example, a transition from a b -quark to an s - or d -quarks would be flavor changing neutral, but not so a transition from a b -quark to a c - or u -quark. Let’s review flavor changing transitions in the SM:

- 1. Tree level. Only interactions with the charged vector bosons W^\pm change flavor; *cf.* (11). The photon and Z couple diagonally in flavor space, so these “neutral currents” are flavor conserving.

For example, $n \rightarrow pe\bar{\nu}$ is



- 2. 1-loop. Can we have FCNCs at 1-loop? Say, $b \rightarrow s\gamma$? Answer: YES. Here is a diagram:



Hence, FCNC are suppressed in the SM by a 1-loop factor of $\sim \frac{g_2^2}{16\pi^2} \sim \frac{\alpha}{4\pi c_W^2}$ relative to the flavor changing charged currents.

Exercises

Exercise 1.5-1: Just in case you have never computed the μ -lifetime, verify that

$$\tau_\mu^{-1} \approx \Gamma(\mu \rightarrow e\nu_\mu\bar{\nu}_e) = \frac{G_F^2 m_\mu^5}{192\pi^3}$$

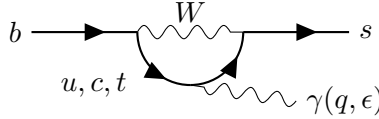
neglecting m_e , at lowest order in perturbation theory.

Exercise 1.5-2: Compute the amplitude for $Z \rightarrow b\bar{s}$ in the SM to lowest order in perturbation theory (in the strong and electroweak couplings). Don’t bother to compute integrals explicitly, just make sure they are finite (so you could evaluate them numerically if need be). Of course, if you can express the result in closed analytic form, you should. See Ref. [10].

1.6 GIM-mechanism: more suppression of FCNC

1.6.1 Old GIM

Let's imagine a world with a light top and a hierarchy $m_u < m_c < m_t \ll M_W$. Just in case you forgot, the real world is not like this, but rather it has $m_u \ll m_c \ll M_W \approx \frac{1}{2}m_t$. We can make a lot of progress towards the computation of the Feynman graph for $b \rightarrow s\gamma$ discussed previously without computing any integrals explicitly:



$$= eq_\mu \epsilon_\nu \bar{u}(p_s) \sigma^{\mu\nu} \left(\frac{1+\gamma_5}{2} \right) u(p_b) \frac{m_b}{M_W^2} \frac{g_2^2}{16\pi^2} \cdot I$$

where

$$I = \sum_{i=u,c,t} V_{ib} V_{is}^* F\left(\frac{m_i^2}{M_W^2}\right)$$

and $F(x)$ is some function that results from doing the integral explicitly, and we expect it to be of order 1. The coefficient of this unknown integral can be easily understood. First, it has the obvious loop factor ($g_2^2/16\pi^2$), photon coupling constant (e) and KM factors $V_{ib} V_{is}^*$ from the charged current interactions. Next, in order to produce a real (on-shell) photon the interaction has to be of the transition magnetic-moment form, $F_{\mu\nu} \bar{s} \sigma^{\mu\nu} b$, which translates into the Dirac spinors $u(p)$ for the quarks combining with the photon's momentum q and polarization vector (ϵ) through $q_\mu \epsilon_\nu \bar{u}(p_s) \sigma^{\mu\nu} u(p_b)$.⁴ Finally, notice that the external quarks interact with the rest of the diagram through a weak interaction, which involves only left-handed fields. This would suggest getting an amplitude proportional to $\bar{u}(p_s) \left(\frac{1+\gamma_5}{2} \right) \sigma^{\mu\nu} \left(\frac{1-\gamma_5}{2} \right) u(p_b)$ which, of course, vanishes. So we need one or the other of the external quarks to flip its chirality, and only then interact. A chirality flip produces a factor of the mass of the quark and we have chosen to flip the chirality of the b quark because $m_b \gg m_s$. This explains both the factor of m_b and the projector $\frac{1+\gamma_5}{2}$ acting on the spinor for the b -quark. The correct units (dimensional analysis) are made up by the factor of $1/M_W^2$.

Now, since we are pretending $m_u < m_c < m_t \ll M_W$, let's expand in a Taylor series, $F(x) = F(0) + xF'(0) + \dots$

$$I = \left(\sum_{i=u,c,t} V_{ib} V_{is}^* \right) F(0) + \left(\sum_{i=u,c,t} V_{ib} V_{is}^* \frac{m_i^2}{M_W^2} \right) F'(0) + \dots$$

Unitarity of the KM matrix gives $\sum_{i=u,c,t} V_{ib} V_{is}^* = 0$ so the first term vanishes. Moreover, we can rewrite the unitarity relation as giving one term as a combination of the other two, for example,

$$V_{tb} V_{ts}^* = - \sum_{i=u,c} V_{ib} V_{is}^*$$

giving us

$$I \approx -F'(0) \sum_{i=u,c} V_{ib} V_{is}^* \frac{m_t^2 - m_i^2}{M_W^2}$$

We have uncovered additional FCNC suppression factors. Roughly,

$$I \sim V_{ub} V_{us}^* \frac{m_t^2 - m_u^2}{M_W^2} + V_{cb} V_{cs}^* \frac{m_t^2 - m_c^2}{M_W^2} \sim \epsilon^4 \frac{m_t^2}{M_W^2} + \epsilon^2 \frac{m_t^2}{M_W^2}.$$

⁴The other possibility, that the photon field A_μ couples to a flavor changing current, $A_\mu \bar{b} \gamma^\mu s$, is forbidden by electromagnetic gauge invariance. Were you to expand the amplitude in powers of q/M_Z you could in principle obtain at lowest order the contribution, $\epsilon^\mu \bar{u}(p_s) \gamma^\mu u(p_b)$. But this should be invariant (gauge invariance) under $\epsilon^\mu \rightarrow \epsilon^\mu + q^\mu$, where $q = p_b - p_s$.

So in addition the 1-loop suppression, there is a mass suppression (m_t^2/M_W^2) and a mixing angle suppression (ϵ^2). This combination of suppression factors was uncovered by Glashow, Iliopoulos and Maiani (hence ‘‘GIM’’) [11] back in the days when we only knew about the existence of three flavors, u , d and s . They studied neutral kaon mixing, which involves a FCNC for s to d transitions and realized that theory would grossly over-estimate the mixing rate unless a fourth quark existed (the charm quark, c) that would produce the above type of cancellation (in the 2-generation case). Not only did they explain kaon mixing and predicted the existence of charm, they even gave a rough upper bound for the mass of the charm quark, which they could do since the contribution to the FCNC grows rapidly with the mass, as shown above. We will study kaon mixing in some detail later, and we will see that the top quark contribution to mixing is roughly as large as that of the charm quark: Glashow, Iliopoulos and Maiani were a bit lucky, the parameters of the SM-CKM could have easily favored top quark mediated dominance in kaon mixing and their bound could have been violated. As it turns out, the charm was discovered shortly after their work, and the mass turned out to be close to their upper bound.

1.6.2 Modern GIM

We have to revisit the above story, since $m_t \ll M_W$ is not a good approximation. Consider our example above, $b \rightarrow s\gamma$. The function $F(x)$ can not be safely Taylor expanded when the argument is the top quark mass. However, I is invariant under $F(x) \rightarrow F(x) + \text{constant}$, so we may choose without loss of generality $F(0) = 0$. Then

$$\begin{aligned} I &= -V_{cb}V_{cs}^* \left(F\left(\frac{m_t^2}{M_W^2}\right) - F'(0)\frac{m_c^2}{M_W^2} \right) - V_{ub}V_{us}^* \left(F\left(\frac{m_t^2}{M_W^2}\right) - F'(0)\frac{m_u^2}{M_W^2} \right) + \dots \\ &= F\left(\frac{m_t^2}{M_W^2}\right)V_{tb}V_{ts}^* + F'(0)\sum_{i=u,c} V_{ib}V_{is}^* \frac{m_i^2}{M_W^2} + \dots \\ &\sim \epsilon^2 F\left(\frac{m_t^2}{M_W^2}\right) \end{aligned}$$

We expect $F(x)$ to be order 1. This is indeed the case, $F(x)$ is a slowly increasing function of x that is of order 1 at the top quark mass. The contributions from u and c quarks to I are completely negligible, and virtual top-quark exchange dominates this amplitude.

Exercises

Exercise 1.6.2-1: Consider $s \rightarrow d\gamma$. Show that the above type of analysis suggests that virtual top quark exchange no longer dominates, but that in fact the charm and top contributions are roughly equally important.

Note: For this you need to know the mass of charm relative to M_W . If you don't, look it up!

1.7 Bounds on New Physics

Now let's bring together all we have learned. Let's stick to the process $b \rightarrow s\gamma$, which in fact places some of the most stringent constraints on models of new physics (NP). Let's model the contribution of NP by adding a dimension 6 operator to the Lagrangian,⁵

$$\Delta\mathcal{L} = \frac{C}{\Lambda^2} e F_{\mu\nu} H \bar{q}_L \sigma^{\mu\nu} b_R = \frac{evC}{\sqrt{2}\Lambda^2} F_{\mu\nu} \bar{s}_L \sigma^{\mu\nu} b_R + \dots$$

I have assumed the left handed doublet belongs in the second generation. The coefficient of the operator is C/Λ^2 : C is dimensionless and we assume it is of order 1, while Λ has dimensions of mass and

⁵The field strength should be the one for weak hypercharge, and the coupling constant should be g_1 . This is just a distraction and does not affect the result; in the interest of pedagogy I have been intentionally sloppy.

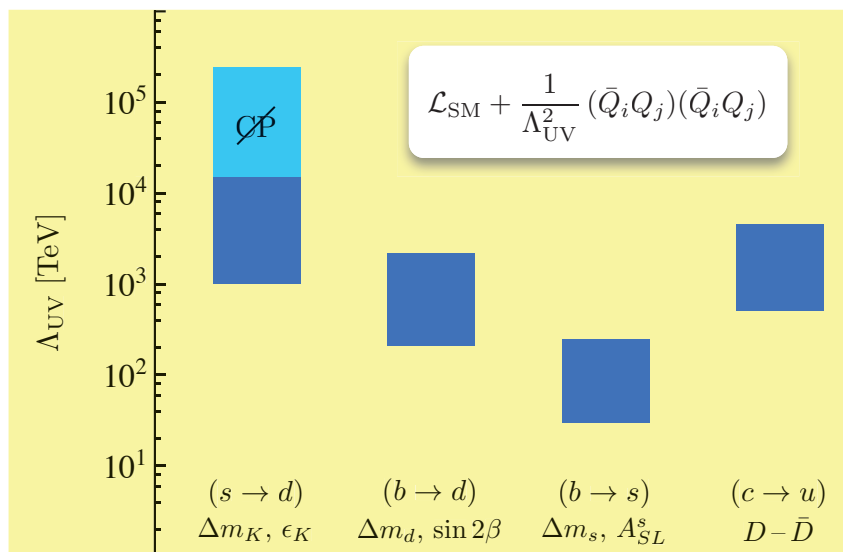


Fig. 5: Bounds on the NP scale from various processes. The NP is modeled as dimension 6 operators. No accidental suppression of the coefficient (as in MFV) is included. The $b \rightarrow s$ case is consistent with the explicit $b \rightarrow s\gamma$ example worked out in these notes. The figure is taken from M. Neubert’s talk at EPS 2011.

indicates the energy scale of the NP. It is easy to compute this term’s contribution to the amplitude. It is even easier to roughly compare it to that of the SM,

$$\frac{\mathcal{A}_{\text{NP}}}{\mathcal{A}_{\text{SM}}} \sim \frac{\frac{vC}{\sqrt{2}\Lambda^2}}{|V_{tb}V_{ts}^*| \frac{\alpha}{4\pi s_W^2} \frac{m_b}{M_W^2}}$$

Require this ratio be less than, say, 10%, since the SM prediction agrees at that level with the measurement. This gives,

$$C^{-1}\Lambda^2 \gtrsim \frac{vM_W^2 s_W^2}{\sqrt{2}m_b|V_{tb}V_{ts}^*| \frac{\alpha}{4\pi}} \cdot \frac{1}{0.1} \Rightarrow \Lambda \gtrsim 70 \text{ TeV}.$$

This bound is extraordinarily strong. The energy scale of 70 TeV is much higher than that of any existing or planned particle physics accelerator facility.

In the numerical bound above we have taken $C \sim 1$, but clearly a small coefficient would help bring the scale of NP closer to experimental reach. The question is what would make the coefficient smaller. One possibility is that the NP is weakly coupled and the process occurs also at 1-loop but with NP mediators in the loop. Then we can expect $C \sim \alpha/4\pi s_W^2$, which brings the bound on the scale of new physics down to about 4 TeV.

Figure 5 shows bounds on the scale of NP from various processes. The NP is modeled as dimension 6 operators, just as in our discussion above. The coefficients of the operators C/Λ^2 are assumed to have $C \approx 1$. The $b \rightarrow s$ case is consistent with our discussion above.

1.7.1 Minimal Flavor Violation

Suppose we extend the SM by adding terms (local,⁶ Lorentz invariant and gauge invariant) to the Lagrangian. Since the SM already includes all possible monomials (“operators”) of dimension 4 or smaller, we consider adding operators of $\text{dim} \geq 5$. We are going to impose an additional constraint, and we will

⁶By “local” we mean a product of fields all evaluated at the same spacetime point.

investigate its consequence. We will require that these operators be invariant under the flavor transformations, comprising the group G_F . We will include the Yukawa matrices as spurions:

$$q_L \rightarrow U_q q_L, \quad u_R \rightarrow U_u u_R, \quad d_R \rightarrow U_d d_R, \quad \lambda_U \rightarrow U_q \lambda_U U_u^\dagger, \quad \lambda_D \rightarrow U_q \lambda_D U_d^\dagger. \quad (17)$$

We add some terms to the Lagrangian

$$\mathcal{L} \rightarrow \mathcal{L} + \Delta\mathcal{L}, \quad \Delta\mathcal{L} = \sum_i c_i O_i$$

with O_i operators of $\dim \geq 5$ invariant under (17). For example,

$$O_1 = G_{\mu\nu}^a H \bar{u}_R T^a \sigma^{\mu\nu} \lambda_U q_L, \\ O_2 = \bar{q}_L \gamma^\mu \lambda_U^\dagger \lambda_U q_L \bar{d}_R \gamma_\mu \lambda_D \lambda_D^\dagger d_R,$$

where $G_{\mu\nu}^a$ is the field strength for the $SU(3)_c$ gauge field (which is quite irrelevant for our discussion, so don't be distracted). Consider these operators when we rotate to the basis in which the mass matrices are diagonal. Start with the first:

$$O_1 \rightarrow G_{\mu\nu}^a H \bar{u}_R T^a \sigma^{\mu\nu} V_{u_R}^\dagger \lambda_U \begin{pmatrix} V_{u_L} u_L \\ V_{d_L} d_L \end{pmatrix} \\ = G_{\mu\nu}^a H \bar{u}_R T^a \sigma^{\mu\nu} (V_{u_R}^\dagger \lambda_U V_{u_L}) \begin{pmatrix} u_L \\ V_{u_L}^\dagger V_{d_L} d_L \end{pmatrix} \\ = G_{\mu\nu}^a H \bar{u}_R T^a \sigma^{\mu\nu} \lambda'_U \begin{pmatrix} u_L \\ V d_L \end{pmatrix}$$

We see that the only flavor-changing interaction is governed by the off-diagonal components of $\lambda'_U V$. Similarly

$$O_2 \rightarrow \bar{q}'_L \gamma^\mu (\lambda'_U)^\dagger q'_L \bar{d}_R \gamma_\mu (\lambda'_D)^\dagger d_R, \quad \text{where } q'_L = \begin{pmatrix} u_L \\ V d_L \end{pmatrix}.$$

This construction, restricting the higher dimension operators by the flavor symmetry with the Yukawa couplings treated as spurions, goes by the name of the *principle of Minimal Flavor Violation* (MFV). Extensions of the SM in which the only breaking of G_F is by λ_U and λ_D automatically satisfy MFV. As we will see they are much less constrained by flavor changing and CP-violating observables than models with generic breaking of G_F .

Exercises

Exercise 1.7.1-1: Had we considered an operator like O_1 but with $\tilde{H} \bar{d}_R$ instead of $H \bar{u}_R$ the flavor off-diagonal terms would have been governed by $\lambda'_D V^\dagger$. Show this is generally true, that is, that flavor change in any operator is governed by V and powers of λ' .

Exercise 1.7.1-2: Exhibit examples of operators of dimension 6 that produce flavor change without involving $\lambda_{U,D}$. Can these be such that only quarks of charge $+2/3$ are involved? (These would correspond to Flavor Changing Neutral Currents; see Sec. 1.5 below).

Now let's consider the effect of the principle of MFV on the process $b \rightarrow s\gamma$. Our first attempt is

$$\Delta\mathcal{L} = \frac{C}{\Lambda^2} e F_{\mu\nu} H \bar{q}_L \lambda_D \sigma^{\mu\nu} d_R.$$

This gives no flavor changing interaction when we go to the field basis that diagonalizes the mass matrices (which can be seen from the analysis above, or simply by noting that this term has the same form, as far

as flavor is concerned, as the mass term in the Lagrangian). To get around this we need to construct an operator which either contains more fields, which will give a loop suppression in the amplitude plus an additional suppression by powers of Λ , or additional factors of spurions. We try the latter. Consider, then

$$\Delta\mathcal{L} = \frac{C}{\Lambda^2} e F_{\mu\nu} H \bar{q}_L \lambda_U \lambda_U^\dagger \lambda_D \sigma^{\mu\nu} d_R.$$

When you rotate the fields to diagonalize the mass matrix you get, for the charge neutral quark bi-linear,

$$\lambda_U \lambda_U^\dagger \lambda_D \rightarrow V_{d_L}^\dagger \lambda_U \lambda_U^\dagger \lambda_D V_{d_R} = V_{d_L}^\dagger V_{u_L} (\lambda'_U)^2 V_{u_L}^\dagger V_{d_L} \lambda'_D = V^\dagger (\lambda'_U)^2 V \lambda'_D, \quad (18)$$

our estimate of the NP amplitude is suppressed much like in the SM, by the mixing angles and the square of the “small” quark masses. Our bound now reads

$$C^{-1} \Lambda^2 \gtrsim \frac{M_W^2 s_W^2}{\sqrt{2} \frac{\alpha}{4\pi}} \cdot \frac{1}{0.1} \Rightarrow C^{-1/2} \Lambda \gtrsim 4 \text{ TeV}$$

This is within the reach of the LHC (barely), even if $C \sim 1$ which should correspond to a strongly coupled NP sector. If for a weakly coupled sector C is one loop suppressed, Λ could be interpreted as a mass M_{NP} of the NP particles in the loop, and the analysis gives $M_{\text{NP}} \gtrsim 200 \text{ GeV}$. The moral is that if you want to build a NP model to explain putative new phenomena at the Tevatron or the LHC you can get around constraints from flavor physics if your model incorporates the principle of MFV (or some other mechanism that suppresses FCNC).

Exercises

Exercise 1.7.1-3: Determine how much each of the bounds in Fig. 5 is weakened if you assume MFV. You may not be able to complete this problem if you do not have some idea of what the symbols ΔM_K , ϵ_K , etc, mean or what type of operators contribute to each process; in that case you should postpone this exercise until that material has been covered later in these lectures.

1.7.2 Examples

This section may be safely skipped: it is not used elsewhere in these notes. The examples presented here require some background knowledge. Skip the first one if you have not studied supersymmetry yet.

1. *The supersymmetrized SM.* I am not calling this the MSSM, because the discussion applies as well to the zoo of models in which the BEH sector has been extended, *e.g.*, the NMSSM. In the absence of SUSY breaking this model satisfies the principle of MFV. The Lagrangian is

$$\mathcal{L} = \int d^4\theta [\bar{Q} e^V Q + \bar{U} e^V U + \bar{D} e^V D] + \text{gauge \& } H \text{ kinetic terms} + \int d^2\theta W + \text{h.c.}$$

with superpotential

$$W = H_1 U y_U Q + H_2 D y_D Q + \text{non-quark-terms}$$

Here V stands for the vector superfields⁷ and Q, D, U, H_1 and H_2 are chiral superfields with the following quantum numbers:

$$\begin{aligned} Q &\sim (3, 2)_{1/6} & H_1 &\sim (1, 2)_{1/2} \\ U &\sim (\bar{3}, 1)_{-2/3} & H_2 &\sim (1, 2)_{-1/2} \\ D &\sim (\bar{3}, 1)_{1/3} \end{aligned}$$

⁷Since I will not make explicit use of vector superfields, there should be no confusion with the corresponding symbol for the the KM matrix, which is used ubiquitously in these lectures.

The fields on the left column come in three copies, the three generations we call flavor. We are again suppressing that index (as well as the gauge and Lorentz indices). Unlike the SM case, this Lagrangian is not the most general one for these fields once renormalizability, Lorentz and gauge invariance are imposed. In addition one needs to impose, of course, supersymmetry. But even that is not enough. One has to impose an R -symmetry to forbid dangerous baryon number violating renormalizable interactions.

When the Yukawa couplings are neglected, $y_U = y_D = 0$, this theory has a $SU(3)^3$ flavor symmetry. The symmetry is broken only by the couplings and we can keep track of this again by treating the couplings as spurions. Specifically, under $SU(3)^3$,

$$Q \rightarrow U_q Q, \quad U \rightarrow S_U U, \quad D \rightarrow S_D D, \quad y_U \rightarrow S_U^* y_U U_q^\dagger, \quad y_D \rightarrow S_D^* y_D U_q^\dagger$$

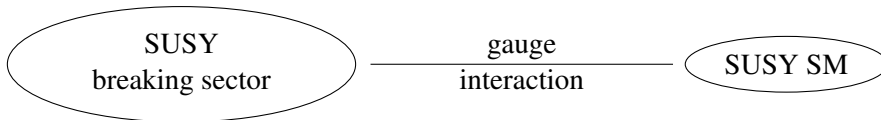
Note that this has both quarks and squarks transforming together. The transformations on quarks may look a little different than the transformation in the SM, Eq. (17). But they are the same, really. The superficial difference is that here the quark fields are all written as left-handed fields, which are obtained by charge-conjugation from the right handed ones in the standard representation of the SM. So in fact, the couplings are related by $y_U = \lambda_U^\dagger$ and $y_D = \lambda_D^\dagger$, and the transformations on the right handed fields by $S_U = U_u^*$ and $S_D = U_d^*$. While the relations are easily established, it is worth emphasizing that we could have carried out the analysis in the new basis without need to connect to the SM basis. All that matters is the way in which symmetry considerations restrict certain interactions.

Now let's add soft SUSY breaking terms. By ‘‘soft’’ we mean operators of dimension less than 4. Since we are focusing on flavor, we only keep terms that include fields that carry flavor:

$$\Delta\mathcal{L}_{\text{SUSY-bkg}} = \phi_q^* \mathcal{M}_q^2 \phi_q + \phi_u^* \mathcal{M}_u^2 \phi_u + \phi_d^* \mathcal{M}_d^2 \phi_d + (\phi_{h_1} \phi_u g_U \phi_q + \phi_{h_2} \phi_d g_D \phi_q + \text{h.c.}) \quad (19)$$

Here ϕ_X is the scalar SUSY-partner of the quark X . This breaks the flavor symmetry unless $\mathcal{M}_{q,u,d}^2 \propto \mathbf{1}$ and $g_{U,D} \propto y_{U,D}$ (see, however, Exercise 1.7.2-4). And unless these conditions are satisfied new flavor changing interactions are generically present and large. The qualifier ‘‘generically’’ is because the effects can be made small by lucky coincidences (fine tunings) or if the masses of scalars are large.

This is the motivation for gauge mediated SUSY-breaking [12]:



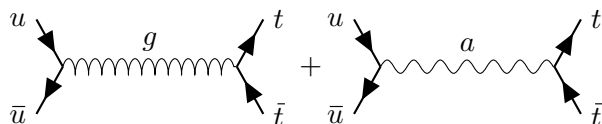
The gauge interactions, *e.g.*, $\bar{Q}e^V Q$, are diagonal in flavor space. In theories of supergravity mediated supersymmetry breaking the flavor problem is severe. To repeat, this is why gauge mediation and its variants were invented.

2. *MFV Fields*. Recently CDF and D0 reported a larger than expected forward-backward asymmetry in $t\bar{t}$ pairs produced in $p\bar{p}$ collisions [13]. Roughly speaking, define the forward direction as the direction in which the protons move, and classify the outgoing particles of a collision according to whether they move in the forward or backward direction. You can be more careful and define this relative to the CM of the colliding partons, or better yet in terms of rapidity, which is invariant under boosts along the beam direction. But we need not worry about such subtleties: for our purposes we want to understand how flavor physics plays a role in this process that one would

have guessed is dominated by SM interactions [14]. Now, we take this as an educational example, but I should warn you that by the time you read this the reported effect may have evaporated. In fact, since the lectures were given D0 has revised its result and the deviation from the SM expected asymmetry is now much smaller [15].

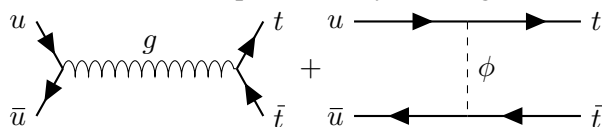
There are two types of BSM models that explain this asymmetry, classified according to the type of new particle exchange that produces the asymmetry:

- (i) s -channel. For example an “axi-gluon,” much like a gluon but massive and coupling to axial currents of quarks. The interference between vector and axial currents,



produces a FB-asymmetry. It turns out that it is best to have the sign of the axigluon coupling to t -quarks be opposite that of the coupling to u quarks, in order to get the correct sign of the FB-asymmetry without violating constraints from direct detection at the LHC. But different couplings to u and t means flavor symmetry violation and by now you should suspect that any complete model will be subjected to severe constraints from flavor physics.

- (ii) t -channel: for example, one may exchange a scalar, and the amplitude now looks like this:



This model has introduced a scalar ϕ with a coupling $\phi\bar{t}u$ (plus its hermitian conjugate). This clearly violates flavor symmetry. Not only we expect that the effects of this flavor violating coupling would be directly observable but, since the coupling is introduced in the mass eigenbasis, we suspect there are also other couplings involving the charge- $+2/3$ quarks, as in $\phi\bar{c}u$ and $\phi\bar{t}u$ and flavor diagonal ones. This is because even if we started with only one coupling in some generic basis of fields, when we rotate the fields to go the mass eigenstate basis we will generate all the other couplings. Of course this does not have to happen, but it will, generically, unless there is some underlying reason, like a symmetry. Moreover, since couplings to a scalar involve both right and left handed quarks, and the left handed quarks are in doublets of the electroweak group, we may also have flavor changing interactions involving the charge- $(-1/3)$ quarks in these models.

One way around these difficulties is to build the model so that it satisfies the principle of MFV, by design. Instead of having only a single scalar field, as above, one may include a multiplet of scalars transforming in some representation of G_F . So, for example, one can have a charged scalar multiplet ϕ transforming in the $(\mathbf{3}, \bar{\mathbf{3}}, 1)$ representation of $SU(3)_q \times SU(3)_u \times SU(3)_d$, with gauge quantum numbers $(1, 2)_{-1/2}$ and with interaction term

$$\lambda \bar{q}_L \phi u_R \quad \text{with} \quad \phi \rightarrow U_{qL} \phi U_{uR}^\dagger.$$

Note that the coupling λ is a single number (if we want invariance under flavor). This actually works! See [16].

Exercises

Exercise 1.7.2-4: Below Eq. (19) we said, “This breaks the flavor symmetry unless $\mathcal{M}_{q,u,d}^2 \propto \mathbf{1}$ and $g_{U,D} \propto y_{U,D}$.” This is not strictly correct (or, more bluntly, it is a lie). While not correct it is the simplest choice. Why? Exhibit alternatives, that is, other forms for $\mathcal{M}_{q,u,d}^2$ and $g_{U,D}$ that respect the symmetry. *Hint:* See (18).

Exercise 1.7.2-5: Classify all possible dim-4 interactions of Yukawa form in the SM. To this end list all possible Lorentz scalar combinations you can form out of pairs of SM quark fields. Then give explicitly the transformation properties of the scalar field, under the gauge and flavor symmetry groups, required to make the Yukawa interaction invariant. Do this first without including the SM Yukawa couplings as spurions and then including also one power of the SM Yukawa couplings.

2 Neutral Meson Mixing and CP Asymmetries

2.1 Why Study This?

Yeah, why? In particular why bother with an old subject like neutral- K meson mixing? I offer you an incomplete list of perfectly good reasons:

- (i) CP violation was discovered in neutral- K meson mixing.
- (ii) Best constraints on NP from flavor physics are from meson mixing. Look at Fig. 5, where the best constraint is from CP violation in neutral- K mixing. In fact, other than A_{SL}^s , all of the other observables in the figure involve mixing.
- (iii) It's a really neat phenomenon (and that should be sufficient reason for wanting to learn about it, I hope you will agree).
- (iv) It's an active field of research both in theory and in experiment. I may be just stating the obvious, but the LHCb collaboration has been very active and extremely successful, and even CMS and ATLAS have performed flavor physics analysis. And, of course, there are also several non-LHC experiments ongoing or planned; see, *e.g.*, [17].

But there is another reason you should pay attention to this, and more generally to the “phenomenology” (as opposed to “theory” or “model building”) part of these lectures. Instead of playing with Lagrangians and symmetries we will use these to try to understand dynamics, that is, the actual physical phenomena the Lagrangian and symmetries describe. As an experimentalist, or even as a model builder, you can get by without an understanding of this. Sort of. There are enough resources today where you can plug in the data from your model and obtain a prediction that can be tested against experiment. Some of the time. And all of the time without understanding what you are doing. You may get it wrong, you may miss effects. As a rule of thumb, if you are doing something good and interesting, it is novel enough that you may not want to rely on calculations you don't understand and therefore don't know if applicable. Besides, the more you know the better equipped you are to produce interesting physics.

2.2 What is mixing?

Suppose you have a \bar{B}_s meson with flavor quantum numbers $\bar{s}b$. If $b \rightarrow c\bar{u}d$, so that $\bar{s}b \rightarrow \bar{s}[c\bar{u}d] = (\bar{s}c)(\bar{u}d)$ you can have a decay $\bar{B}_s \rightarrow D_s^+\pi^-$. Now, the decay is not immediate: the \bar{B}_s meson has a non-zero lifetime. So if you somehow determined that you produced a \bar{B}_s at $t = 0$ and measure the probability of decaying into $D_s^+\pi^-$ as a function of time you get the oscillating function with an exponential envelope depicted by the red line in Fig. 6. Moreover, if you measure its decay probability into $D_s^-\pi^+$ you obtain the blue line in that same figure. The sum of the two curves is the exponentially decaying black curve. The final state $D_s^-\pi^+$ is what you expect from a decay of a B_s meson, rather than a \bar{B}_s .

We guess that as \bar{B}_s evolves we have transmutations of flavor, $\bar{B}_s \rightarrow B_s \rightarrow \bar{B}_s \rightarrow B_s \rightarrow \dots$. We can model this by assuming the time evolution of the state is

$$|\bar{B}_s(t)\rangle = e^{-\frac{1}{2}\Gamma t} [\cos(\omega t)|\bar{B}_s\rangle + \sin(\omega t)|B_s\rangle]$$

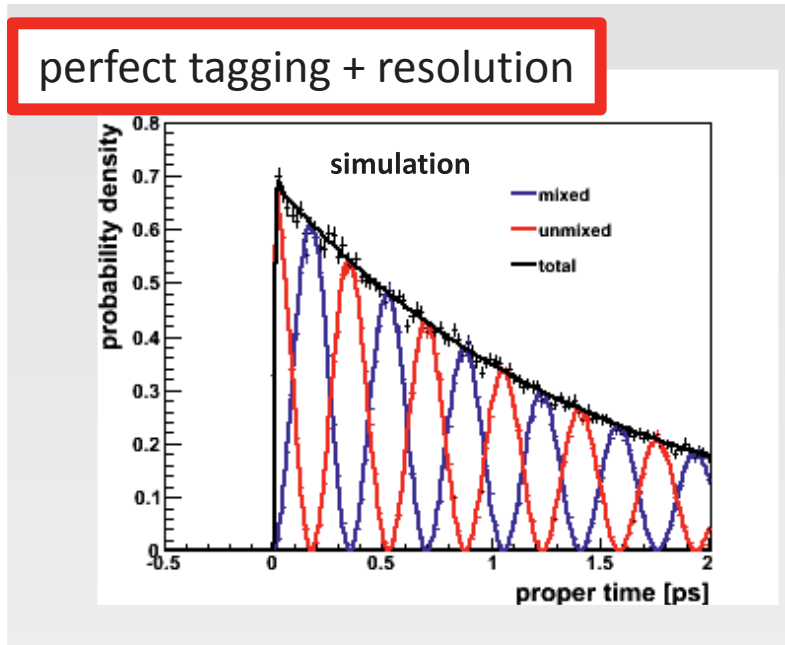
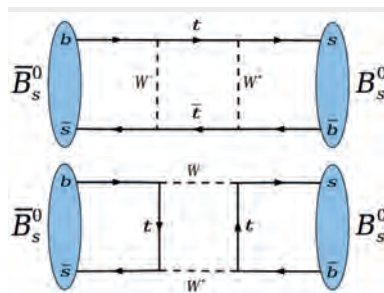


Fig. 6: Decay probability of a \bar{B}_s meson as a function of proper time in a perfect world (perfect tagging and resolution) from Ref. [18]. The red and blue lines correspond to $D_s^+ \pi^-$ and $D_s^- \pi^+$ final states, respectively, and the black is the sum. “Unmixed” refers to the fact that the tagging determined that initially the state is \bar{B}_s .

where the \bar{B}_s and B_s states of the right hand side are defined as having the quantum numbers $\bar{s}b$ and $\bar{b}s$, respectively. How can a \bar{B}_s turn into a B_s ? Weak interactions can do that: Feynman graphs producing the transition are shown here:



This must be a very small effect. It is a weak interaction. And it is further suppressed by being a 1-loop effect and by CKM mixing angles (modern GIM).

Let’s ignore the fact that there is a finite life-time for the moment and concentrate on the mixing aspect of these states. In quantum mechanics the state of a free \bar{B}_s at rest evolves according to Schrödinger’s equation,

$$i \frac{d}{dt} \bar{B}_s(t) = M \bar{B}_s(t)$$

where I have used the mass, M , of the state as its energy at rest, and similarly for the B_s state which, incidentally, has the same mass. The small perturbation introduced by the Feynman diagrams above couples the evolution of the two states. We can model this by coupling the two Schrödinger equations as follows:

$$i \frac{d}{dt} \begin{pmatrix} \bar{B}_s(t) \\ B_s(t) \end{pmatrix} = M \begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix} \begin{pmatrix} \bar{B}_s(t) \\ B_s(t) \end{pmatrix}$$

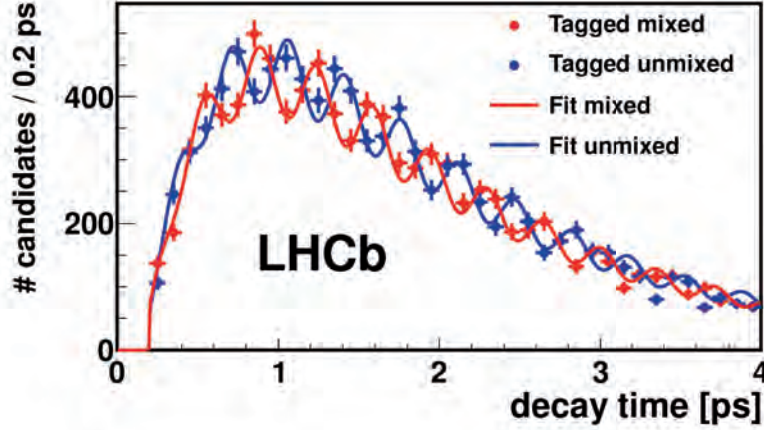


Fig. 7: As in Fig. 6 but with finite resolution and imperfect tagging [18]. This time, however, the figure shows data measured at LHCb rather than a computer simulation.

The matrix $\begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix}$ has eigenvalues $1 \pm \epsilon$, but no matter how small ϵ is the eigenvectors $\begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}$ are maximally mixed! The solution to the differential equation is straightforward,

$$\bar{B}_s(t) = e^{-iMt} [\cos(\epsilon Mt)\bar{B}_s(0) - i \sin(\epsilon Mt)B_s(0)].$$

This is the magic of meson-mixing: a very small perturbation gives a large effect (full mixing). The smallness of ϵ shows up in the frequency of oscillation, but the oscillation turns the initial \bar{B}_s into 100% B_s in half a period of oscillation.

Before we go on to a more complete treatment of this phenomenon let's take a look at real data and understand how one can determine that the initial state is in fact a \bar{B}_s , as opposed to a B_s . Fig. 7 shows LHCb data that corresponds to the ideal case of Fig. 6. The difference between the two figures is well understood as arising from imperfect resolution and tagging. Tagging is the method by which the experiment determines the initial state is in fact a \bar{B}_s . Figure 8 is a diagrammatic representation of a B_s meson (with a \bar{b} -quark) produced on the "same side." At the primary vertex one may observe a K^+ signaling the presence of the \bar{s} quark and hence a tag that the B -meson produced contains an s -quark. The opposite side must contain a state with a b quark. If it decays semileptonically, $b \rightarrow c\ell^-\nu$ it will produce a negatively charged lepton; e^- or μ^- also tag the B_s . When the opposite side b quark decays it is highly likely that it will produce a c -quark, and this one, in turn, an s quark, so a K^- signals the presence of a b quark on the opposite side, giving a third tag.

2.3 Mixing: Formalism

We present the Weisskopf-Wigner mixing formalism for a generic neutral meson-antimeson system, denoted by $X^0 - \bar{X}^0$. We can apply this to the cases $X^0 = K^0, D^0, B^0$ and B_s . Under charge conjugation (C) and spatial inversions (or parity, P) states with a single pseudoscalar meson at rest transform as

$$\begin{aligned} P|X^0\rangle &= -|X^0\rangle & P|\bar{X}^0\rangle &= -|\bar{X}^0\rangle \\ C|X^0\rangle &= |\bar{X}^0\rangle & C|\bar{X}^0\rangle &= |X^0\rangle \end{aligned}$$

Of course, there is an implicit transformation of the momentum of the state under P . We will be interested in CP-violation. The combination of the above transformations gives

$$CP|\bar{X}^0\rangle = -|X^0\rangle \quad \text{and} \quad CP|X^0\rangle = -|\bar{X}^0\rangle.$$

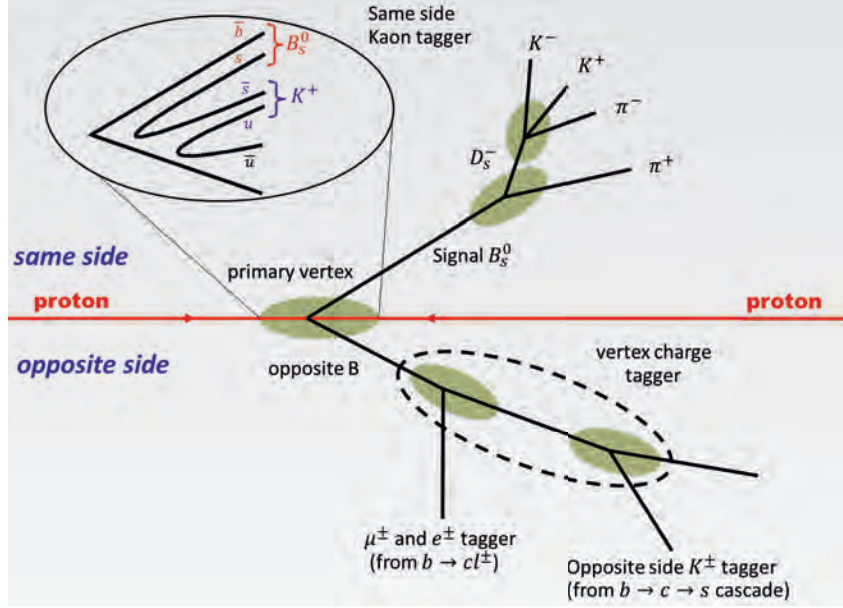


Fig. 8: Tagging from lepton charge or opposite side K charge for $B_s \rightarrow D_s^- \pi^+$ decays. Figure from Ref. [18].

As in our guess in the previous section we study this system allowing for mixing between the two states in their rest frame. But now we want to incorporate finite life-time effects. So for the time evolution we need a Hamiltonian that contains a term that corresponds to the width. In other words, since these one particle states may evolve into states that are not accounted for in the two state Hamiltonian, the evolution will not be unitary and the Hamiltonian will not be Hermitian. Keeping this in mind we write, for this effective Hamiltonian

$$\mathbf{H} = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} = \begin{pmatrix} M - \frac{i}{2}\Gamma & M_{12} - \frac{i}{2}\Gamma_{12} \\ M_{12}^* - \frac{i}{2}\Gamma_{12}^* & M - \frac{i}{2}\Gamma \end{pmatrix} \quad (20)$$

where $\mathbf{M}^\dagger = \mathbf{M}$ and $\mathbf{\Gamma}^\dagger = \mathbf{\Gamma}$. Also we have taken $|1\rangle = |X^0\rangle$ and $|2\rangle = |\bar{X}^0\rangle$. We have insisted on CPT: $(CPT)^{-1} \mathbf{H} (CPT) = \mathbf{H}^\dagger \Rightarrow H_{11} = H_{22}$. Studies of CPT invariance relax this assumption; see Ref. [19].

Exercises

Exercise 2.3-1: Show that CPT implies $H_{11} = H_{22}$.

CP invariance requires $M_{12}^* = M_{12}$ and $\Gamma_{12}^* = \Gamma_{12}$. Therefore *either* $\text{Im}M_{12} \neq 0$ *or* $\text{Im}\Gamma_{12} \neq 0$, or both, signal that CP is violated. Now, to study the time evolution of the system we solve Schrödinger's equation. To this end we first solve the eigensystem for the effective Hamiltonian. The physical eigenstates are labeled conventionally as Heavy and Light

$$|X_H\rangle = p|X^0\rangle + q|\bar{X}^0\rangle, \quad |X_L\rangle = p|X^0\rangle - q|\bar{X}^0\rangle \quad (21)$$

and the corresponding eigenvalues are defined as

$$M_{X_H} - \frac{i}{2}\Gamma_{X_H} = M - \frac{i}{2}\Gamma \pm \frac{1}{2}(\Delta M - \frac{i}{2}\Delta\Gamma).$$

Note that for $q = p$ these are CP -eigenstates: $CP|X_H\rangle = \mp|X_H\rangle$.

We still have to give the eigenvalues and coefficients p, q in terms of the entries in the Hamiltonian. From the eigenstate equation we read off,

$$\frac{p}{q} = 2 \frac{M_{12} - \frac{i}{2}\Gamma_{12}}{\Delta M - \frac{i}{2}\Delta\Gamma} = \frac{1}{2} \frac{\Delta M - \frac{i}{2}\Delta\Gamma}{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}$$

From this we can write simple non-linear equations giving ΔM and $\Delta\Gamma$:

$$\begin{aligned} (\Delta M)^2 - \frac{1}{4}(\Delta\Gamma)^2 &= 4|M_{12}|^2 - |\Gamma_{12}|^2 \\ \Delta M\Delta\Gamma &= 4\text{Re}(M_{12}\Gamma_{12}^*) \end{aligned} \quad (22)$$

For Kaons it is standard practice to label the states differently, with Long and Short instead of Heavy and Light: the eigenvalues of the 2×2 Hamiltonian are

$$M_{K_S} - \frac{i}{2}\Gamma_{K_S} = M - \frac{i}{2}\Gamma \pm \frac{1}{2}(\Delta M - \frac{i}{2}\Delta\Gamma)$$

and the corresponding eigenvectors are

$$|K_S\rangle = \frac{1}{\sqrt{2(1+|\epsilon|^2)}} \left[(1+\epsilon)|K^0\rangle \pm (1-\epsilon)|\bar{K}^0\rangle \right] \quad (23)$$

If $\epsilon = 0$ these are CP -eigenstates: $CP|K_L\rangle = -|K_L\rangle$ and $CP|K_S\rangle = |K_S\rangle$. Since $CP|\pi\pi\rangle_{\ell=0} = |\pi\pi\rangle_{\ell=0}$ and $CP|\pi\pi\pi\rangle_{\ell=0} = -|\pi\pi\pi\rangle_{\ell=0}$ we see that if CP were a good symmetry the decays $K_L \rightarrow \pi\pi\pi$ and $K_S \rightarrow \pi\pi$ are allowed, but not so the decays $K_L \rightarrow \pi\pi$ and $K_S \rightarrow \pi\pi\pi$. Barring CP violation in the decay amplitude, observation of $K_L \rightarrow \pi\pi$ or $K_S \rightarrow \pi\pi\pi$ indicates $\epsilon \neq 0$, that is, CP -violation in mixing.

This is very close to what is observed:

$$\begin{aligned} \text{Br}(K_S \rightarrow \pi\pi) &= 100.00 \pm 0.24\% \\ \text{Br}(K_L \rightarrow \pi\pi) &= 0.297 \pm 0.023\% \\ \text{Br}(K_L \rightarrow \pi\pi\pi) &= 33.9 \pm 1.2\% \end{aligned} \quad (24)$$

Hence, we conclude (i) ϵ is small, and (ii) CP is not a symmetry. The longer life-time of K_L is accidental. To understand this notice that $3m_\pi \sim 3(140) \text{ MeV} = 420 \text{ MeV}$ while $m_K \sim 490 \text{ MeV}$, leaving little phase space for the decays $K \rightarrow \pi\pi\pi$. This explains why K_L is much longer lived than K_S ; the labels “L” and “S” stand for “long” and “short,” respectively:

$$\begin{aligned} \tau_{K_S} &= 0.59 \times 10^{-10} \text{ s} \\ \tau_{K_L} &= 5.18 \times 10^{-8} \text{ s} \end{aligned}$$

This is no longer the case for heavy mesons for which there is a multitude of possible decay modes and only a few multi-particle decay modes are phase-space suppressed.

Eventually we will want to connect this effective 2×2 Hamiltonian to the underlying fundamental physics we are studying. This can be done using perturbation theory (in the weak interactions) and is an elementary exercise in Quantum Mechanics (see, *e.g.*, Messiah’s textbook, p.994 – 1001 [20]). With $|X^0\rangle = |1\rangle$ and $|\bar{X}^0\rangle = |2\rangle$ one has

$$M_{ij} = M\delta_{ij} + \langle i|H|j\rangle + \sum'_n \text{PP} \frac{\langle i|H|n\rangle\langle n|H|j\rangle}{M - E_n} + \dots \quad (25)$$

$$\Gamma_{ij} = 2\pi \sum'_n \delta(M - E_n) \langle i|H|n\rangle\langle n|H|j\rangle + \dots \quad (26)$$

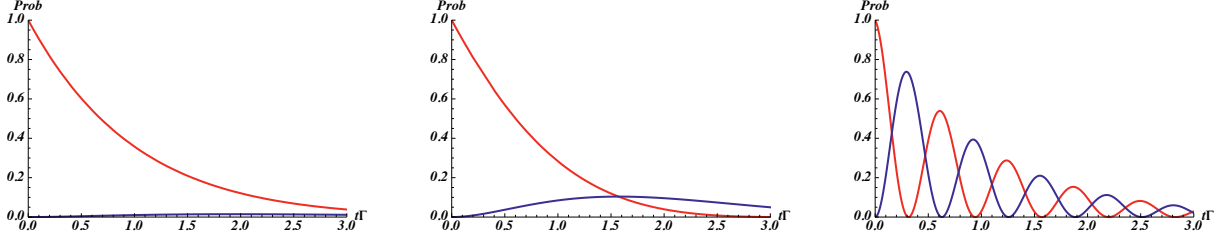


Fig. 9: Mixing probability in $X^0 - \bar{X}^0$ mixing as a function of Γt for $\Delta M/\Gamma = 1/3, 1$ and 3 in left, center and right panels, respectively, assuming $\Delta\Gamma = 0$ and $|p/q| = 1$. In red is the probability for the unmixed state and in blue for the mixed state.

Here the prime in the summation sign means that the states $|1\rangle$ and $|2\rangle$ are excluded and PP stands for “principal part.” Beware the states are assumed discrete and normalized to unity. Also, H is a Hamiltonian, not a Hamiltonian density \mathcal{H} ; $H = \int d^3x \mathcal{H}$. It is the part of the SM Hamiltonian that can produce flavor changes. In the absence of H the states $|X^0\rangle = |1\rangle$ and $|\bar{X}^0\rangle = |2\rangle$ would be stable eigenstates of the Hamiltonian and their time evolution would be by a trivial phase. It is assumed that this flavor-changing interaction is weak, while there may be other much stronger interactions (like the strong one that binds the quarks together). The perturbative expansion is in powers of the weak interaction while the matrix elements are computed non-perturbatively with respect to the remaining (strong) interactions. Of course the weak flavor changing interaction is, well, the Weak interaction of the electroweak model, and below we denote the Hamiltonian by H_w .

2.4 Time Evolution in $X^0 - \bar{X}^0$ mixing.

We have looked at processes involving the ‘physical’ states K_L and K_S . As these are eigenvectors of H their time evolution is quite simple

$$i \frac{d}{dt} |X_{H,L}\rangle = (M_{H,L} - \frac{i}{2}\Gamma_{H,L}) |X_{H,L}\rangle \quad \Rightarrow \quad |X_{H,L}(t)\rangle = e^{-iM_{H,L}t} e^{-\frac{1}{2}\Gamma_{H,L}t} |X_{H,L}(0)\rangle$$

Since $|X_{H,L}\rangle$ are eigenvectors of \mathbf{H} , they do not mix as they evolve. But often one creates X^0 or \bar{X}^0 in the lab. These, of course, mix with each other since they are linear combinations of X_H and X_L .

The time evolution of $X_{H,L}$ is trivially given by

$$|X_{H,L}(t)\rangle = e^{-iM_{H,L}t} e^{-\frac{1}{2}\Gamma_{H,L}t} |X_{H,L}(0)\rangle.$$

Now we can invert,

$$\begin{aligned} |X^0\rangle &= \frac{1}{2p} (|X_H\rangle + |X_L\rangle), \\ |\bar{X}^0\rangle &= \frac{1}{2q} (|X_H\rangle - |X_L\rangle). \end{aligned} \quad (27)$$

Hence,

$$|X^0(t)\rangle = \frac{1}{2p} \left[e^{-iM_H t} e^{-\frac{1}{2}\Gamma_H t} |X_H(0)\rangle + e^{-iM_L t} e^{-\frac{1}{2}\Gamma_L t} |X_L(0)\rangle \right]$$

and using (21) for the states at $t = 0$ we obtain

$$|X^0(t)\rangle = f_+(t) |X^0\rangle + \frac{q}{p} f_-(t) |\bar{X}^0\rangle \quad (28)$$

where

$$\begin{aligned} f_{\pm}(t) &= \frac{1}{2} \left[e^{-iM_H t} e^{-\frac{1}{2}\Gamma_H t} \pm e^{-iM_L t} e^{-\frac{1}{2}\Gamma_L t} \right] \\ &= \frac{1}{2} e^{-iM_H t} e^{-\frac{1}{2}\Gamma_H t} \left[1 \pm e^{i\Delta M t} e^{\frac{1}{2}\Delta\Gamma t} \right] \\ &= \frac{1}{2} e^{-iM_L t} e^{-\frac{1}{2}\Gamma_L t} \left[e^{-i\Delta M t} e^{-\frac{1}{2}\Delta\Gamma t} \pm 1 \right] \end{aligned} \quad (29)$$

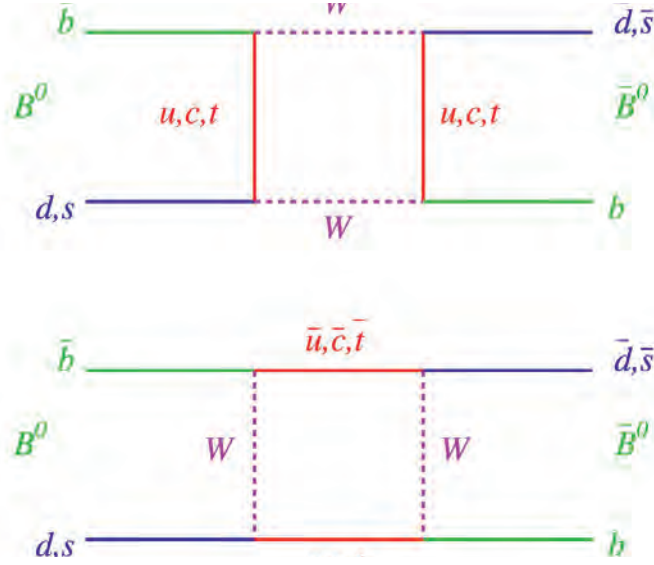


Fig. 10: Box diagrams contributing to $B_{d,s}$ -mixing.

Similarly,

$$|\bar{X}^0(t)\rangle = \frac{p}{q} f_-(t) |X^0\rangle + f_+(t) |\bar{X}^0\rangle. \quad (30)$$

2.4.1 Mixing: Slow vs Fast

Fig. 9 shows in red the probability of finding an X^0 as a function of time (in units of lifetime, $1/\Gamma$) if the starting state is X^0 . In blue is the probability of starting with X^0 and finding \bar{X}^0 at time t . In all three panels $\Delta\Gamma = 0$ and $|p/q| = 1$ is assumed. In the left panel $\Delta M = \frac{1}{3}\Gamma$ so the oscillation is slow, while in the right panel $\Delta M = 3\Gamma$, the oscillation is fast. The middle panel is in-between, $\Delta M = \Gamma$. The three panels qualitatively show what is seen for D^0 , B^0 and B_s as we go from left to right.

To understand how the SM accounts for the slow *versus* fast oscillation behavior of the different neutral meson systems we need to look at the underlying process. Consider the box diagrams in Fig. 10. First note that each of the two fermion lines in each diagram will produce a modern GIM: the diagrams come with a factor of $(V_{qb}V_{qd,s}^*)^2$ with $q = u, c, t$, times m_q^2 dependent functions.

Next, let's recall the connection between the parameters of the 2×2 Hamiltonian and fundamental theory, Eqs. (25) and (26). In particular the presence of the delta function in Eq. (26) indicates that Γ_{12} originates in graphs where the intermediate states are on-shell. In the top box graph the intermediate states are W^+W^- which are much heavier than $B_{d,s}$ and therefore never on-shell. The upper panel box cannot contribute to Γ_{12} . Then modern GIM dictates the graph is dominated by the top quark exchange. The bottom panel box graph is a little different. It does not contribute to Γ_{12} when the intermediate state is $t\bar{t}$, but it does for $c\bar{c}$ and $u\bar{u}$. However, these contributions are much smaller than the ones with $t\bar{t}$ or the ones in the upper panel graph. So we conclude that Γ_{12} is negligible (compared to M_{12}) for $B^0 = B_d$ and B_s . From (22) we see that

$$\Gamma_{12} = 0 \quad \Rightarrow \quad \Delta M = 2|M_{12}| \quad \Rightarrow \quad \frac{p}{q} = \frac{M_{12}}{|M_{12}|}$$

That is p/q is a pure phase, $|p/q| = 1$. Moreover, the phase originates in the KM factors in the Feynman graph, because there is no imaginary part produced by the loop integration since intermediate states

cannot go on-shell (the very same reason $\Gamma_{12} = 0$). So we can read off the phase immediately:

$$\left(\frac{p}{q}\right)_{B^0} = \frac{(V_{tb}V_{td}^*)^2}{|V_{tb}V_{td}^*|^2}, \quad \left(\frac{p}{q}\right)_{B_s} = \frac{(V_{tb}V_{ts}^*)^2}{|V_{tb}V_{ts}^*|^2}.$$

Of course, we cannot compute ΔM fully, but we can compare this quantity for B^0 and B_s . In particular, in the flavor- $SU(3)$ symmetry limit the strong interactions treat the B^0 and B_s identically, so the only difference in the evaluation of M_{12} stems from the KM factors. So to the accuracy that $SU(3)$ may hold (typically 20%), we have

$$\frac{(\Delta M)_{B_s}}{(\Delta M)_{B^0}} = \left| \frac{V_{ts}}{V_{td}} \right|^2$$

Let's look back at Fig. 5. We can understand a lot of it now. For example, the most stringent bound is from CP violation in $K^0 - \bar{K}^0$ mixing. We have seen that this requires $\text{Im}M_{12} \neq 0$ or $\text{Im}\Gamma_{12} \neq 0$. Now we can write, roughly, that the imaginary part of the box diagram for K^0 mixing gives

$$\text{Im}M_{12} \approx \text{Im} \left(\bar{K}^0 \text{---} \left[\begin{array}{c} \xrightarrow{s} \quad \xrightarrow{u, c, t} \quad \xrightarrow{d} \\ \text{---} W \text{---} \quad \text{---} W \text{---} \\ \xleftarrow{d} \quad \xleftarrow{u, c, t} \quad \xleftarrow{s} \end{array} \right] K^0 \right) \sim$$

$$\text{Im} \left[\frac{G_F^2 M_W^2}{4\pi^2} \sum_{q, q' = u, c, t} V_{qd}^* V_{qs} V_{q'd} V_{q's} f(m_q, m_{q'}) \langle K^0 | \bar{d}_L \gamma^\mu s_L \bar{d}_L \gamma_\mu s_L | \bar{K}^0 \rangle \right]$$

Here f is a dimensionless function that is computed from a Feynman integral of the box diagram and depends on M_W implicitly. Note that the diagram has a double GIM, one per quark line. In the second line above, the non-zero imaginary part is from the phase in the KM-matrix. In the standard parametrization V_{ud} and V_{us} are real, so we need at least one heavy quark in the Feynman diagram to get a non-zero imaginary part. One can show that the diagram with one u quark and one heavy, c or t , quark is suppressed. We are left with c and t contributions only. Notice also that KM-unity gives $\sum_q V_{qd}^* V_{qs} = 0$, and since $\text{Im}V_{ud}^* V_{us} = 0$, we have a single common coefficient, $\text{Im}V_{cd}^* V_{cs} = -\text{Im}V_{td}^* V_{ts} = A^2 \lambda^5 \eta$ in terms of the Wolfenstein parametrization. Taking only the top contribution we can compare with the contribution from new physics which we parametrize as

$$\frac{1}{\Lambda^2} \langle K^0 | \bar{d}_L \gamma^\mu s_L \bar{d}_L \gamma_\mu s_L | \bar{K}^0 \rangle$$

Comparing to the SM results and assuming the SM approximately accounts for the observed quantity, this gives

$$\Lambda^2 \gtrsim \frac{4\pi^2}{G_F^2 M_W^2} \frac{1}{|V_{td}^* V_{ts}|} \approx \left[\frac{6}{(10^{-5})(10^2)} \frac{1}{(0.04)(0.004)} \text{GeV} \right]^2 \approx [4 \times 10^4 \text{TeV}]^2$$

Exercises

Exercise 2.4.1-1: Challenge: Can you check the other three mixing bounds in Fig. 5 (assuming the SM gives about the right result).

2.5 CPV

We now turn our attention to CP violation, or CPV for short. There are several ways of measuring CPV. Some of them are associated with mixing, some with decay and some with both at once. We will take a look at each of these.

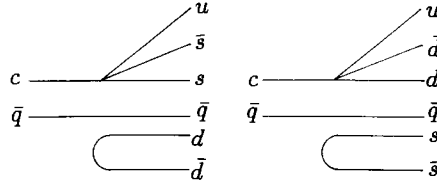


Fig. 11: Sample Feynman diagrams for some D -meson decay.

2.5.1 CPV in Decay

We begin by looking at CPV in decay. This has nothing to do with mixing per-se. It is conceptually simple but the price we pay for this simplicity is that they are hard to compute from first principles. We will see later that in some cases CPV in interference between mixing and decay can be accurately predicted.

Very generally we define an asymmetry as

$$\mathcal{A} = \frac{\Gamma - \bar{\Gamma}}{\Gamma + \bar{\Gamma}}$$

where Γ is some rate for some process and $\bar{\Gamma}$ is the rate for the process conjugated under something, like C , or P or $\theta \rightarrow \pi - \theta$ (Forward-backward asymmetry). For a CP decay asymmetry in the decay $X \rightarrow f$ we have

$$\mathcal{A} = \frac{|\langle f|X\rangle|^2 - |\langle \bar{f}|\bar{X}\rangle|^2}{|\langle f|X\rangle|^2 + |\langle \bar{f}|\bar{X}\rangle|^2}$$

where the \bar{X} and \bar{f} are the CP conjugates of X and f respectively.

Fig. 11 shows diagrams for a D -meson decay. The two diagrams produce the same final state, so they both contribute to the decay amplitude. The W exchange is shown as a 4-fermion point vertex. The first diagram contains a KM factor of $V_{cs}^*V_{us}$ while the second has a factor of $V_{cd}^*V_{ud}$. So in preparation for a computation of the CPV decay asymmetry we write

$$\begin{aligned} \langle f|X\rangle &= aA + bB \\ \langle \bar{f}|\bar{X}\rangle &= a^*\bar{A} + b^*\bar{B} \end{aligned}$$

where $a = V_{cs}^*V_{us}$ and $b = V_{cd}^*V_{ud}$ and the rest are matrix elements computed in the presence of strong interactions

$$\begin{aligned} A &= \langle f|(\bar{u}_L\gamma^\mu s_L)(\bar{s}_L\gamma_\mu c_L)|D\rangle \\ B &= \langle f|(\bar{u}_L\gamma^\mu d_L)(\bar{d}_L\gamma_\mu c_L)|D\rangle. \end{aligned}$$

While we cannot compute these, we can say something useful about them. Assuming the strong interactions are invariant under CP we have $\bar{A} = A$ and $\bar{B} = B$. This is easy to show:

$$\begin{aligned} A &= \langle f|(\bar{u}_L\gamma^\mu s_L)(\bar{s}_L\gamma_\mu c_L)|D\rangle \\ &= \langle f|(CP)^{-1}(CP)(\bar{u}_L\gamma^\mu s_L)(\bar{s}_L\gamma_\mu c_L)(CP)^{-1}(CP)|D\rangle \\ &= \langle \bar{f}|(\bar{s}_L\gamma^\mu u_L)(\bar{c}_L\gamma_\mu s_L)|\bar{D}\rangle \\ &= \bar{A} \end{aligned}$$

Using this and plugging into the above definition of the asymmetry \mathcal{A} we have

$$\mathcal{A} = \frac{2\text{Im}(a^*b)\text{Im}(A^*B)}{|aA|^2 + |bB|^2 + 2\text{Re}(a^*b)\text{Re}(A^*B)} \quad (31)$$

In order that CP be violated in the decay it is necessary that we have a relative phase between a and b and also between A and B . The first one is from the KM matrix, but the second requires computation of non-trivial strongly interaction matrix elements. Note that

$$\text{Im}(a^*b) = \text{Im}((V_{cs}^*V_{us})^*V_{cd}^*V_{ud}) = \text{Im}(V_{cs}V_{cd}^*V_{ud}V_{us}^*) = J$$

so, as promised, the Jarlskog determinant must be non-zero in order to see CPV.

There are numerous CPV decay asymmetries listed in the PDG. It is too bad we cannot use them to extract the KM angles precisely, let alone test for new physics (because of our inability to compute the strong interaction matrix elements).

2.5.2 CPV in Mixing

We will look at the case of kaons first and come back to heavy mesons later. This is partly because CPV was discovered through CPV in mixing in kaons. But also because it offers a special condition not found in other neutral meson mixing: the vast difference in lifetimes between eigenstates allows clean separation between them.

This allows us to meaningfully define the K_L semileptonic decay charge-asymmetry, which is a measure of CP violation:

$$\delta = \frac{\Gamma(K_L \rightarrow \pi^- e^+ \nu) - \Gamma(K_L \rightarrow \pi^+ e^- \bar{\nu})}{\Gamma(K_L \rightarrow \pi^- e^+ \nu) + \Gamma(K_L \rightarrow \pi^+ e^- \bar{\nu})}$$

In order to compute this we use the expansion of K_L in terms of flavor eigenstates K^0 and \bar{K}^0 of Eq. (23), and note that the underlying process is $s \rightarrow ue^- \bar{\nu}$ (or $\bar{s} \rightarrow \bar{u}e^+ \nu$) so that we assume $\langle \pi^- e^+ \nu | H_W | \bar{K}^0(t) \rangle = 0 = \langle \pi^+ e^- \bar{\nu} | H_W | K^0(t) \rangle$. Moreover, we assume CPV is in the mixing only (through the parameter ϵ) and therefore assume that CP is a good symmetry of the decay amplitude: $\langle \pi^- e^+ \nu | H_W | K^0(t) \rangle = \langle \pi^+ e^- \bar{\nu} | H_W | \bar{K}^0(t) \rangle$.

Exercises

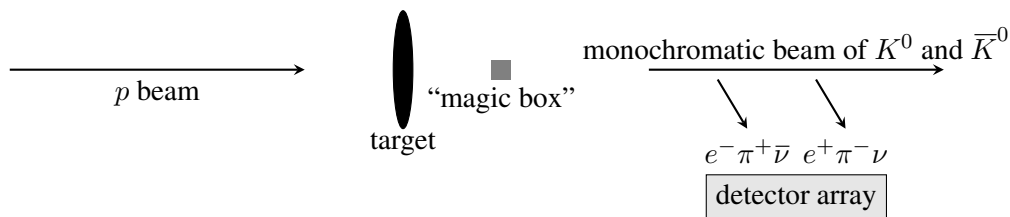
Exercise 2.5.2-1: With these assumptions show

$$\delta = \frac{|1 + \epsilon|^2 - |1 - \epsilon|^2}{|1 + \epsilon|^2 + |1 - \epsilon|^2} \approx 2\text{Re}\epsilon$$

Experimental measurement gives $\delta_{\text{exp}} = 0.330 \pm 0.012\%$, from which $\text{Re}\epsilon \simeq 1.65 \times 10^{-3}$.

2.5.2.1 Example: Time dependent asymmetry in semileptonic K decay (“ $K_{\ell 3}$ decay”).

This is the time dependent analogue of δ above. The experimental set-up is as follows:



The proton beam hits a target, and the magic box produces a clean monochromatic beam of neutral K mesons. These decay in flight and the semileptonic decays are registered in the detector array. We denote by N_{K^0} the number of K^0 -mesons, and by $N_{\bar{K}^0}$ that of \bar{K}^0 -mesons, from the beam. Measure

$$\delta(t) = \frac{N^+ - N^-}{N^+ + N^-}$$

as a function of distance from the beam (which can be translated into time from production at the magic box). Here N^\pm refers to the total number of $K_{\ell 3}$ events observed with charge \pm lepton. In reality “ π^\pm ” really stands for “hadronic stuff” since only the electrons are detected. We have then,

$$\delta(t) = \frac{N_{K^0} [\Gamma(K^0(t) \rightarrow \pi^- e^+ \nu) - \Gamma(K^0(t) \rightarrow \pi^+ e^- \bar{\nu})] + N_{\bar{K}^0} [\Gamma(\bar{K}^0(t) \rightarrow \pi^- e^+ \nu) - \Gamma(\bar{K}^0(t) \rightarrow \pi^+ e^- \bar{\nu})]}{N_{K^0} [\Gamma(K^0(t) \rightarrow \pi^- e^+ \nu) + \Gamma(K^0(t) \rightarrow \pi^+ e^- \bar{\nu})] + N_{\bar{K}^0} [\Gamma(\bar{K}^0(t) \rightarrow \pi^- e^+ \nu) + \Gamma(\bar{K}^0(t) \rightarrow \pi^+ e^- \bar{\nu})]}$$

The calculation of $\delta(t)$ in terms of the mixing parameters q and p and the mass and width differences is much like the calculation of δ above so, again, I leave it as an exercise:

Exercises

Exercise 2.5.2-2: Use $\Gamma(K^0(t) \rightarrow \pi^- e^+ \nu) \propto |\langle \pi^- e^+ \nu | H_W | K^0(t) \rangle|^2$ and the assumptions that

- (i) $\langle \pi^- e^+ \nu | H_W | \bar{K}^0(t) \rangle = 0 = \langle \pi^+ e^- \nu | H_W | K^0(t) \rangle$
- (ii) $\langle \pi^- e^+ \nu | H_W | K^0(t) \rangle = \langle \pi^+ e^- \nu | H_W | \bar{K}^0(t) \rangle$

to show that

$$\delta(t) = \frac{(N_{K^0} - N_{\bar{K}^0}) \left[|f_+(t)|^2 - |f_-(t)|^2 \frac{1}{2} \left(\left| \frac{q}{p} \right|^2 + \left| \frac{p}{q} \right|^2 \right) \right] + \frac{1}{2} (N_{K^0} + N_{\bar{K}^0}) |f_-(t)|^2 \left(\left| \frac{p}{q} \right|^2 - \left| \frac{q}{p} \right|^2 \right)}{(N_{K^0} + N_{\bar{K}^0}) \left[|f_+(t)|^2 + |f_-(t)|^2 \frac{1}{2} \left(\left| \frac{q}{p} \right|^2 + \left| \frac{p}{q} \right|^2 \right) \right] - \frac{1}{2} (N_{K^0} - N_{\bar{K}^0}) |f_-(t)|^2 \left(\left| \frac{p}{q} \right|^2 - \left| \frac{q}{p} \right|^2 \right)}$$

Justify assumptions (i) and (ii).

The formula in the exercise is valid for any X^0 - \bar{X}^0 system. We can simplify further for kaons, using $p/q = (1 + \epsilon)/(1 - \epsilon)$, $a \equiv (N_{K^0} - N_{\bar{K}^0})/(N_{K^0} + N_{\bar{K}^0})$ and $\Delta\Gamma \approx -\Gamma_S$. Then

$$\begin{aligned} \delta(t) &= \frac{a \left[|f_+(t)|^2 - |f_-(t)|^2 \right] + 4\text{Re}(\epsilon) |f_-(t)|^2}{\left[|f_+(t)|^2 + |f_-(t)|^2 \right] - 4a\text{Re}(\epsilon) |f_-(t)|^2} \\ &\approx \frac{2ae^{-\frac{1}{2}\Gamma_S t} \cos(\Delta M t) + (1 + e^{-\Gamma_S t} - 2e^{-\frac{1}{2}\Gamma_S t} \cos(\Delta M t)) 2 \left(1 + \frac{a}{2}\right) \text{Re}(\epsilon)}{1 + e^{-\Gamma_S t}} \end{aligned} \quad (32)$$

Figure 12 shows the experimental measurement of the asymmetry [21]. The solid curve is a fit to the formula (32) from which the parameters Γ_S , ΔM , a and $\text{Re}(\epsilon)$ are extracted. The fit to this figure gives $\Delta M_K = (0.5287 \pm 0.0040) \times 10^{10} \text{ s}^{-1}$. The current value, from the PDG is $\Delta M_K = (0.5293 \pm 0.0009) \times 10^{10} \text{ s}^{-1}$.

2.6 CP-Asymmetries: Interference of Mixing and Decay

We have seen in (31) that in order to generate a non-vanishing CP-asymmetry we need two amplitudes that can interfere. One way to get an interference is to have two “paths” from $|in\rangle$ to $|out\rangle$. For example,

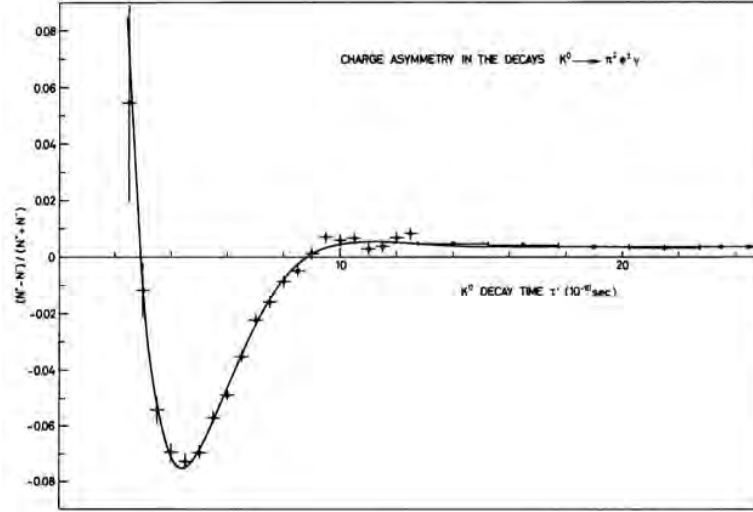
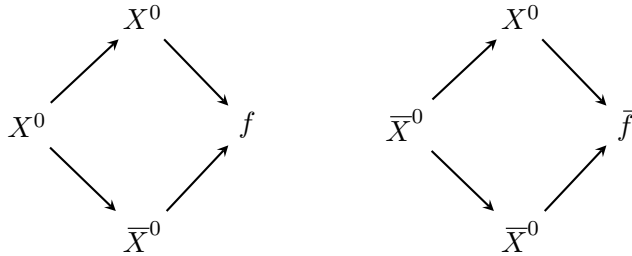


Fig. 12: Charge asymmetry in semi-leptonic neutral kaon decays, from an experiment by Gjesdal *et al*, [21]. The solid curve is a fit to the formula (32) from which the parameters Γ_S , ΔM , a and $\text{Re}(\epsilon)$ are extracted.

consider an asymmetry constructed from $\Gamma = \Gamma(X^0 \rightarrow f)$ and $\bar{\Gamma} = \Gamma(\bar{X}^0 \rightarrow \bar{f})$, where f stands for some final state and \bar{f} for its CP conjugate. Then Γ may get contributions either from a direct decay $X^0 \rightarrow f$ or it may first oscillate into \bar{X}^0 and then decay $\bar{X}^0 \rightarrow f$. Note that this requires that both X^0 and its antiparticle, \bar{X}^0 , decay to the same common state. Similarly for $\bar{\Gamma}$ we may get contributions from both $\bar{X}^0 \rightarrow \bar{f}$ and the oscillation of \bar{X}^0 into X^0 followed by a decay into \bar{f} . In pictures,



Concretely,

$$\begin{aligned} \Gamma(X^0(t) \rightarrow f) &\propto |f_+(t)\langle f|H_w|X^0\rangle + f_-(t)\frac{q}{p}\langle f|H_w|\bar{X}^0\rangle|^2 \\ &\equiv |f_+(t)A_f + f_-(t)\frac{q}{p}\bar{A}_f|^2 \\ \Gamma(\bar{X}^0(t) \rightarrow \bar{f}) &\propto |f_-(t)\frac{p}{q}\langle \bar{f}|H_w|X^0\rangle + f_+(t)\langle \bar{f}|H_w|\bar{X}^0\rangle|^2 \\ &\equiv |\frac{p}{q}f_-(t)A_{\bar{f}} + f_+(t)\bar{A}_{\bar{f}}|^2 \end{aligned}$$

I hope the notation, which is pretty standard, is not just self-explanatory, but fairly explicit. The bar over an amplitude A refers to the decaying state being \bar{X}^0 , while the decay product is explicitly given by the subscript, *e.g.*, $\bar{A}_{\bar{f}} = \langle \bar{f}|H_w|\bar{X}^0\rangle$.

Exercises

Exercise 2.6-1: If f is an eigenstate of the strong interactions, show that CPT implies $|A_f|^2 = |\bar{A}_{\bar{f}}|^2$ and $|A_{\bar{f}}|^2 = |\bar{A}_f|^2$

The time dependent asymmetry is

$$\mathcal{A}(t) = \frac{\Gamma(\bar{X}^0(t) \rightarrow \bar{f}) - \Gamma(X^0(t) \rightarrow f)}{\Gamma(\bar{X}^0(t) \rightarrow \bar{f}) + \Gamma(X^0(t) \rightarrow f)}$$

and the time integrated asymmetry is

$$a = \frac{\Gamma(\bar{X}^0 \rightarrow \bar{f}) - \Gamma(X^0 \rightarrow f)}{\Gamma(\bar{X}^0 \rightarrow \bar{f}) + \Gamma(X^0 \rightarrow f)}$$

where $\Gamma(X^0 \rightarrow f) \equiv \int_0^\infty dt \Gamma(X^0(t) \rightarrow f)$, and likewise for the CP conjugate. These are analogs of the quantities we called $\delta(t)$ and δ we studied for kaons.

2.6.1 Semileptonic

We take $f = e^- + \text{any}$. Note that we are taking the wrong sign decay of X^0 . That is, $\bar{b} \rightarrow \bar{c}e^+\nu$ implies $X^0 \rightarrow e^+ + \text{any}$ so that $A_f = 0$. Similarly, $b \rightarrow ce^-\bar{\nu}$ implies $\bar{X}^0 \rightarrow e^- + \text{any}$ so that $\bar{A}_{\bar{f}} = 0$. Therefore we have $\Gamma(X^0(t) \rightarrow f) = |\frac{q}{p}f_-(t)\bar{A}_f|^2$ and $\Gamma(\bar{X}^0(t) \rightarrow \bar{f}) = |\frac{p}{q}f_-(t)A_{\bar{f}}|^2$. We obtain

$$\mathcal{A}_{\text{SL}}(t) = \frac{\left| \frac{p}{q} \right|^2 - \left| \frac{q}{p} \right|^2}{\left| \frac{p}{q} \right|^2 + \left| \frac{q}{p} \right|^2}$$

Comments:

- (i) This is useful because it directly probes $|q/p|$ without contamination from other quantities, in particular from those that require knowledge of strong interactions.
- (ii) We started off with an *a priori* time dependent quantity, but discovered it is time independent.
- (iii) We already saw that in the SM this is expected to vanish to high accuracy for B mesons, because Γ_{12} is small.
- (iv) It is not expected to vanish identically because Γ_{12} while small is non-vanishing. We can guesstimate,

$$B^0 : \mathcal{A}_{\text{SL}}^d = \mathcal{O} [(m_c^2/m_t^2) \sin \beta] \lesssim 10^{-3}, \quad B^s : \mathcal{A}_{\text{SL}}^s = \mathcal{O} [(m_c^2/m_t^2) \sin \beta_s] \lesssim 10^{-4}.$$

(v) Experiment:

$$\begin{aligned} \mathcal{A}_{\text{SL}}^d &= (+0.7 \pm 2.7) \times 10^{-3} & \Rightarrow & |q/p| = 0.9997 \pm 0.0013 \\ \mathcal{A}_{\text{SL}}^s &= (-17.1 \pm 5.5) \times 10^{-3} & \Rightarrow & |q/p| = 1.0086 \pm 0.0028 \end{aligned}$$

For the rest of this section we will make the approximation that $|q/p| = 1$. In addition, we will assume $\Delta\Gamma$ is negligible. We have seen why this is a good approximation. In fact, for the case of B^0 , $\Delta\Gamma/\Gamma \sim 10^{-2}$, while for B_s the ratio is about 10%. This simplifies matters because in this approximation

$$f_{\pm}(t) \approx e^{-iMt} e^{-\frac{1}{2}\Gamma t} \begin{cases} \cos(\frac{1}{2}\Delta Mt) \\ -i \sin(\frac{1}{2}\Delta Mt) \end{cases}$$

2.6.2 CPV in interference between a decay with mixing and a decay without mixing

Assume $\bar{f} = \pm f$. Such self-conjugate states are easy to come by. For example D^+D^- or, to good approximation, $J/\psi K_S$. Now, in this case we have $A_{\bar{f}} = \pm A_f$ and $\bar{A}_{\bar{f}} = \pm \bar{A}_f$. Our formula for the asymmetry now takes the form

$$\mathcal{A}_{f_{CP}} = \frac{|\frac{p}{q}f_-(t)A_f + f_+(t)\bar{A}_f|^2 - |f_+(t)A_f + \frac{q}{p}f_-(t)\bar{A}_f|^2}{|\frac{p}{q}f_-(t)A_f + f_+(t)\bar{A}_f|^2 + |f_+(t)A_f + \frac{q}{p}f_-(t)\bar{A}_f|^2}$$

Now, dividing by $A_f|^2$ and defining

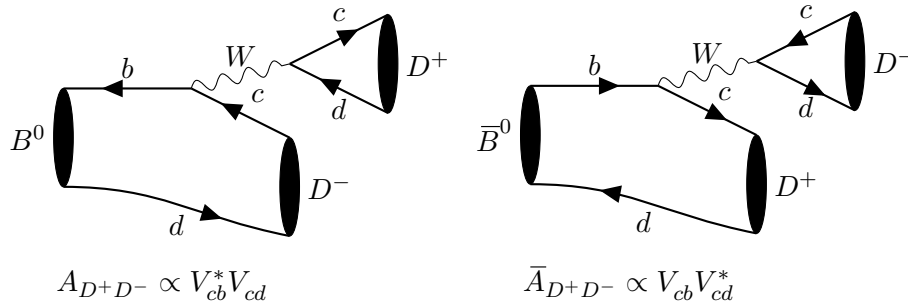
$$\lambda_f = \frac{q}{p} \frac{\bar{A}_f}{A_f}$$

we have

$$\begin{aligned} \mathcal{A}_{f_{CP}} &= \frac{|f_-(t) + f_+(t)\lambda_f|^2 - |f_+(t) + f_-(t)\lambda_f|^2}{|f_-(t) + f_+(t)\lambda_f|^2 + |f_+(t) + f_-(t)\lambda_f|^2} \\ &= -\frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2} \cos(\Delta Mt) + \frac{2\text{Im}\lambda_f}{1 + |\lambda_f|^2} \sin(\Delta Mt) \\ &\equiv -C_f \cos(\Delta Mt) + S_f \sin(\Delta Mt) \end{aligned}$$

Here is what is amazing about this formula, for which Bigi and Sanda [22] were awarded the Sakurai Prize for Theoretical Particle Physics: the coefficients C_f and S_f can be computed in terms of KM elements only. They are independent of non-computable, non-perturbative matrix elements. The point is that what most often frustrates us in extracting fundamental parameters from experiment is our inability to calculate in terms of the parameters to be measured and, at most, other known parameters. I now explain the claim that C_f and S_f are calculable and its range of validity.

The leading contributions to the processes $B^0 \rightarrow f$ and $\bar{B}^0 \rightarrow f$ in the case $f = D^+D^-$ are shown in the following figures:



Either using CP symmetry of the strong interactions or noting that as far as the strong interactions are concerned the two diagrams are identical, we have

$$\frac{\bar{A}_{D^+D^-}}{A_{D^+D^-}} = \frac{V_{cb} V_{cd}^*}{V_{cb}^* V_{cd}}$$

Since $|\bar{A}_{D^+D^-}/A_{D^+D^-}| = 1$, this is a pure phase, and we see that the phase is given purely in terms of KM elements.

To complete the argument we need q/p . But we have already seen that Γ_{12} is negligible. Hence

$$\frac{p}{q} = \frac{2M_{12}}{\Delta M} = \frac{\Delta M}{2M_{12}^*} = \frac{M_{12}}{|M_{12}|} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}$$

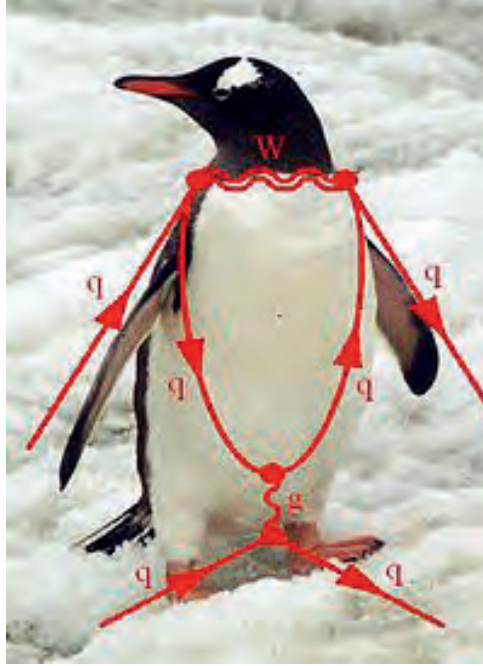


Fig. 13: Penguin Feynman diagram.

Collecting results

$$\text{Im}(\lambda_{D^+D^-}) = \text{Im}\left(\frac{V_{cb}V_{cd}^*V_{tb}^*V_{td}}{V_{cb}^*V_{cd}V_{tb}V_{td}^*}\right) = \text{Im}(e^{2i\beta}) = \sin(2\beta)$$

and the asymmetry parameters are $C_{D^+D^-} = 0$ and $S_{D^+D^-} = \sin(2\beta)$. Measurements of the asymmetry gives (twice the sine of) one of the angles of the unitarity triangle without hadronic uncertainties!

More generally, precisely as in the case of direct CPV we can have several terms contributing to A_f , each with different combinations of KM elements:

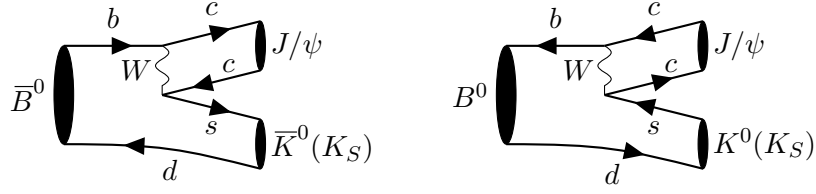
$$\begin{aligned} A_f &= aT + bP, \\ \bar{A}_f &= a^*T + b^*P, \end{aligned}$$

where a and b are KM elements and T and P are matrix elements. A word about notation. T stands for “tree” because we have in mind a contribution that at the quark level and before dressing up with gluons is a Feynman diagram at tree level. P stands for “penguin” and represents a contribution that at the quark level starts at 1-loop. *Digression:* I do not know why this is called so. I have heard many stories. It was certainly first introduced in the context we are studying. Fig. 13 shows a penguin-like depiction of the diagram. *End digression.* The trick is to find processes where the penguin contribution is expected to be suppressed. Suppose $|P|=0$. Then

$$\lambda_f = \frac{q}{p} \frac{a^*}{a}.$$

This is the same result as above, only emphasizing the hidden assumption.

The most celebrated case is $B \rightarrow J/\psi K_S$. Here are the leading diagrams:



Generally we should write

$$\frac{\bar{A}_{\psi K_S}}{A_{\psi K_S}} = -\frac{(V_{cb}V_{cs}^*)T + (V_{ub}V_{us}^*)P}{(V_{cb}^*V_{cs})T + (V_{ub}^*V_{us})P} \times \frac{V_{cd}^*V_{cs}}{V_{cd}V_{cs}^*}$$

The novelty here is the last factor which arises from projecting the K^0 and \bar{K}^0 states onto K_S . Using Using (27) with L and S for H and L , respectively, this is just $-q/p = -V_{cd}^*V_{cs}/V_{cd}V_{cs}^*$. Now in this case the penguin contribution is suppressed by a 1-loop factor relative to the tree level contribution and in addition the KM factor of the penguin contribution is very suppressed relative to that in the tree contribution: counting powers of Wolfenstein's λ parameter $|V_{ub}V_{us}^*|/|V_{cb}V_{cs}^*| \sim \lambda^2$. Safely neglecting P we have

$$\lambda_{\psi K_S} = -e^{-2i\beta} \quad S_{\psi K_S} = \sin(2\beta), \quad C_{\psi K_S} = 0$$

The PDG values are

$$S_{\psi K_S} = +0.682 \pm 0.019, \quad C_{\psi K_S} = (0.5 \pm 2.0) \times 10^{-2}.$$

The vanishing of $C_{\psi K_S}$ is reassuring, we must know what we are doing!

How about other angles? We can get $\sin(2\alpha)$ from $B \rightarrow \pi\pi$ if the penguin can be neglected in

$$\frac{\bar{A}_{\pi\pi}}{A_{\pi\pi}} = \frac{(V_{ub}V_{ud}^*)T + (V_{tb}V_{td}^*)P}{(V_{ub}^*V_{ud})T + (V_{tb}^*V_{td})P}$$

It was realized well before the experiment was performed that the penguin here cannot be expected to be negligible [23]. The PDG gives the measured value $C_{\pi^+\pi^-} = -0.31 \pm 0.05$ confirming this expectation. This can be fixed by determining P/T from an isospin analysis and measurement of several rates and asymmetries [24]. But the analysis is difficult and compromises the precision in the determination of α . The moral is that you must have a good reason to neglect P before you can claim a clean determination of the angles of the unitarity triangle.

Exercises

Exercise 2.6.2-2: The following table is reproduced from the PDG.

$\bar{b} \rightarrow \bar{q}q\bar{q}'$	$B^0 \rightarrow f$	$B_s \rightarrow f$	CKM dependence of A_f	Suppression
$\bar{b} \rightarrow \bar{c}c\bar{s}$	ψK_S	$\psi\phi$	$(V_{cb}^*V_{cs})T + (V_{ub}^*V_{us})P^u$	loop $\times \lambda^2$
$\bar{b} \rightarrow \bar{s}s\bar{s}$	ϕK_S	$\phi\phi$	$(V_{cb}^*V_{cs})P^c + (V_{ub}^*V_{us})P^u$	λ^2
$\bar{b} \rightarrow \bar{u}u\bar{s}$	$\pi^0 K_S$	$K^+ K^-$	$(V_{cb}^*V_{cs})P^c + (V_{ub}^*V_{us})T$	λ^2/loop
$\bar{b} \rightarrow \bar{c}c\bar{d}$	$D^+ D^-$	ψK_S	$(V_{cb}^*V_{cd})T + (V_{tb}^*V_{td})P^t$	loop
$\bar{b} \rightarrow \bar{s}s\bar{d}$	$K_S K_S$	ϕK_S	$(V_{tb}^*V_{td})P^t + (V_{cb}^*V_{cd})P^c$	$\lesssim 1$
$\bar{b} \rightarrow \bar{u}u\bar{d}$	$\pi^+ \pi^-$	$\rho^0 K_S$	$(V_{ub}^*V_{ud})T + (V_{tb}^*V_{td})P^t$	loop

The columns from left to right give the underlying quark process, the final state in B^0 decay, the final state in B_s decay, an expression for the amplitude including KM factors and T or P for whether the underlying process is tree level or penguin, and lastly, suppression factor of the sub-leading contribution to the amplitude relative to the leading one. Note that in some cases both contributions to the amplitude are from 1-loop diagrams, so they are both labeled P . Reproduce the last column (we have done the first line already). Find S_f in each case, assuming you can neglect the suppressed amplitude.

Acknowledgments

I would like to thank the organizers of and support staff for CLASHEP for their support throughout the process of preparing the lectures, traveling, presenting and finally producing this report, and in particular Martijn Mulders, Nick Ellis, Kate Ross and Edgar Carrera. This work is supported in part by the US Department of Energy under grant DE-SC0009919.

Bibliography

- [1] CKMfitter Group (J. Charles et al.), Eur. Phys. J. C **41**, 1-131 (2005) [hep-ph/0406184], updated results and plots available at: <http://ckmfitter.in2p3.fr>
- [2] K.A. Olive et al. (Particle Data Group), Chin. Phys. C, , 090001 (2014).
- [3] B. Grinstein, Ann. Rev. Nucl. Part. Sci. **42** (1992) 101.
- [4] M. E. Luke, Phys. Lett. B **252**, 447 (1990).
- [5] N. Isgur, D. Scora, B. Grinstein and M. B. Wise, Phys. Rev. D **39**, 799 (1989).
- [6] J. Chay, H. Georgi and B. Grinstein, Phys. Lett. B **247**, 399 (1990).
- [7] A. V. Manohar and M. B. Wise, Phys. Rev. D **49**, 1310 (1994) [hep-ph/9308246].
- [8] I. I. Y. Bigi, N. G. Uraltsev and A. I. Vainshtein, Phys. Lett. B **293**, 430 (1992) [Erratum-ibid. B **297**, 477 (1993)] [hep-ph/9207214].
- [9] C. W. Bauer, Z. Ligeti, M. Luke, A. V. Manohar and M. Trott, Phys. Rev. D **70**, 094017 (2004) [hep-ph/0408002].
- [10] M. Clements, C. Footman, A. S. Kronfeld, S. Narasimhan and D. Photiadis, Phys. Rev. D **27**, 570 (1983).
- [11] S. L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. D **2**, 1285 (1970).
- [12] M. Dine and A. E. Nelson, Phys. Rev. D **48** (1993) 1277 [hep-ph/9303230]. M. Dine, A. E. Nelson and Y. Shirman, Phys. Rev. D **51** (1995) 1362 [hep-ph/9408384]. M. Dine, A. E. Nelson, Y. Nir and Y. Shirman, Phys. Rev. D **53** (1996) 2658 [hep-ph/9507378].

- [13] T. Aaltonen *et al.* [CDF Collaboration], Phys. Rev. D **83**, 112003 (2011) [arXiv:1101.0034 [hep-ex]]. V. M. Abazov *et al.* [D0 Collaboration], Phys. Rev. D **84**, 112005 (2011) [arXiv:1107.4995 [hep-ex]].
- [14] J. H. Kuhn and G. Rodrigo, JHEP **1201**, 063 (2012) [arXiv:1109.6830 [hep-ph]]. W. Hollik and D. Pagani, Phys. Rev. D **84**, 093003 (2011) [arXiv:1107.2606 [hep-ph]].
- [15] S. Leone [CDF and D0 Collaboration], Nuovo Cim. C **037**, no. 02, 40 (2014).
- [16] B. Grinstein, A. L. Kagan, M. Trott and J. Zupan, Phys. Rev. Lett. **107**, 012002 (2011) [arXiv:1102.3374 [hep-ph]]. *idem*, JHEP **1110**, 072 (2011) [arXiv:1108.4027 [hep-ph]].
- [17] S. Venditti [NA62 Collaboration], Nuovo Cim. B **123**, 844 (2008). M. Akashi-Ronquest [KTeV Collaboration], arXiv:1003.5574 [hep-ex]. A. V. Artamonov *et al.* [BNL-E949 Collaboration], Phys. Rev. D **79**, 092004 (2009) [arXiv:0903.0030 [hep-ex]]. T. Abe *et al.* [Belle-II Collaboration], arXiv:1011.0352 [physics.ins-det].
- [18] S. Wandernoth, “Precision measurement of the oscillation frequency in the $B_s^0 - \bar{B}_s^0$ system,” Proceedings of the 48th Rencontres de Moriond on Electroweak Interactions and Unified Theories. La Thuile, Italy, March 2-9, 2013,
- [19] P. Huet and M. E. Peskin, Nucl. Phys. B **434**, 3 (1995) [hep-ph/9403257].
- [20] A. Messiah, *Quantum Mechanics*, Volume II, North Holland Publishing Company, 1965.
- [21] S. Gjesdal, G. Presser, T. Kamae, P. Steffen, J. Steinberger, F. Vannucci, H. Wahl and F. Eisele *et al.*, Phys. Lett. B **52**, 113 (1974).
- [22] I. I. Y. Bigi and A. I. Sanda, Nucl. Phys. B **193**, 85 (1981).
- [23] M. Gronau, Phys. Rev. Lett. **63**, 1451 (1989). D. London and R. D. Peccei, Phys. Lett. B **223**, 257 (1989). B. Grinstein, Phys. Lett. B **229**, 280 (1989).
- [24] M. Gronau and D. London, Phys. Rev. Lett. **65**, 3381 (1990).

Neutrino Physics

P. Hernández

IFIC, Universidad de València and CSIC, E-46071 Valencia, Spain

Abstract

This is the writeup of the lectures on neutrino physics delivered at various schools: TASI and Trieste in 2013 and the CERN-Latin American School in 2015. The topics discussed in this lecture include: general properties of neutrinos in the SM, the theory of neutrino masses and mixings (Dirac and Majorana), neutrino oscillations both in vacuum and in matter, as well as an overview of the experimental evidence for neutrino masses and of the prospects in neutrino oscillation physics. We also briefly review the relevance of neutrinos in leptogenesis and in beyond-the-Standard-Model physics.

Keywords

Neutrino; particle physics; lectures; neutrino oscillations; mixing; standard model.

1 Introduction

The history of neutrinos is tightly linked to that of the Standard Model. The discovery of neutrinos and the measurement of their tiny masses has been a scientific *tour de force*.

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



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

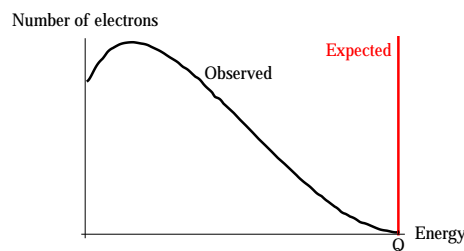


Fig. 1: Electron spectrum of β -decay.

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

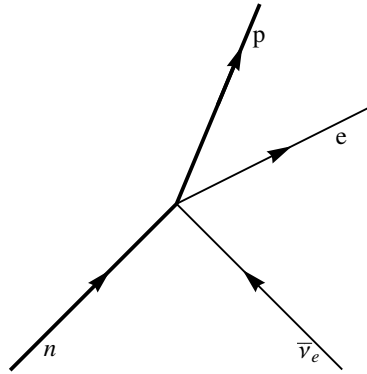


Fig. 2: Fermi four-fermion coupling responsible for β -decay.

a four-fermion interaction with strength given by G_F , the fermi constant. Such interaction implies that neutrinos should also scatter off matter through the inverse beta process, $\bar{\nu} p \rightarrow n e^+$. Bethe and Pearls [2] estimated the cross section for such process to be

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

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

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

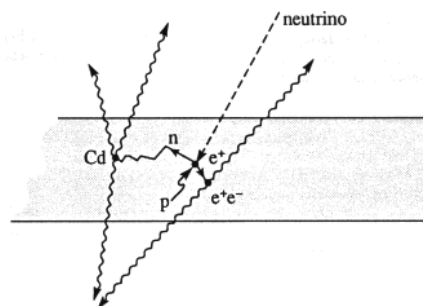


Fig. 3: Detection technique in the Reines-Cowan experiment.

Soon after anti-neutrinos were discovered, it was realised that they come in flavours. The muon had been discovered in cosmic rays much earlier, but it took a long time to understand that this particle was a heavier version of the electron and not the pion. The analogous of the β -process involving muons is pion decay

$$\pi^- \rightarrow \mu^- \bar{\nu}_\mu. \quad (3)$$

It was understood that also in this case a neutrino was being emitted but that such neutrino, accompanying the μ , had a different identity to that in β -decay. Since the energies involved in this process are higher than in β -decay and neutrino cross-sections grow fast with energy in the Fermi theory, it would actually be easier to detect this new type of neutrinos.

In 1962 Lederman, Schwartz and Steinberger (LSS) achieved this goal by creating the first accelerator neutrino beam [6]. In such a beam, an boosted proton beam hits a target producing pions and other hadrons that decay into neutrinos and other particles, mimicking what happens in cosmic rays. If a thick shield intercepts the beam, all particles except the neutrinos can be stopped, see Fig. 4. A neutrino detector is located behind the shield. A neutrino event will be seen from the appearance of a muon in the detector. Again this was such a great idea that we are still making discoveries with the modern versions of the LSS experiment, in the so-called conventional accelerator neutrino beams.

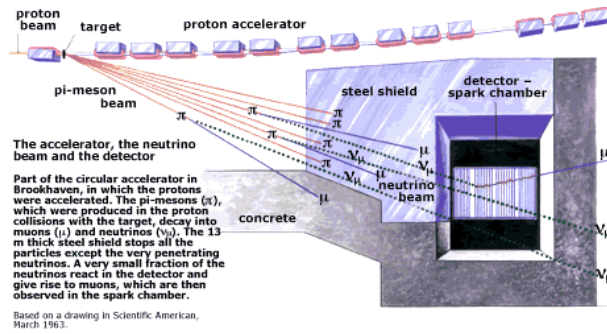


Fig. 4: Lederman, Schwartz, Steinberger experiment.

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



Fig. 5 shows the effect of a neutrino mass in the end-point electron energy spectrum in this decay. The best limit has been obtained by the Mainz and Troitsk experiments. The PDG combination gives [7]:

$$m_{\nu_e} < 2 \text{ eV} (95\% \text{CL}) . \tag{5}$$

The direct limits from processes involving μ, τ leptons are much weaker. The best limit on the ν_μ mass ($m_{\nu_\mu} < 170 \text{ keV}$ [8]) was obtained from the end-point spectrum of the decay $\pi^+ \rightarrow \mu^+ \nu_\mu$, while that

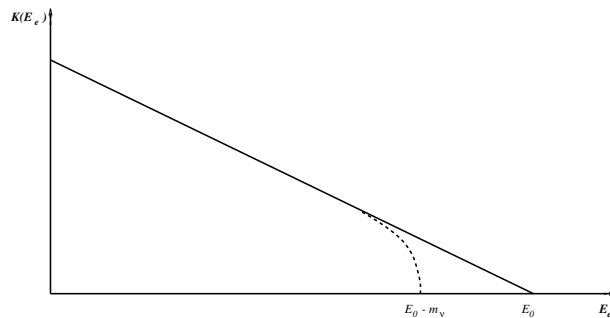


Fig. 5: Effect of a neutrino mass in the end-point of the lepton energy spectrum in β decay.

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

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

on the ν_τ mass was obtained at LEP ($m_{\nu_\tau} < 18.2$ MeV [9]) from the decay $\tau \rightarrow 5\pi\nu_\tau$. Neutrinos in the SM where therefore conjectured to be massless.

2 Neutrinos in the Standard Model

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

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

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

2.1 Chiral structure of the weak interactions

The left and right entries in table 1 have well defined chirality, negative and positive respectively. They are two-component spinors or Weyl fermions, that is the smallest irreducible representation of the Lorentz group representing spin 1/2 particles. Only fields with negative chirality (i.e. eigenvalue of γ_5 minus one) carry the $SU(2)$ charge. For free fermions moving at the speed of light (i.e., massless), it is easy to see that the chiral projectors are equivalent to the projectors on helicity components:

$$P_{R,L} \equiv \frac{1 \pm \gamma_5}{2} = \frac{1}{2} \left(1 \pm \frac{\mathbf{s} \cdot \mathbf{p}}{|\mathbf{p}|} \right) + O\left(\frac{m_i}{E}\right), \quad (6)$$

where the helicity operator $\Sigma = \frac{\mathbf{s} \cdot \mathbf{p}}{|\mathbf{p}|}$ measures the component of the spin in the direction of the momentum. Therefore for massless fermions only the left-handed states (with the spin pointing in the opposite direction to the momentum) carry $SU(2)$ charge. This is not inconsistent with Lorentz invariance, since for a fermion travelling at the speed of light, the helicity is the same in any reference frame. In other words, the helicity operator commutes with the Hamiltonian for a massless fermion and is thus a good quantum number.

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

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

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

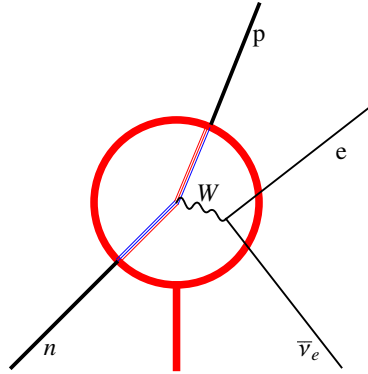


Fig. 6: β -decay process in the SM.

Neutrinos interact in the SM via charged and neutral currents:

$$\mathcal{L}_{SM} = \frac{g}{\sqrt{2}} \sum_{\alpha} \bar{\nu}_{\alpha} \gamma_{\mu} P_L l_{\alpha} W_{\mu}^{+} - \frac{g}{2 \cos \theta_W} \sum_{\alpha} \bar{\nu}_{\alpha} \gamma_{\mu} P_L \nu_{\alpha} Z_{\mu}^{+} + h.c. \quad (7)$$

The weak current is therefore $V-A$ since it only couples to the left fields: $\gamma_{\mu} P_L \propto \gamma_{\mu} - \gamma_{\mu} \gamma_5$. This structure is clearly seen in the kinematics of weak decays involving neutrinos, such as the classic example of pion decay to $e \bar{\nu}_e$ or $\mu \bar{\nu}_{\mu}$. In the limit of vanishing electron or muon mass, this decay is forbidden, because the spin of the initial state is zero and thus it is impossible to conserve simultaneously momentum and angular momentum if the two recoiling particles must have opposite helicities, as shown in Fig. 7. Thus the ratio of the decay rates to electrons and muons, in spite of the larger phase space in the former, is strongly suppressed by the factor $\left(\frac{m_e}{m_{\mu}}\right)^2 \sim 2 \times 10^{-5}$.

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

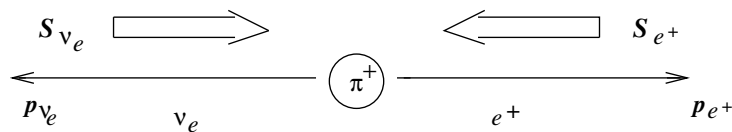


Fig. 7: Kinematics of pion decay.

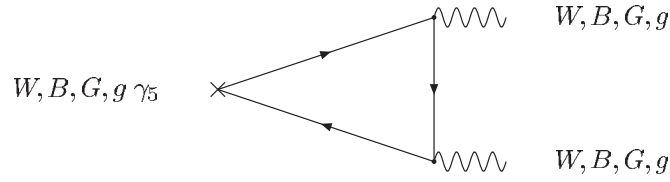


Fig. 8: Triangle diagrams that can give rise to anomalies. W, B, G are the gauge bosons associated to the $SU(2), U_Y(1), SU(3)$ gauge groups, respectively, and g is the graviton.

$$\underbrace{\sum_{i=\text{quarks}} Y_i^L - Y_i^R}_{GGB} = \underbrace{\sum_{i=\text{doublets}} Y_i^L}_{WWB} = \underbrace{\sum_i Y_i^L - Y_i^R}_{Bgg} = \underbrace{\sum_i (Y_i^L)^3 - (Y_i^R)^3}_{B^3} = 0, \quad (8)$$

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

2.2 Family structure

Concerning the family structure, we know, thanks to neutrinos, that there are exactly three families in the SM. An extra SM family with quarks and charged leptons so heavy that cannot be produced, would also have massless neutrinos that would contribute to the invisible Z^0 decay:

$$Z^0 \rightarrow \bar{\nu}_\alpha \nu_\alpha. \quad (9)$$

The invisible width of the Z^0 has been measured at LEP with an impressive precision, as shown in Fig. 9. This measurement excludes any number of standard families different from three [7]:

$$N_\nu = \frac{\Gamma_{\text{inv}}}{\Gamma_{\bar{\nu}\nu}} = 2.984 \pm 0.008. \quad (10)$$

3 Massive Neutrinos

Neutrinos are ubiquitous in our surroundings. If we open our hand, it will be crossed each second by about $\mathcal{O}(10^{12})$ neutrinos from the sun, about $\mathcal{O}(10)$ from the atmosphere, about $\mathcal{O}(10^9)$ from natural radioactivity in the Earth and even $\mathcal{O}(10^{12})$ relic neutrinos from the Big Bang. In 1987, the Kamiokande detector in Japan observed the neutrino burst from a SuperNova that exploded in the Large Magellanic Cloud, at a distance of 168 thousand light years from Earth. For a few seconds, the supernova flux was of the same order of magnitude as the flux of solar neutrinos!

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

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

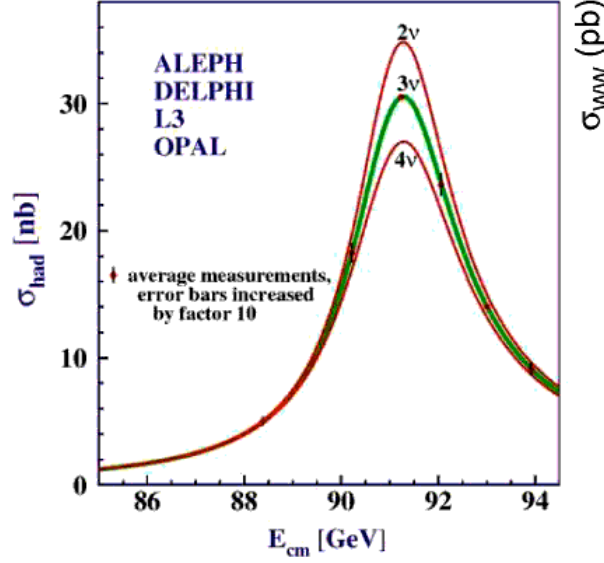


Fig. 9: Z^0 resonance from the LEP experiments. Data are compared to the case of $N_\nu = 2, 3$ and 4.

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

$$-\mathcal{L}_m^{Dirac} = m\bar{\psi}\psi = m(\overline{\psi_L + \psi_R})(\psi_L + \psi_R) = m(\overline{\psi_L}\psi_R + \overline{\psi_R}\psi_L). \quad (11)$$

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

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

$$-\mathcal{L}_m^{Majorana} = \frac{m}{2}\bar{\psi}^c\psi + \frac{m}{2}\bar{\psi}\psi^c = \frac{m}{2}\psi^T C\psi + \frac{m}{2}\bar{\psi}C\bar{\psi}^T, \quad (12)$$

where

$$\psi^c \equiv C\bar{\psi}^T = C\gamma_0\psi^*. \quad (13)$$

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

- 1) It can be constructed with a two-component spinor or Weyl fermion: if $\psi = P_L\psi \equiv (\psi_L, 0)$

$$\psi^T C\psi = \psi_L^T i\sigma_2\psi_L, \quad (14)$$

and it does not vanish in the absence of the right chiral component.

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

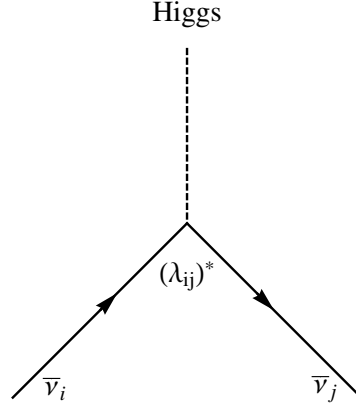
$$\psi \rightarrow e^{-\frac{i}{4}\omega_{\mu\nu}\sigma^{\mu\nu}}\psi \equiv S(\Lambda)\psi, \quad \psi^c \rightarrow S(\Lambda)\psi^c, \quad (15)$$

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

- 3) The equation of motion derived from eq. (12) for a free majorana fermion has plane wave solutions satisfying the relativistic relation for a massive fermion:

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

In the SM none of the mass terms of eqs. (11) and (12) are gauge invariant. Spontaneous symmetry breaking allows to generate the Dirac mass term from Yukawa couplings for all fermions in the SM, while the Majorana mass term can only be generated for neutrinos.


Fig. 10: Neutrino Yukawa coupling.

3.1 Massive Dirac neutrinos

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

$$-\mathcal{L}_m^{Dirac} = \bar{L} \lambda \tilde{\Phi} \nu_R + \text{h.c.} \quad (16)$$

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

$$-\mathcal{L}_m^{Dirac} \rightarrow -\bar{\nu}_L \lambda \frac{v}{\sqrt{2}} \nu_R + \text{h.c.} \quad (17)$$

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

$$m_\nu = \lambda \frac{v}{\sqrt{2}}. \quad (18)$$

There are two important consequences of Dirac neutrinos. First, there is a new hierarchy problem in the SM to be explained: why are neutrinos so much lighter than the remaining leptons, even those in the same family (see Fig. 11), if they get the mass in the same way? Secondly, an accidental global symmetry, lepton number L , that counts the number of leptons minus that of antilepton, remains exactly conserved at the classical level¹, just as baryon number, B , is.

3.2 Massive Majorana neutrinos

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

$$-\mathcal{L}_m^{Majorana} = \bar{L}\tilde{\phi} \alpha C \tilde{\phi}^T \bar{L}^T + \text{h.c.}, \quad (19)$$

is gauge invariant. This term, first written down by Weinberg [10], gives rise to a Majorana mass term for neutrinos upon spontaneous symmetry breaking:

$$-\mathcal{L}_m^{Majorana} \rightarrow \bar{\nu}_L \alpha \frac{v^2}{2} C \bar{\nu}_L^T + \text{h.c.}, \quad (20)$$

¹As usual $B + L$ is broken by the anomaly and only $B - L$ remains exact at all orders.

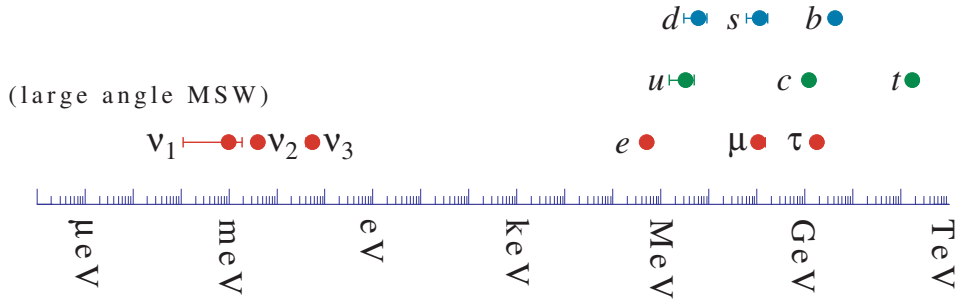


Fig. 11: Fermion spectrum in the Standard Model.

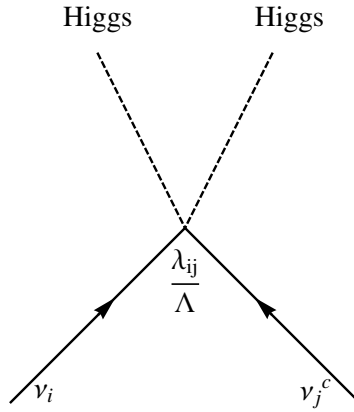


Fig. 12: Weinberg operator.

The neutrino mass matrix in this case is given by:

$$m_\nu = \alpha v^2. \tag{21}$$

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

$$\alpha = \frac{\lambda}{\Lambda}, \tag{22}$$

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

The consequences of the SM neutrinos being massive Majorana particles are profound.

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

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

Finally, it is interesting to note that the anomaly cancellation conditions fix all the hypercharges in this case (i.e., there is only one possible choice for the hypercharges that satisfies Eqs. (8)), which implies that electromagnetic charge quantization is the only possibility in a field theory with the same matter content as the SM.

3.3 Neutrino masses and physics beyond-the-standard-model

Any new physics beyond the standard model (BSM) characterized by a high scale, Λ , will induce effects at low energies $E \ll \Lambda$ that can be described by an effective field theory [11, 12]:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{\alpha_i}{\Lambda} O_i^{d=5} + \sum_i \frac{\beta_i}{\Lambda^2} O_i^{d=6} + \dots \quad (23)$$

It is the most general Lagrangian which includes the SM and an infinite tower of operators constructed out of the SM fields respecting Lorentz and gauge symmetries. In principle such a theory depends on infinite new couplings, one per new independent operator, and it is therefore not predictive. However, if the energy we are interested in effects at a given finite order, n , in $(\frac{E}{\Lambda})^n$, we can truncate the series to include only operators of dimension $d \leq n + 4$. The operators of lowest dimension are the most relevant at low energies.

It turns out that there is only one such operator of the lowest possible dimension, $d = 5$, which is precisely the Weinberg operator of eq. (19). In this perspective, it is natural to expect that the first indication of BSM physics is precisely Majorana neutrino masses, and while many types of BSM theories can give rise to neutrino masses, generically they will induce other new physics effects represented by the operators of $d = 6$ or higher.

4 Neutrino masses and lepton mixing

Neutrino masses, whether Dirac or Majorana, imply lepton mixing [13, 14]. The Yukawa couplings in eq. (16) is a generic complex matrix in flavour space, while that in eq. (19) is a generic complex symmetric matrix, and therefore the same holds for the corresponding mass matrices:

$$-\mathcal{L}_m^{\text{Dirac}} = \overline{\nu}_L^i (M_\nu)_{ij} \nu_R^j + \overline{l}_L^i (M_l)_{ij} l_R^j + \text{h.c.} \quad (24)$$

$$-\mathcal{L}_m^{\text{Majorana}} = \frac{1}{2} \overline{\nu}_L^i (M_\nu)_{ij} \nu_L^{cj} + \overline{l}_L^i (M_l)_{ij} l_R^j + \text{h.c.} \quad (25)$$

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

$$M_\nu = U_\nu^\dagger \text{Diag}(m_1, m_2, m_3) V_\nu, \quad M_l = U_l^\dagger \text{Diag}(m_e, m_\mu, m_\tau) V_l, \quad (26)$$

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

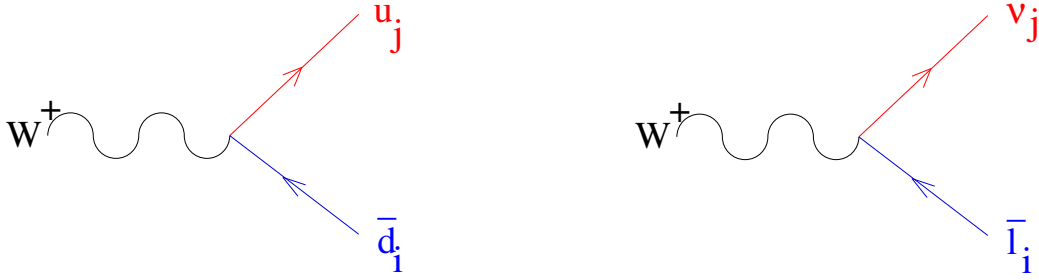
$$M_\nu = U_\nu^\dagger \text{Diag}(m_1, m_2, m_3) U_\nu^*. \quad (27)$$

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

$$\nu'_R = V_\nu \nu_R, \quad \nu'_L = U_\nu \nu_L, \quad l'_R = V_l l_R, \quad l'_L = U_l l_L. \quad (28)$$

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

$$\mathcal{L}_{CC}^{\text{lepton}} = -\frac{g}{\sqrt{2}} \overline{l}'_i \gamma_\mu P_L W_\mu^+ \underbrace{(U_l^\dagger U_\nu)_{ij}}_{U_{\text{PMNS}}} \nu'_j + \text{h.c.} \quad (29)$$


Fig. 13: Quark and lepton mixing.

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

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

	Yukawas	Field redefinitions	$No. m$	$No. \theta$	$No. \phi$
Dirac	λ_l, λ_ν $4n^2$	$\frac{U(n)^3/U(1)}{2}, \frac{3(n^2+n)-1}{2}$	$2n$	$\frac{n^2-n}{2}$	$\frac{(n-2)(n-1)}{2}$
Majorana	$\lambda_l, \alpha_\nu^T = \alpha_\nu$ $3n^2 + n$	$U(n)^2$ $n^2 - n, n^2 + n$	$2n$	$\frac{n^2-n}{2}$	$\frac{n^2-n}{2}$

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

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

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

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

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

$$\begin{aligned}
 U_{\text{PMNS}} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\
 \tilde{U}_{\text{PMNS}} &= U_{\text{PMNS}}(\theta_{12}, \theta_{13}, \theta_{23}, \delta) \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\alpha_1} & 0 \\ 0 & 0 & e^{i\alpha_2} \end{pmatrix}, \tag{30}
 \end{aligned}$$

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

5 Majorana versus Dirac

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

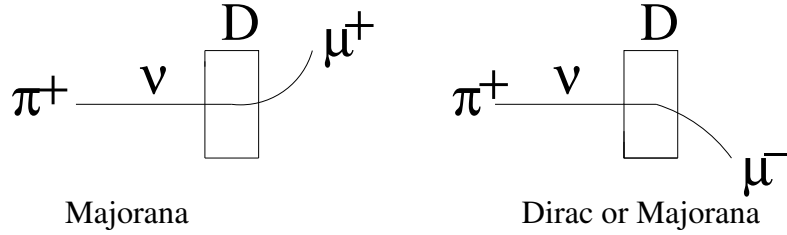


Fig. 14: A neutrino beam from π^+ decay (ν_μ) could interact in the magnetized detector producing a μ^+ only if neutrinos are Majorana.

The best hope of observing a rare process of this type seems to be the search for neutrinoless double-beta decay ($2\beta 0\nu$), the right diagram of Fig. 15. The background to this process is the standard double-beta decay depicted on the left of Fig. 15, which has been observed to take place for various isotopes with a lifetime of $T_{2\beta 2\nu} > 10^{19}-10^{21}$ years.

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

$$T_{2\beta 0\nu}^{-1} \simeq \underbrace{G^{0\nu}}_{\text{Phase}} \underbrace{|M^{0\nu}|^2}_{\text{Nuclear M.E.}} \underbrace{\left| \sum_i (\tilde{U}_{\text{PMNS}}^{ei})^2 m_i \right|^2}_{|m_{ee}|^2}. \tag{31}$$

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



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

Experiment	Nucleus	$ m_{ee} $
EXO-200	^{136}Xe	$< 0.19\text{--}0.45$ eV
NEMO-3	^{100}Mo	$< 0.33\text{--}0.87$ eV
GERDA	^{76}Ge	$< 0.2\text{--}0.4$ eV
KamLAND-Zen	^{136}Xe	$< 0.12\text{--}0.25$ eV
CUORICINO	^{130}Te	$< 0.2\text{--}0.7$ eV

Table 3: Present bounds at 90%CL from some recent neutrinoless double-beta-decay experiments [7].

6 Neutrino Oscillations

The most spectacular implication of neutrino masses and mixings is the macroscopic quantum phenomenon of neutrino oscillations, first introduced by B. Pontecorvo [15]. The Nobel prize in 2015 has been awarded to T. Kajita (from the SuperKakiokande collaboration) and A. B. McDonald (from the SNO collaboration) for the *discovery of neutrino oscillations, which shows that neutrinos have a mass*.

We have seen that the neutrino flavour fields (ν_e, ν_μ, ν_τ) that couple via CC to the leptons (e, μ, τ) are unitary combinations of the mass eigenstates fields (ν_1, ν_2, ν_3):

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U_{\text{PMNS}}(\theta_{12}, \theta_{13}, \theta_{23}, \text{phases}) \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \quad (32)$$

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

Two ingredients are mandatory for this phenomenon to take place:

- neutrinos must keep quantum coherence in propagation over macroscopic distances, which is only possible because they are so weakly interacting



Fig. 16: Neutrino oscillations.

- there is sufficient uncertainty in momentum at production and detection so that a coherent flavour state can be produced²

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

$$P(\nu_\alpha \rightarrow \nu_\beta) = \sum_{i,j} U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^* e^{-i \frac{\Delta m_{ji}^2 L}{2|\mathbf{p}|}}, \quad (33)$$

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

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

6.1 Plane wave derivation

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

$$|\nu_\alpha(t_0)\rangle = \sum_i U_{\alpha i}^* |\nu_i(\mathbf{p})\rangle. \quad (34)$$

The mass eigenstates are eigenstates of the free Hamiltonian:

$$\hat{H} |\nu_i(\mathbf{p})\rangle = E_i(\mathbf{p}) |\nu_i(\mathbf{p})\rangle, \quad E_i(\mathbf{p})^2 = \mathbf{p}^2 + m_i^2. \quad (35)$$

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

$$|\nu_\alpha(t)\rangle = e^{-i\hat{H}(t-t_0)} |\nu_\alpha(t_0)\rangle = \sum_i U_{\alpha i}^* e^{-iE_i(\mathbf{p})(t-t_0)} |\nu_i(\mathbf{p})\rangle. \quad (36)$$

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

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

$$P(\nu_\alpha \rightarrow \nu_\beta)(t) = |\langle \nu_\beta | \nu_\alpha(t) \rangle|^2 = \left| \sum_i U_{\beta i} U_{\alpha i}^* e^{-iE_i(\mathbf{p})(t-t_0)} \right|^2, \quad (37)$$

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

Since the neutrinos are ultrarelativistic, we can approximate

$$E_i(\mathbf{p}) - E_j(\mathbf{p}) \simeq \frac{1}{2} \frac{m_i^2 - m_j^2}{|\mathbf{p}|} + \mathcal{O}(m^4), \quad (38)$$

and $L \simeq (t - t_0)$, so that the master formula in eq. (33) is recovered.

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

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

6.2 Wave packet derivation

Many authors have derived the master formula treating neutrinos involved as wave packets. For some recent references see [16, 17].

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

$$|\nu_\alpha(t)\rangle = \sum_i U_{\alpha i}^* \int_{\mathbf{p}} f_i^S(\mathbf{p}) e^{-iE_i(\mathbf{p})t} |\nu_i(\mathbf{p})\rangle. \quad (39)$$

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

$$f_i^S(\mathbf{p}) \propto e^{-(\mathbf{p}-\mathbf{Q}_i)^2/2\sigma_S^2}. \quad (40)$$

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

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

$$|\nu_\beta(t)\rangle = \sum_j U_{\beta j}^* \int_{\mathbf{p}} f_j^D(\mathbf{p}) e^{-iE_j(\mathbf{p})(t-T)} e^{-i\mathbf{p}\mathbf{L}} |\nu_j(\mathbf{p})\rangle, \quad (41)$$

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

$$f_j^D(\mathbf{p}) \propto e^{-(\mathbf{p}-\mathbf{Q}'_j)^2/2\sigma_D^2}. \quad (42)$$

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

$$\mathcal{A}(\nu_\alpha \rightarrow \nu_\beta) = \langle \nu_\beta(t) | \nu_\alpha(t) \rangle = \sum_i U_{\alpha i}^* U_{\beta i} \int_{\mathbf{p}} e^{-iE_i(\mathbf{p})T} e^{i\mathbf{p}\mathbf{L}} f_i^S(\mathbf{p}) f_i^{D*}(\mathbf{p}) \quad (43)$$

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

$$f_i^{D*}(\mathbf{p})f_i^S(\mathbf{p}) \propto f_i^{ov}(\mathbf{p})e^{-(\mathbf{Q}_i-\mathbf{Q}'_i)^2/4(\sigma_S^2+\sigma_D^2)}, \quad (44)$$

where the overlap wave packet

$$f_i^{ov}(\mathbf{p}) \equiv e^{-(\mathbf{p}-\bar{\mathbf{Q}}_i)^2/2\sigma_{ov}^2}, \quad \bar{\mathbf{Q}}_i \equiv \left(\frac{\mathbf{Q}_i}{\sigma_S^2} + \frac{\mathbf{Q}'_i}{\sigma_D^2} \right) \sigma_{ov}^2, \quad \sigma_{ov}^2 \equiv \frac{1}{1/\sigma_S^2 + 1/\sigma_D^2}. \quad (45)$$

The momentum integral in eq. (43) can be done analytically if we approximate

$$E_i(\mathbf{p}) \simeq E_i(\bar{\mathbf{Q}}_i) + \sum_k \left. \frac{\partial E_i}{\partial p_k} \right|_{\bar{\mathbf{Q}}_i} (p_k - (\bar{\mathbf{Q}}_i)_k) + \dots = E_i(\bar{\mathbf{Q}}_i) + \mathbf{v}_i(\mathbf{p} - \bar{\mathbf{Q}}_i) + \dots, \quad (46)$$

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

The amplitude obtained is

$$\mathcal{A}(\nu_\alpha \rightarrow \nu_\beta) \propto \sum_i U_{\alpha i}^* U_{\beta i} e^{-iE_i(\bar{\mathbf{Q}}_i)T} e^{i\bar{\mathbf{Q}}_i \mathbf{L}} e^{-(\mathbf{Q}_i-\mathbf{Q}'_i)^2/4(\sigma_S^2+\sigma_D^2)} e^{-(\mathbf{L}-\mathbf{v}_i T)^2 \sigma_{ov}^2/2}. \quad (47)$$

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

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

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta) &\propto \int_{-\infty}^{\infty} dT |\mathcal{A}(\nu_\alpha \rightarrow \nu_\beta)|^2 \\ &\propto \sum_{i,j} U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^* e^{-i \frac{\Delta m_{ji}^2 L}{2|\mathbf{p}|}} \underbrace{e^{-\left(\frac{L}{L_{coh}(i,j)}\right)^2}}_{\text{coherence}} \underbrace{e^{-\left(\frac{E_i(\bar{\mathbf{Q}}_i) - E_j(\bar{\mathbf{Q}}_j)}{2\sigma_{ov}}\right)^2}}_{\text{momentum uncertainty}} \end{aligned} \quad (48)$$

where the coherence length

$$L_{coh}(i,j) \simeq \sigma_{ov} \frac{|\mathbf{v}_i - \mathbf{v}_j|}{\sqrt{|\mathbf{v}_i^2 + \mathbf{v}_j^2}}, \quad (49)$$

represents the distance travelled by the two wave packets, moving at slightly different group velocities v_i and v_j , such that the center of the two wave packets have separated spacially a distance of the order of the spatial uncertainty σ_{ov}^{-1} . For $L \geq L_{coh}(i,j)$ the coherence between the wave packets i, j is lost and the corresponding terms in the oscillation probability exponentially suppressed. The last exponential factor in eq. (48) leads to a suppression of the oscillation probability when the difference in average energies of the two wave packets i, j is larger than the momentum uncertainty of the overlap wave packet, σ_{ov} . Note that σ_{ov} is dominated by the smallest of the production and detection uncertainties, and therefore both should be large enough to ensure that the wave packets of the different mass eigenstates remain coherent. To the extent that $L \ll L_{coh}$ and $|E_i - E_j| \ll \text{Min}(\sigma_S, \sigma_D)$, the probability reduces to the master formula, with one caveat: we have lost the normalization along the way. This is usually unavoidable in the wave packet derivation. The right normalization can be imposed only a posteriori, for example, from unitarity, $\sum_\beta P(\nu_\alpha \rightarrow \nu_\beta) = 1$.

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

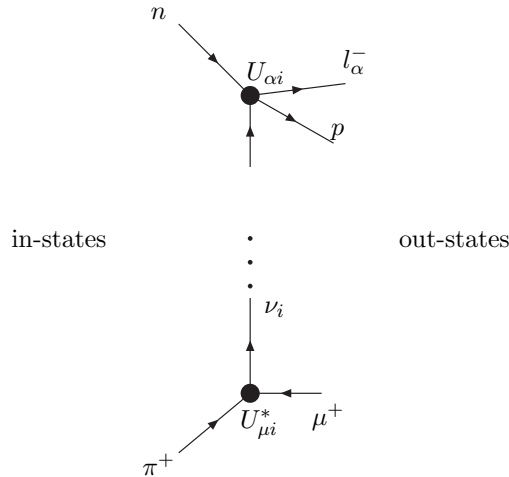


Fig. 17: Neutrino oscillations in QFT.

6.3 QFT derivation

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

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

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

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

$$\pi n \rightarrow p \mu l_{\beta}. \quad (50)$$

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

By explicit calculation [18], it is possible to show that such factorization does indeed take place as long as kinematical effects of neutrino masses can be neglected. The oscillation probability defined as the ratio of the probability for the whole process and the product of the neutrino flux from pion decay and the neutrino scattering cross-section is properly normalized.

6.4 Neutrino oscillations in vacuum

Let us analyse more closely the master formula eq. (33). The probability is a superposition of oscillatory functions of the baseline with wavelengths that depend on the neutrino mass differences $\Delta m_{ij}^2 = m_j^2 - m_i^2$, and amplitudes that depend on different combinations of the mixing matrix elements. Defining $W_{\alpha\beta}^{ij} \equiv [U_{\alpha i} U_{\beta i}^* U_{\alpha j}^* U_{\beta j}]$ and using the unitarity of the mixing matrix, we can rewrite the probability in the more familiar form:

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta) &= \delta_{\alpha\beta} - 4 \sum_{j>i} \text{Re}[W_{\alpha\beta}^{ij}] \sin^2\left(\frac{\Delta m_{ij}^2 L}{4E_\nu}\right) \\ &\mp 2 \sum_{j>i} \text{Im}[W_{\alpha\beta}^{ij}] \sin\left(\frac{\Delta m_{ij}^2 L}{2E_\nu}\right), \end{aligned} \quad (51)$$

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

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

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

$$U_{\text{PMNS}} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}, \quad (52)$$

and there is only one mass square difference Δm^2 . The oscillation probability of Eq. (51) simplifies to the well-known expression where we have introduced convenient physical units:

$$\begin{aligned} P(\nu_\alpha \rightarrow \nu_\beta) &= \sin^2 2\theta \sin^2\left(1.27 \frac{\Delta m^2 (\text{eV}^2) L (\text{km})}{E_\nu (\text{GeV})}\right), \quad \alpha \neq \beta. \\ P(\nu_\alpha \rightarrow \nu_\alpha) &= 1 - P(\nu_\alpha \rightarrow \nu_\beta). \end{aligned} \quad (53)$$

The probability is the same for neutrinos and antineutrinos, because there cannot be CP violation when there are only two families. Indeed CPT implies that the disappearance probabilities are the same for neutrinos and antineutrinos, and therefore according to eq. (53) the same must hold for the appearance probability. The latter is a sinusoidal function of the distance between source and detector, with a period determined by the oscillation length:

$$L_{\text{osc}} (\text{km}) = \pi \frac{E_\nu (\text{GeV})}{1.27 \Delta m^2 (\text{eV}^2)}, \quad (54)$$

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

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

$$E_{\text{max}} (\text{GeV}) = 1.27 \frac{\Delta m^2 (\text{eV}^2) L (\text{km})}{\pi/2}. \quad (55)$$

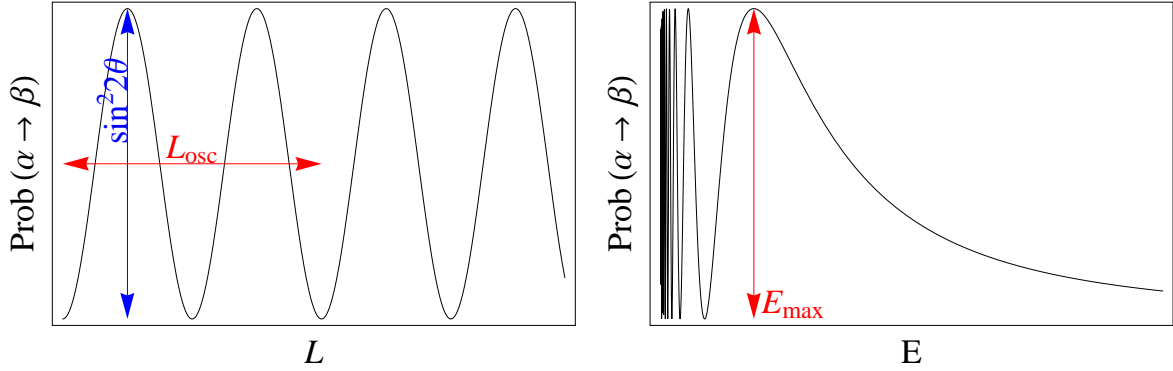


Fig. 18: Left: two-family appearance oscillation probability as a function of the baseline of L at fixed neutrino energy. Right: same probability shown as a function of the neutrino energy for fixed baseline.

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

$$\langle P(\nu_\alpha \rightarrow \nu_\beta) \rangle \simeq \frac{1}{2} \sin^2 2\theta. \quad (56)$$

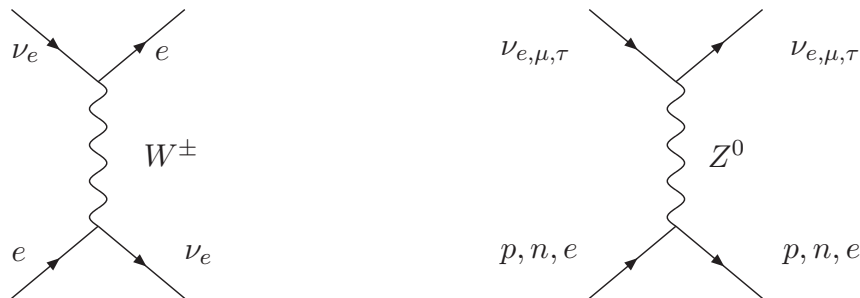
It is interesting, and reassuring, to note that this averaged oscillation regime gives the same result as the flavour transition probability in the case of incoherent propagation ($L \gg L_{\text{coh}}$):

$$P(\nu_\alpha \rightarrow \nu_\beta) = \sum_i |U_{\alpha i} U_{\beta i}|^2 = 2 \cos^2 \theta \sin^2 \theta = \frac{1}{2} \sin^2 2\theta. \quad (57)$$

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

6.5 Neutrino propagation in matter

When neutrinos propagate in matter (Earth, sun, etc.), their propagation is modified owing to coherent forward scattering on electrons and nucleons [19]:



The effective Hamiltonian density resulting from the charged current interaction is

$$\mathcal{H}_{CC} = 2\sqrt{2}G_F [\bar{e}\gamma_\mu P_L \nu_e][\bar{\nu}_e\gamma^\mu P_L e] = 2\sqrt{2}G_F [\bar{e}\gamma_\mu P_L e][\bar{\nu}_e\gamma^\mu P_L \nu_e]. \quad (58)$$

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

$$\langle \bar{e}\gamma_\mu P_L e \rangle_{\text{unpol. medium}} = \delta_{\mu 0} \frac{N_e}{2}. \quad (59)$$

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

$$\langle \mathcal{H}_{CC} + \mathcal{H}_{NC} \rangle_{\text{medium}} = \bar{\nu} V_m \gamma^0 (1 - \gamma_5) \nu \quad (60)$$

$$V_m = \begin{pmatrix} \frac{G_F}{\sqrt{2}} (N_e - \frac{N_n}{2}) & 0 & 0 \\ 0 & \frac{G_F}{\sqrt{2}} (-\frac{N_n}{2}) & 0 \\ 0 & 0 & \frac{G_F}{\sqrt{2}} (-\frac{N_n}{2}) \end{pmatrix}, \quad (61)$$

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

The plane wave solutions to the modified Dirac equation satisfy a different dispersion relation and as a result, the phases of neutrino oscillation phenomena change. The new dispersion relation becomes

$$E - V_m - M_\nu = (\pm|\mathbf{p}| - V_m) \frac{1}{E + M_\nu - V_m} (\pm|\mathbf{p}| - V_m) \quad h = \pm, \quad (62)$$

where $h = \pm$ indicate the two helicity states and we have neglected effects of $\mathcal{O}(VM_\nu)$. This is a reasonable approximation since $m_\nu \gg V_m$. For the positive energy states we then have

$$E > 0 \quad E^2 = |\mathbf{p}|^2 + M_\nu^2 + 4EV_m \quad h = - \quad E^2 = |\mathbf{p}|^2 + M_\nu^2, \quad h = +, \quad (63)$$

while for the negative energy ones $V_m \rightarrow -V_m$ and $h \rightarrow -h$.

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

$$\tilde{M}_\nu^2 = M_\nu^2 \pm 4EV_m. \quad (64)$$

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

$$\begin{pmatrix} \tilde{m}_1^2 & 0 & 0 \\ 0 & \tilde{m}_2^2 & 0 \\ 0 & 0 & \tilde{m}_3^2 \end{pmatrix} = \tilde{V}_{\text{MNS}}^\dagger \left(M_\nu^2 \pm 4E \begin{pmatrix} V_e & 0 & 0 \\ 0 & V_\mu & 0 \\ 0 & 0 & V_\tau \end{pmatrix} \right) \tilde{V}_{\text{MNS}}. \quad (65)$$

The effective mixing angles and masses depend on the energy.

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

6.6 Neutrino oscillations in constant matter

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

$$\Delta \tilde{m}^2 = \sqrt{(\Delta m^2 \cos 2\theta \mp 2\sqrt{2}E G_F N_e)^2 + (\Delta m^2 \sin 2\theta)^2}, \quad (66)$$

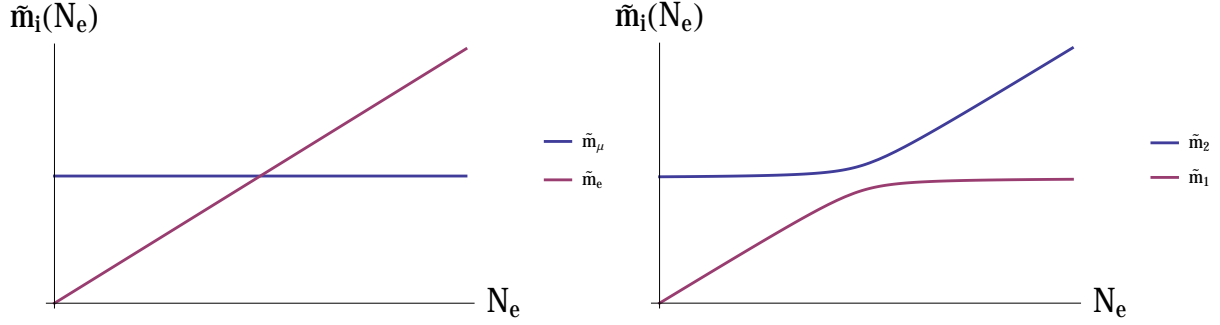


Fig. 19: Mass eigenstates as a function of the electron number density at fixed neutrino energy for $\theta = 0$ (left) and $\theta \neq 0$ (right).

$$\sin^2 2\tilde{\theta} = \frac{(\Delta m^2 \sin 2\theta)^2}{(\Delta \tilde{m}^2)^2} \quad (67)$$

where the sign \mp corresponds to neutrinos/antineutrinos. The corresponding oscillation amplitude has a resonance [20], when the neutrino energy satisfies

$$\sqrt{2} G_F N_e \mp \frac{\Delta m^2}{2E} \cos 2\theta = 0 \quad \Rightarrow \quad \sin^2 2\tilde{\theta} = 1, \quad \Delta \tilde{m}^2 = \Delta m^2 \sin 2\theta. \quad (68)$$

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

We also note that

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

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

6.7 Neutrino oscillations in variable matter

In the sun the density of electrons is not constant. However, if the variation is sufficiently slow, the eigenstates will change slowly with the density, and we can assume that the neutrino produced in an eigenstate in the center of the sun, remains in the same eigenstate along the trajectory. This is the so-called *adiabatic approximation*.

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

$$|\tilde{\nu}_1\rangle = |\nu_e\rangle \cos \tilde{\theta} - |\nu_\mu\rangle \sin \tilde{\theta}, \quad (69)$$

$$|\tilde{\nu}_2\rangle = |\nu_e\rangle \sin \tilde{\theta} + |\nu_\mu\rangle \cos \tilde{\theta}. \quad (70)$$

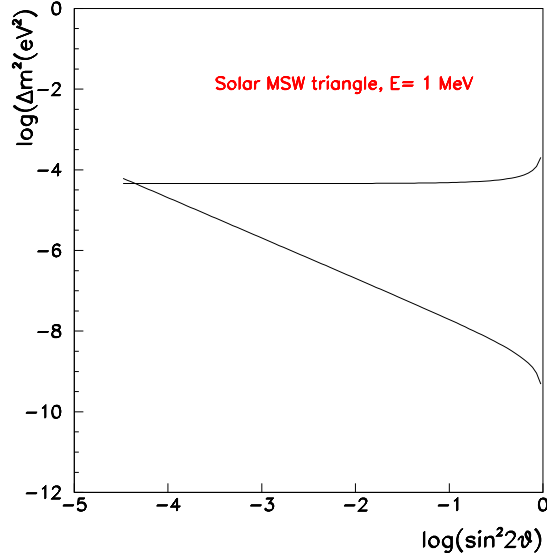


Fig. 20: MSW triangle: in the region between the two lines the resonance and adiabaticity conditions are both satisfied for neutrinos of energy 1 MeV.

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

$$2\sqrt{2}G_F N_e(0) \gg \Delta m^2 \cos 2\theta. \quad (71)$$

Then the diagonalization of the mass matrix at this point gives

$$\tilde{\theta} \simeq \frac{\pi}{2} \Rightarrow |\nu_e\rangle \simeq |\tilde{\nu}_2\rangle, \quad (72)$$

in such a way that an electron neutrino is mostly the second mass eigenstate. When neutrinos exit the sun, at $x = R_\odot$, the matter density falls to zero, $N_e(R_\odot) = 0$, and the local effective mixing angle is the one in vacuum, $\tilde{\theta} = \theta$. If θ is small, the eigenstate $\tilde{\nu}_2$ is mostly ν_μ according to Eq. (70).

Therefore an electron neutrino produced at $x = 0$ is mostly the eigenstate $\tilde{\nu}_2$, but this eigenstate outside the sun is mostly ν_μ . There is maximal $\nu_e \rightarrow \nu_\mu$ conversion if the adiabatic approximation is a good one. This is the famous MSW effect [19,20]. The conditions for this to happen are:

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

$$N_e(0) > \frac{\Delta m^2 \cos 2\theta}{2\sqrt{2}EG_F}. \quad (73)$$

- *Adiabaticity:* the splitting of the levels is large compared to energy injected in the system by the variation of $N_e(r)$. A measurement of this is given by γ which should be much larger than one:

$$\gamma = \frac{\sin^2 2\theta}{\cos 2\theta} \frac{\Delta m^2}{2E} \frac{1}{|\nabla \log N_e(r)|} > \gamma_{\min} > 1, \quad (74)$$

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

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

$$\log(\Delta m^2) < \log\left(\frac{2\sqrt{2}G_F N_e(0)E}{\cos 2\theta}\right) \quad (75)$$

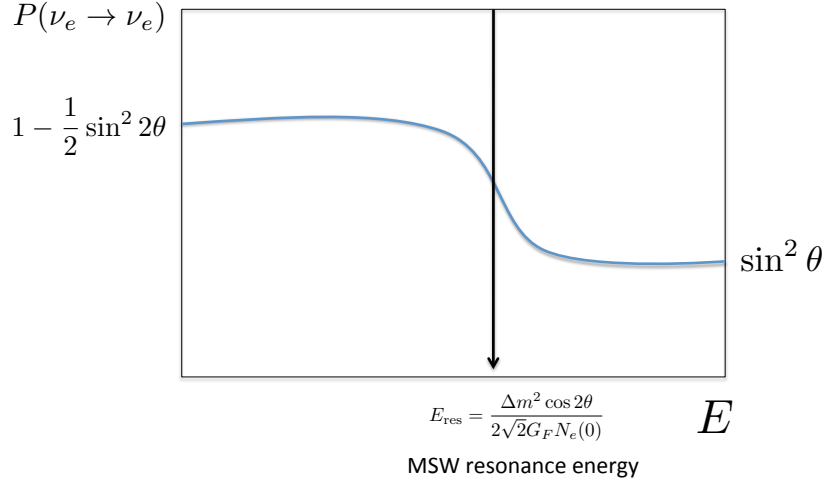


Fig. 21: Schematic survival probability of solar neutrinos as a function of the energy.

$$\log(\Delta m^2) > \log\left(\gamma_{\min} 2E \nabla \log N_e \frac{\cos 2\theta}{\sin^2 2\theta}\right). \quad (76)$$

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

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

$$P(\nu_e \rightarrow \nu_e) = \sum_i |\langle \nu_e | \tilde{\nu}_i(R_\odot) \rangle|^2 |\langle \tilde{\nu}_i(0) | \nu_e \rangle|^2, \quad (77)$$

where $\tilde{\nu}_i(r)$ is the i -th mass eigenstate for the electron number density, $N_e(r)$, at a distance r from the center of the sun. If the mass eigenstates contribute incoherently, how can we measure the neutrino mass splitting? The answer is that the resonance condition of eq. (73) depends on the neutrino energy. If we define

$$E_{\text{res}} \equiv \frac{\Delta m^2 \cos 2\theta}{2\sqrt{2}G_F N_e(0)}, \quad (78)$$

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

$$\begin{aligned} P(\nu_e \rightarrow \nu_e) &\simeq 1 - \frac{1}{2} \sin^2 2\theta, & E \ll E_{\text{res}} \\ P(\nu_e \rightarrow \nu_e) &\simeq \sin^2 \theta, & E \gg E_{\text{res}} \end{aligned} \quad (79)$$

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

7 Evidence for neutrino oscillations

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

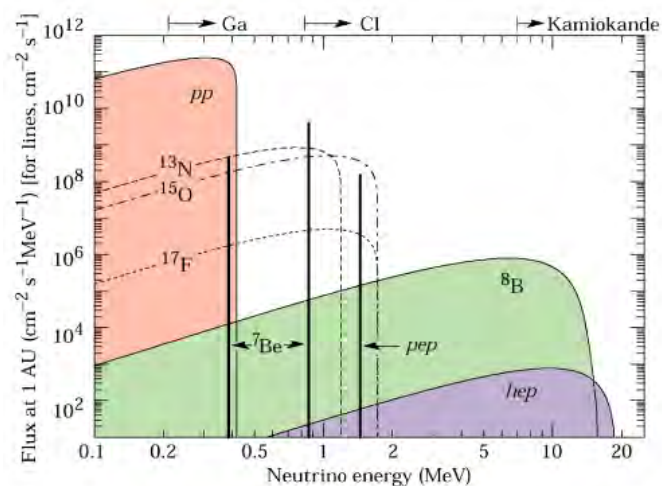


Fig. 22: Spectrum of solar neutrinos [22]. The arrows indicate the threshold of the different detection techniques.

7.1 Solar neutrinos

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



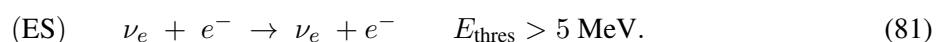
The theory of stellar nucleosynthesis was established at the end of the 30's by H. Bethe [21]. The spectrum of the solar ν_e , for massless neutrinos, is shown in Fig. 22. The prediction of this flux, obtained by J. Bahcall and collaborators [22], is the result of a detailed simulation of the solar interior and has been improved over many years. It is the so-called standard solar model (SSM).

Neutrinos coming from the sun have been detected with several experimental techniques that have a different neutrino energy threshold as indicated in Fig. 22. On the one hand, the radiochemical techniques, used in the experiments Homestake (chlorine, ${}^{37}\text{Cl}$) [23], Gallex/GNO [24] and Sage [25] (using gallium, ${}^{71}\text{Ga}$, and germanium, ${}^{71}\text{Ge}$, respectively), can count the total number of neutrinos with a rather low threshold ($E_\nu > 0.81$ MeV in Homestake and $E_\nu > 0.23$ MeV in Gallex and Sage), but they cannot get any information on the directionality, the energy of the neutrinos, nor the time of the event.

On the other hand, Kamiokande [26] pioneered a new technique to observe solar neutrinos using water Cherenkov detectors that can measure the recoil electron in elastic neutrino scattering on electrons: $\nu_e + e^- \rightarrow \nu_e + e^-$. This is a real-time experiment that provides information on the directionality and the energy of the neutrinos. The threshold on the other hand is much higher, ~ 5 MeV. All these experiments have consistently observed a number of solar neutrinos between 1/3 and 1/2 of the number expected in the SSM and for a long time this was referred to as the *solar neutrino problem or deficit*.

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

1998: The experiment SuperKamiokande [27] measured the solar neutrino deficit with unprecedented precision, using the elastic reaction (ES):



The measurement of the direction of the events demonstrated that the neutrinos measured definitely come from the sun: the left plot of Fig. 23 shows the distribution of the events as a function of the zenith angle of the sun. A seasonal variation of the flux is expected since the distance between the Earth and the sun

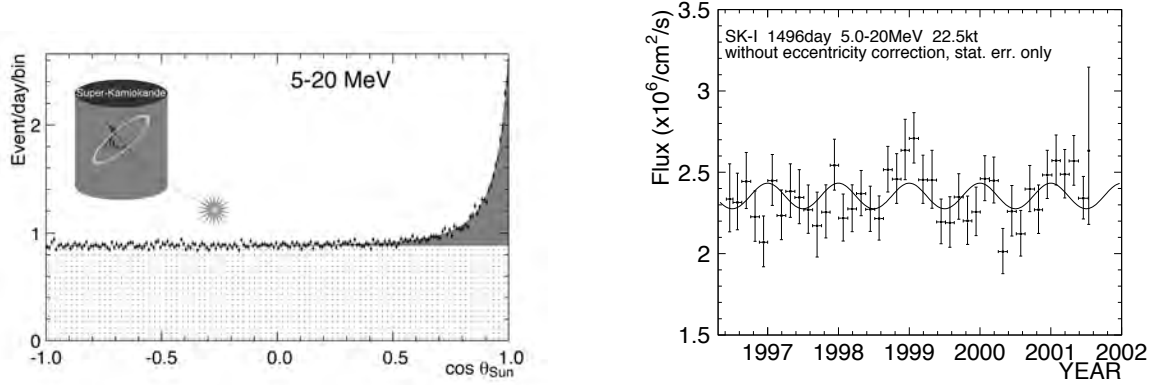


Fig. 23: Left: distribution of solar neutrino events as a function of the zenith angle of the sun. Right: seasonal variation of the solar neutrino flux in SuperKamiokande (from Ref. ([28])).

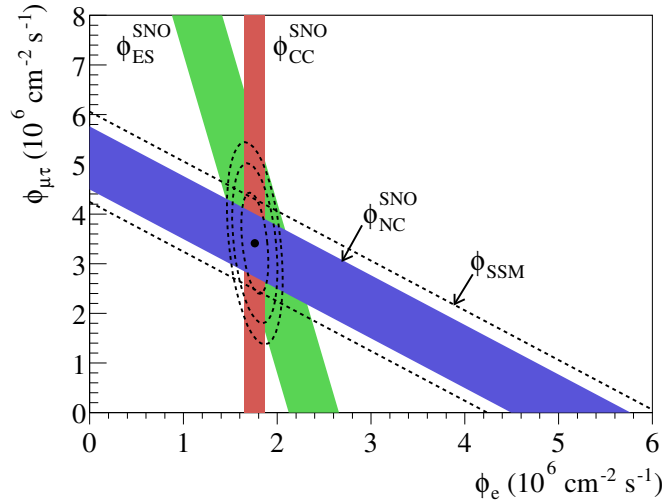
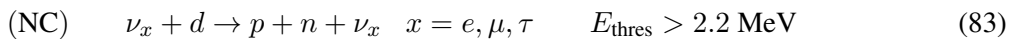


Fig. 24: Flux of ν_μ and ν_τ versus the flux of ν_e in the solar neutrino flux as measured from the three reactions observable in the SNO experiment. The dashed band shows the prediction of the SSM, which agrees perfectly with the flux measured with the NC reaction (from Ref. [30]).

varies seasonally. The right plot of Fig. 23 shows that the measured variation is in perfect agreement with that expectation.

2001: The SNO experiment [29, 30] measured the flux of solar neutrinos using also the two reactions:



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

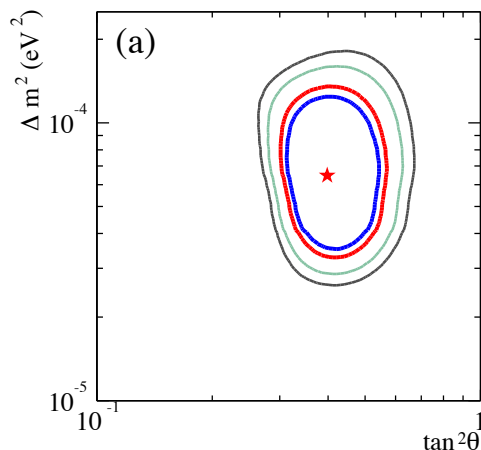


Fig. 25: Analysis of all solar data at SNO in terms of neutrino oscillations (from Ref. [29]).

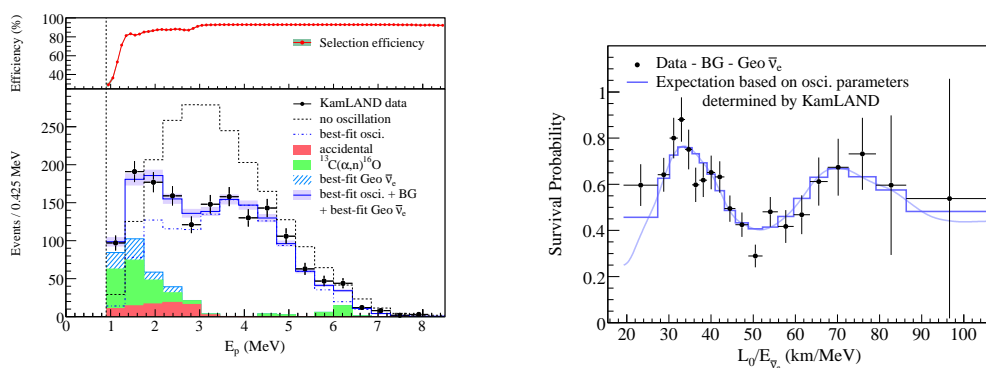


Fig. 26: Spectral distribution of the $\bar{\nu}_e$ events in KamLAND (left) and E_ν/L dependence (right). The data are compared to the expectation in the absence of oscillations and to the best fit oscillation hypothesis (from Ref. [32]).

All solar neutrino data can be interpreted in terms of neutrino masses and mixings. The analysis in terms of two neutrino families is shown in the left plot of Fig. 25. The solar ν_e deficit can be explained for a $\Delta m_{\text{solar}}^2 \simeq 7\text{--}8 \times 10^{-5} \text{eV}$ and a relatively large mixing angle. The fortunate circumstance that

$$\Delta m_{\text{solar}}^2 \sim \langle E_\nu(1 \text{ MeV}) \rangle / L(100 \text{ km}) \quad (84)$$

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

2002: The solar oscillation is confirmed with reactor neutrinos in the KamLAND experiment [31]. This is 1kton of liquid scintillator which measures the flux of reactor neutrinos produced in a cluster of nuclear plants around the Kamioka mine in Japan. The average distance is $\langle L \rangle = 175$ km. Neutrinos are detected via inverse β -decay which has a threshold energy of about 2.6 MeV:

$$\bar{\nu}_e + p \rightarrow e^+ + n \quad E_{\text{th}} > 2.6 \text{ MeV} . \quad (85)$$

Figure 26 shows the KamLAND results [32] on the antineutrino spectrum, as well as the survival probability as a function of the ratio E_ν/L . The low-energy contribution of geoneutrinos is clearly visible. This measurement could have important implications in geophysics.

Concerning the sensitivity to the oscillation parameters, Fig. 27 shows the present determination of the solar oscillation parameters from KamLAND and other solar experiments. The precision in the

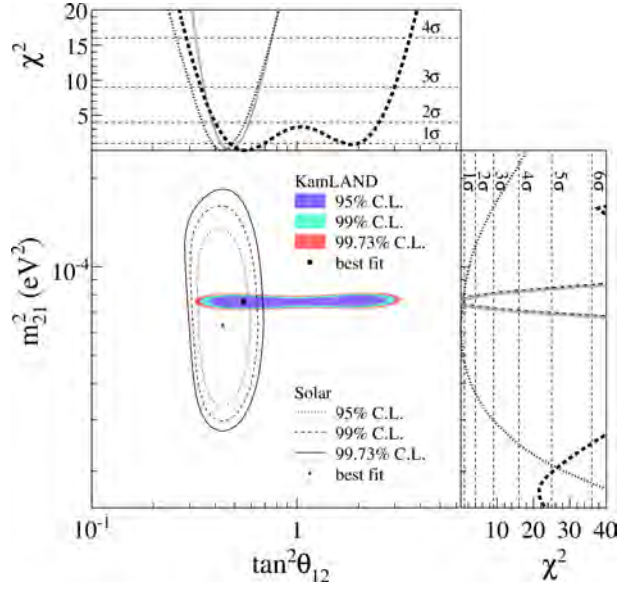


Fig. 27: Analysis of all solar and KamLAND data in terms of oscillations (from Ref. [32]).

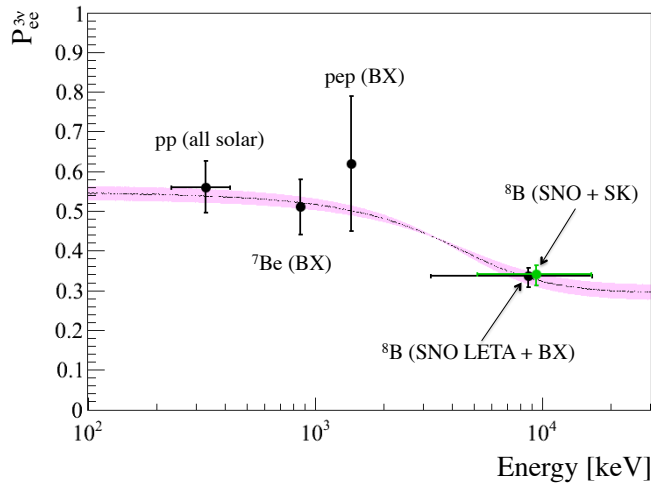


Fig. 28: Comparison of solar neutrino fluxes measured by the different solar neutrino experiments (from Ref. [33]).

determination of $\Delta m_{\text{solar}}^2$ is spectacular and shows that solar neutrino experiments are entering the era of precision physics.

The last addition to this success story is the Borexino [33] experiment. This is the lowest-threshold real-time solar neutrino experiment and the only one capable of measuring the flux of the monochromatic ${}^7\text{Be}$ neutrinos and pep neutrinos. Their recent results are shown in Fig. 28. The result is in agreement with the oscillation interpretation of other solar and reactor experiments and it adds further information to disfavour alternative exotic interpretations of the data. In summary, solar neutrinos experiments have made fundamental discoveries in particle physics and are now becoming useful for other applications, such as a precise understanding of the sun and the Earth.

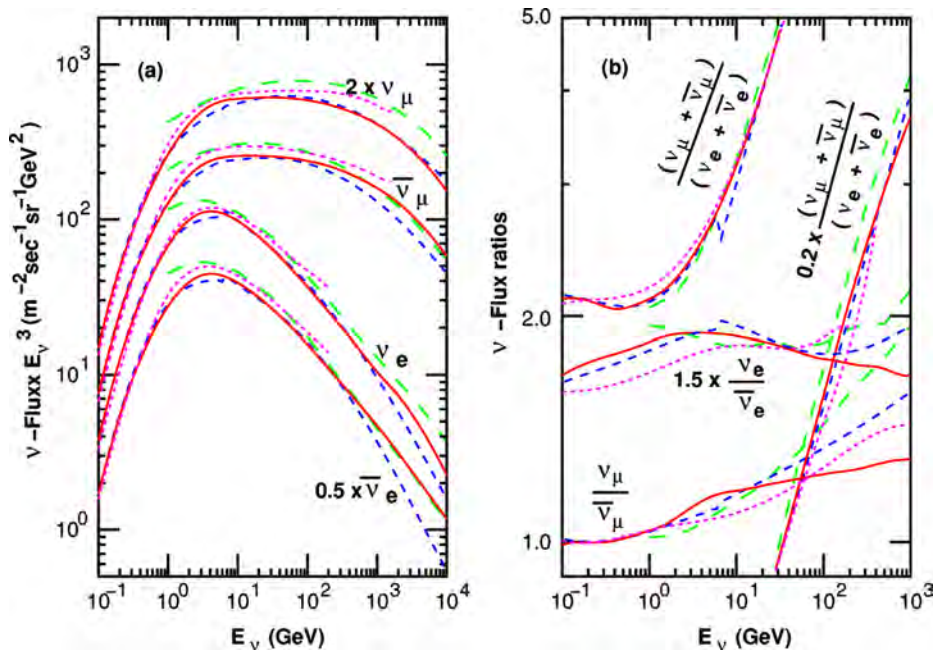


Fig. 29: Comparison of the predictions of different Monte Carlo simulations of the atmospheric neutrino fluxes averaged over all directions (left) and of the flux ratios $(\nu_\mu + \bar{\nu}_\mu)/(\nu_e + \bar{\nu}_e)$, $\nu_\mu/\bar{\nu}_\mu$, and $\nu_e/\bar{\nu}_e$ (right). The solid line corresponds to a recent full 3D simulation. Taken from the last reference in Ref. [34].

7.2 Atmospheric neutrinos

Neutrinos are also produced in the atmosphere when primary cosmic rays impinge on it producing K , π that subsequently decay. The fluxes of such neutrinos can be predicted within a 10–20% accuracy to be those in the left plot of Fig. 29.

Clearly, atmospheric neutrinos are an ideal place to look for neutrino oscillation since the E_ν/L span several orders of magnitude, with neutrino energies ranging from a few hundred MeV to 10^3 GeV and distances between production and detection varying from 10– 10^4 km, as shown in Fig. 30 (right).

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

In 1998, SuperKamiokande clarified the origin of this anomaly [35]. This experiment can distinguish muon and electron events, measure the direction of the outgoing lepton (the zenith angle with respect to the Earth’s axis) which is correlated to that of the neutrino (the higher the energy the higher the correlation), in such a way that they could measure the variation of the flux as a function of the distance travelled by the neutrinos. Furthermore, they considered different samples of events: sub-GeV (lepton with energy below 1 GeV), multi-GeV (lepton with energy above 1 GeV), together with stopping and through-going muons that are produced on the rock surrounding Superkamiokande. The different samples correspond to different parent neutrino energies as can be seen in Fig. 30 (left). The number of events for the different samples as a function of the zenith angle of the lepton are shown in the Nobel-prize-winning plot Fig. 31.

While the electron events observed are in rough agreement with predictions, a large deficit of muon events was found with a strong dependence on the zenith angle: the deficit was almost 50% for those events corresponding to neutrinos coming from below $\cos\theta = -1$, while there is no deficit for those coming from above. The perfect fit to the oscillation hypothesis is rather non-trivial given the sensitivity

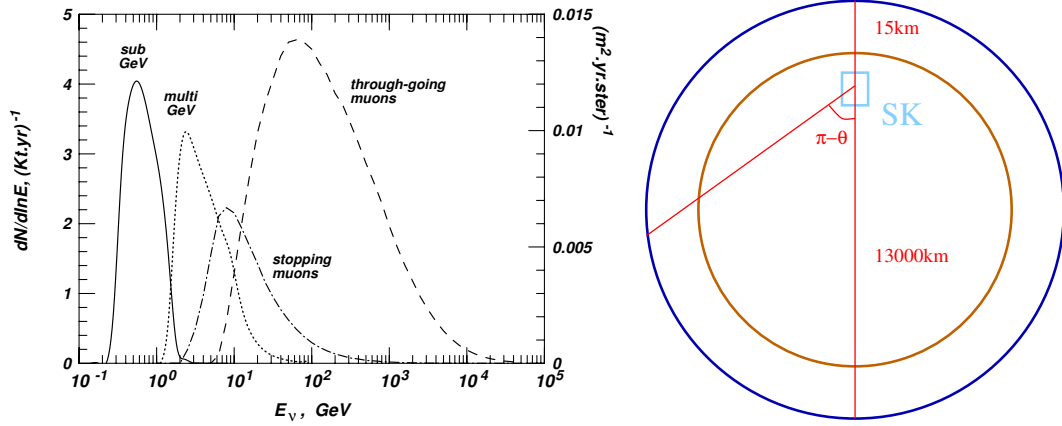


Fig. 30: Left: Parent neutrino energies of the different samples considered in Superkamiokande: sub-GeV, multi-GeV, stopping and through-going muons. Right: Distances travelled by atmospheric neutrinos as a function of the zenith angle.

of this measurement to the E_ν (different samples) and L (zenith angle) dependence. The significance of the E_ν/L dependence has also been measured by the SuperKamiokande Collaboration [37], as shown in Fig. 32. The best fit value of the oscillation parameters indicate $\Delta m^2 \simeq 3 \times 10^{-3} \text{ eV}^2$ and maximal mixing.

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

$$|\Delta m_{\text{atmos}}^2| \sim \frac{E_\nu(1 - 10 \text{ GeV})}{L(10^2 - 10^3 \text{ km})} \sim \frac{E_\nu(1 - 10 \text{ MeV})}{L(0.1 - 1 \text{ km})}. \quad (86)$$

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

$$p \rightarrow \text{Target} \rightarrow \pi^+, K^+ \rightarrow \nu_\mu (\% \nu_e, \bar{\nu}_\mu, \bar{\nu}_e) \quad (87)$$

$$\nu_\mu \rightarrow \nu_x. \quad (88)$$

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

Three conventional beams confirmed the atmospheric oscillation from the measurement of the disappearance of ν_μ neutrinos: K2K ($L = 235 \text{ km}$), MINOS ($L = 730 \text{ km}$) and from the appearance of ν_τ OPERA ($L = 730 \text{ km}$). Fig. 33 shows the measurement of the ν_μ survival probability as a function of the reconstructed neutrino energy in the MINOS experiment.

Three reactor neutrino experiments, Daya Bay [39], RENO [40] and Double Chooz [41], have discovered that the electron neutrino flavour also oscillates with the atmospheric wavelength: electron antineutrinos from reactors disappear at distances of $\mathcal{O}(1 \text{ km})$, but with a small amplitude.

Finally the T2K experiment has measured for the first time the appearance of ν_e in an accelerator ν_μ beam [42] in the atmospheric range.

The agreement of all these measurements with the original atmospheric oscillation signal was excellent.

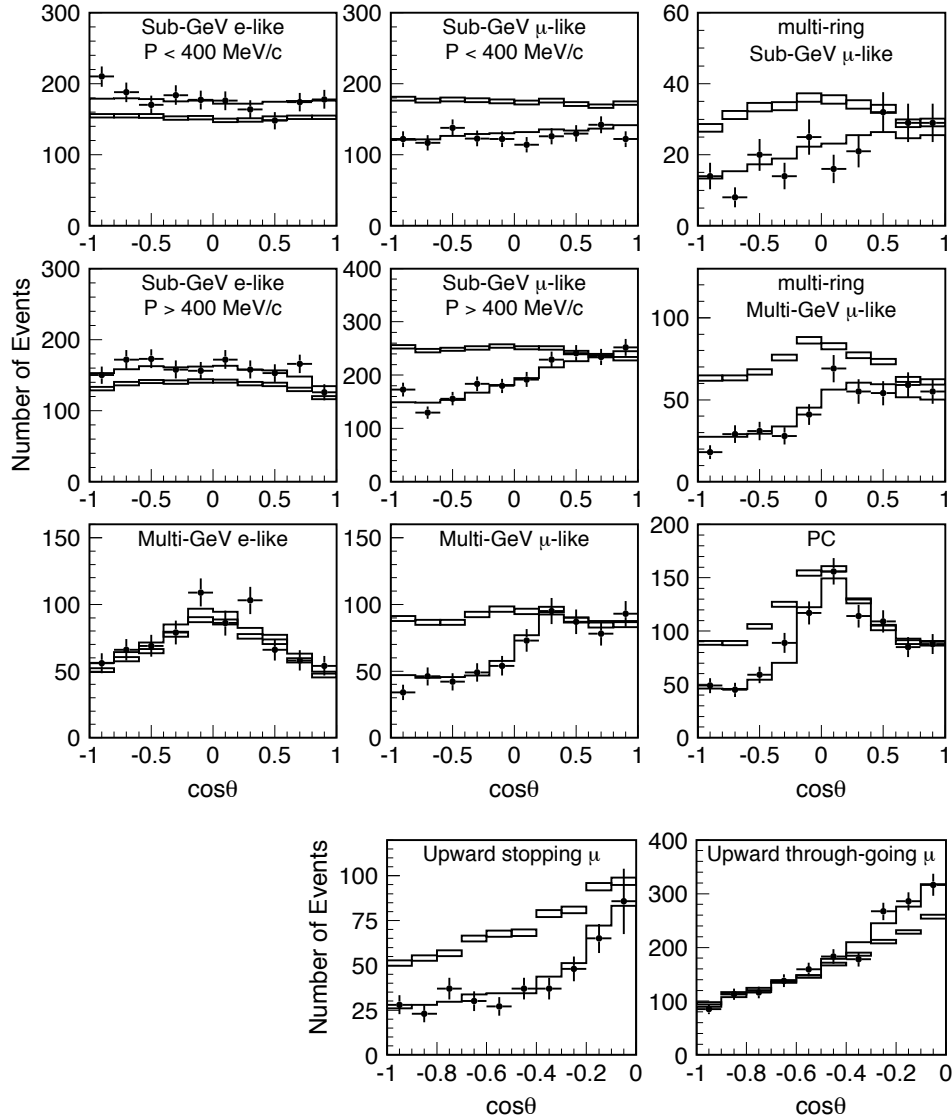


Fig. 31: Zenith angle distribution for fully-contained single-ring e -like and μ -like events, multi-ring μ -like events, partially contained events, and upward-going muons. The points show the data and the boxes show the Monte Carlo events without neutrino oscillations. The solid lines show the best-fit expectations for $\nu_\mu \leftrightarrow \nu_\tau$ oscillations (from Ref. [36]).

8 The three-neutrino mixing scenario

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

$$\underbrace{|\Delta m_{\text{solar}}^2|}_{\sim 8 \cdot 10^{-5} \text{ eV}^2} \ll \underbrace{|\Delta m_{\text{atmos}}^2|}_{\sim 2.5 \cdot 10^{-3} \text{ eV}^2} \quad (89)$$

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

$$\Delta m_{13}^2 = m_3^2 - m_1^2 = \Delta m_{\text{atmos}}^2, \quad \Delta m_{12}^2 = m_2^2 - m_1^2 = \Delta m_{\text{solar}}^2. \quad (90)$$

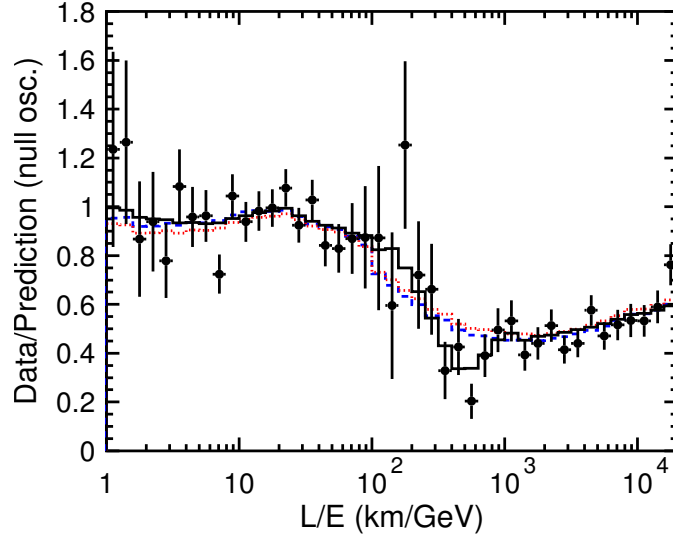


Fig. 32: Ratio of the data to the non-oscillated Monte Carlo events (points) with the best-fit expectation for 2-flavour $\nu_\mu \leftrightarrow \nu_\tau$ oscillations (solid line) as a function of E_ν/L (from Ref. [37]).

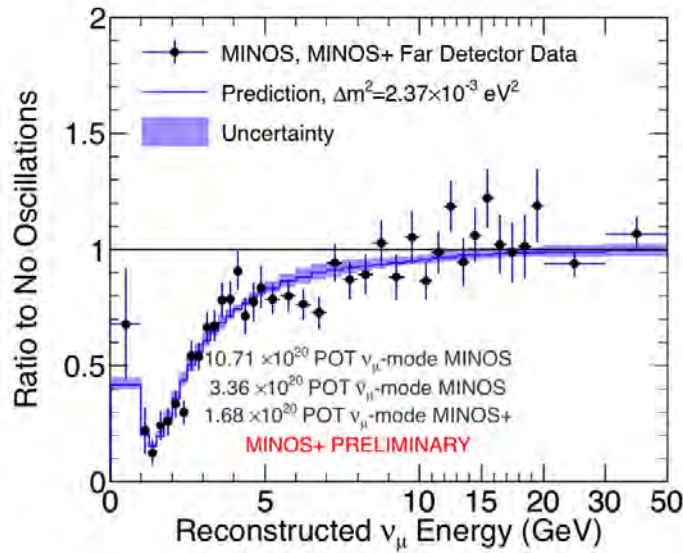


Fig. 33: Ratio of measured to expected (in absence of oscillations) neutrino events in MINOS as a function of neutrino energy compared to the best fit oscillation solution (from Ref. [38]).

The PMNS mixing matrix depends on three angles and one or more CP phases (see eq. (30) for the standard parametrization). Only one CP phase, the so-called Dirac phase δ , appears in neutrino oscillation probabilities.

With this convention, the mixing angles θ_{23} and θ_{12} in the parametrization of Eq. (30) correspond approximately to the ones measured in atmospheric and solar oscillations, respectively. This is because solar and atmospheric anomalies approximately decouple as independent 2-by-2 mixing phenomena thanks to the hierarchy between the two mass splittings, $|\Delta m_{\text{atmos}}^2| \gg |\Delta m_{\text{solar}}^2|$, on the one hand, and the fact that the angle θ_{13} , which measures the electron component of the third mass eigenstate element $\sin \theta_{13} = (U_{\text{PMNS}})_{e3}$, is small.

To see this, let us first consider the situation in which $E_\nu/L \sim |\Delta m_{\text{atmos}}^2|$. We can thus neglect

the solar mass square difference in front of the atmospheric one and E_ν/L . The oscillation probabilities obtained in this limit are given by

$$P(\nu_e \rightarrow \nu_\mu) \simeq s_{23}^2 \sin^2 2\theta_{13} \sin^2 \left(\frac{\Delta m_{13}^2 L}{4E_\nu} \right), \quad (91)$$

$$P(\nu_e \rightarrow \nu_\tau) \simeq c_{23}^2 \sin^2 2\theta_{13} \sin^2 \left(\frac{\Delta m_{13}^2 L}{4E_\nu} \right), \quad (92)$$

$$P(\nu_\mu \rightarrow \nu_\tau) \simeq c_{13}^4 \sin^2 2\theta_{23} \sin^2 \left(\frac{\Delta m_{13}^2 L}{4E_\nu} \right). \quad (93)$$

The results for antineutrinos are the same. All flavours oscillate therefore with the atmospheric frequency, but only two angles enter these formulae: θ_{23} and θ_{13} . The latter is the only one that enters the disappearance probability for ν_e or $\bar{\nu}_e$ in this regime:

$$P(\nu_e \rightarrow \nu_e) = P(\bar{\nu}_e \rightarrow \bar{\nu}_e) = 1 - P(\nu_e \rightarrow \nu_\mu) - P(\nu_e \rightarrow \nu_\tau) \simeq \sin^2 2\theta_{13} \sin^2 \left(\frac{\Delta m_{13}^2 L}{4E_\nu} \right). \quad (94)$$

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

Note that in the limit $\theta_{13} \rightarrow 0$, the only probability that survives in Eq. (93) is the $\nu_\mu \rightarrow \nu_\tau$ one, which has the same form as a 2-family mixing formula Eq. (53) if we identify

$$(\Delta m_{\text{atmos}}^2, \theta_{\text{atmos}}) \rightarrow (\Delta m_{13}^2, \theta_{23}). \quad (95)$$

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

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

$$P(\nu_e \rightarrow \nu_e) = P(\bar{\nu}_e \rightarrow \bar{\nu}_e) \simeq c_{13}^4 \left(1 - \sin^2 2\theta_{12} \sin^2 \left(\frac{\Delta m_{12}^2 L}{4E_\nu} \right) \right) + s_{13}^4. \quad (96)$$

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

$$(\Delta m_{\text{solar}}^2, \theta_{\text{solar}}) \rightarrow (\Delta m_{12}^2, \theta_{12}). \quad (97)$$

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

On the other hand, the fact that reactor experiments have already measured the disappearance of reactor $\bar{\nu}_e$ in the atmospheric range implies that the effects of $\theta_{13} \simeq 9^\circ$ are not negligible, and therefore a proper analysis of all the oscillation data requires performing global fits in the 3-family scenario. Figure 34 shows the $\Delta\chi^2$ as a function of each of the six parameters from a recent global analysis [43]. There are two parameters in which we observe to distinct minima, these corresponds to degeneracies that cannot be resolved with present data. The first corresponds to the neutrino mass ordering or hierarchy: present data cannot distinguish between the normal (NH or NO) and inverted ordering (IH or IO) represented in Fig. 35. Note that we denote by $\Delta m_{13}^2 = \Delta m_{\text{atmos}}^2$ the atmospheric splitting for NO and $\Delta m_{23}^2 = -\Delta m_{\text{atmos}}^2$. The second degeneracy corresponds to the octant choice of θ_{23} . Present data are mostly sensitive to $\sin^2 2\theta_{23}$. If this angle is not maximal, there are two possible choices that are roughly

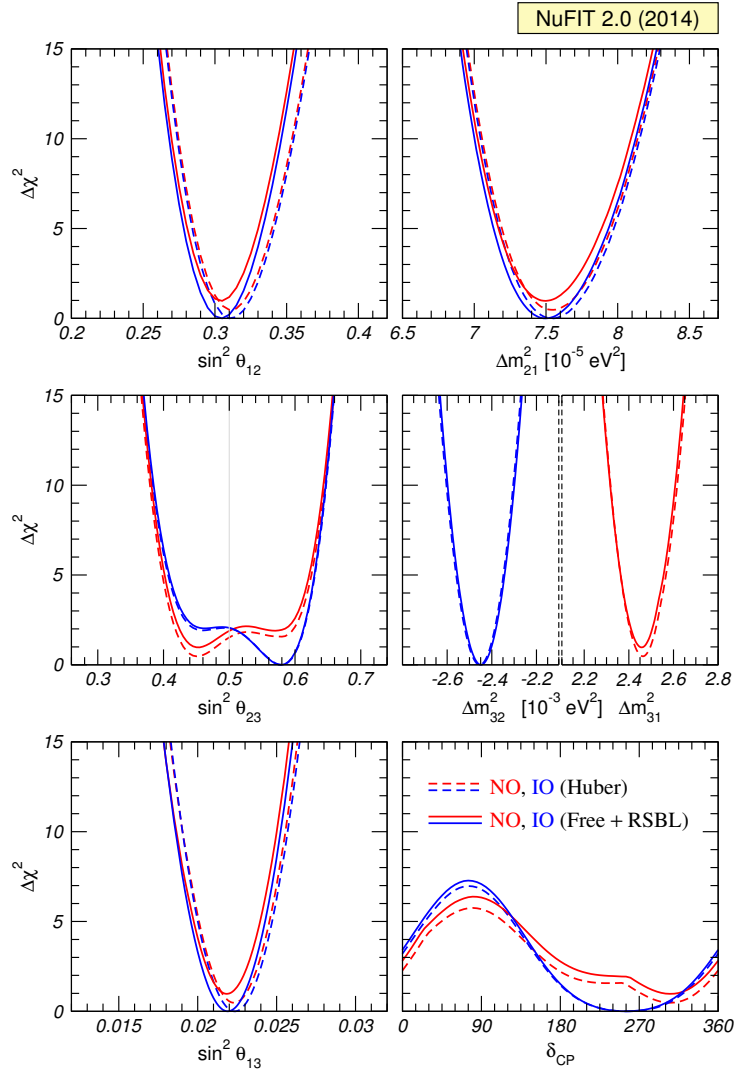


Fig. 34: Fits to the standard 3ν -mixing scenario including all available neutrino oscillation data (from Ref. [43]).

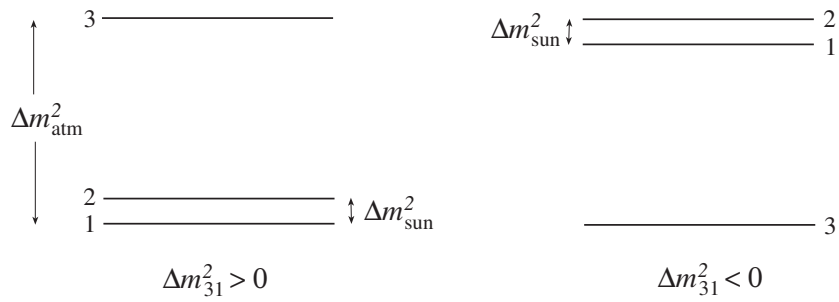


Fig. 35: Possible neutrino spectra consistent with solar and atmospheric data.

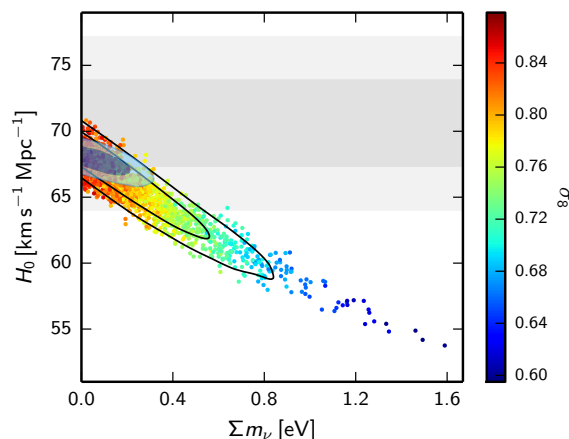


Fig. 36: Constraints on the sum of neutrino masses (in the standard 3ν scenario from cosmology [44]).

equivalent $\theta_{23} \leftrightarrow \pi/4 - \theta_{23}$. Due to this degeneracy, the largest angle is also the one less accurate. The 1σ limits for NO are:

$$\begin{aligned} \theta_{23}/^\circ &= 42_{-1.6}^{+3}, & \theta_{12}/^\circ &= 33.5_{-0.75}^{+0.78}, & \theta_{13}/^\circ &= 8.5(2), \\ \Delta m_{12}^2 &= 7.5(2) \times 10^{-5} \text{ eV}^2, & \Delta m_{13}^2 &= 2.46(5) \times 10^{-3} \text{ eV}^2. \end{aligned} \quad (98)$$

The CP phase δ remains completely unconstrained at 3σ . As we will see, the dependence on the phase requires sensitivity to both frequencies simultaneously. There is however at 2σ some hint of a preference for $\delta > 180^\circ$. For more details see [43].

Neutrino oscillations cannot provide information on the absolute neutrino mass scale. The best sensitivity to this scale is at present coming from cosmology. Indeed neutrinos properties are imprinted in the history of the universe. In particular the features of the cosmic microwave background (CMB) and the large scale galaxy distribution depends sizeably on the sum of neutrino masses. The last results from Planck [44] are shown in Fig. 36. Their conservative limit at 95%CL is impressive:

$$\sum_i m_i \leq 0.23 \text{ eV}. \quad (99)$$

9 Prospects in neutrino oscillation experiments

An ambitious experimental program is underway to pin down the remaining unknowns and reach a 1% precision in the lepton flavour parameters. The neutrino ordering, the octant of θ_{23} and the CP violating phase, δ , can be search for in neutrino oscillation experiments with improved capabilities.

Concerning the neutrino ordering, the best hope to identify the spectrum exploits the MSW effect in the propagation of GeV neutrinos through the Earth matter. In the case of three neutrinos propagating in matter, the ν mass eigenstates as a function of the electron density for vanishing θ_{12}, θ_{13} are depicted in Fig. 37 for NO and IO. For NO we see that there are two level crossings giving rise to two MSW resonances. The first one is essentially the one relevant for solar neutrinos, as it affects the smallest mass splitting, with the resonance condition:

$$E_{\text{res}}^{(1)} = \frac{\Delta m_{12}^2 \cos 2\theta_{12}}{2\sqrt{2}G_F N_e}. \quad (100)$$

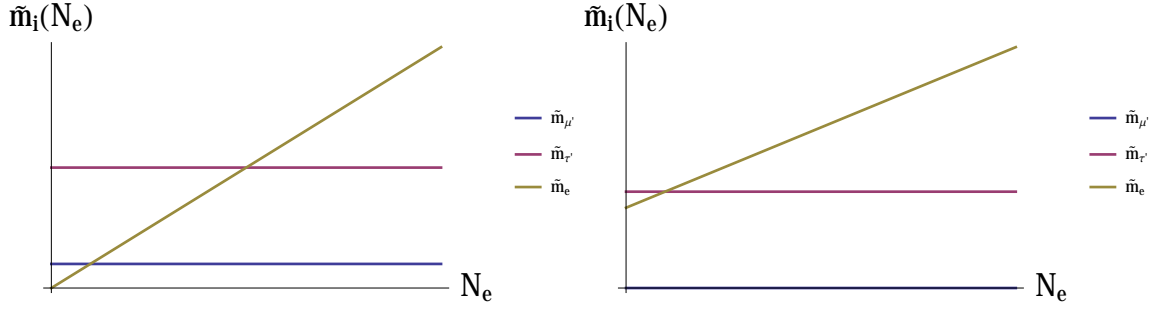


Fig. 37: Level crossings for ν in the three neutrino scenario for NO (left) and IO (right) at vanishing θ_{12} and θ_{13} .

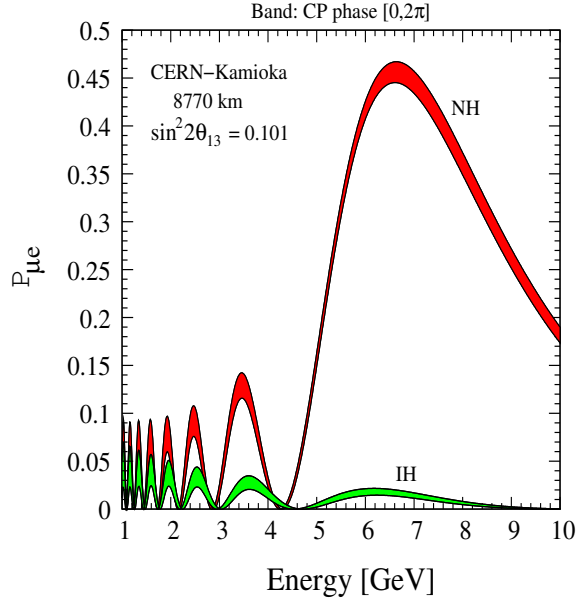


Fig. 38: $P_{\mu e}$ as a function of neutrino energy for L corresponding to the distance CERN-Kamioka for NH/IH. The bands corresponds to the uncertainty in δ (from Ref. [45]).

The second one affects the largest mass splitting

$$E_{\text{res}}^{(2)} = \frac{\Delta m_{13}^2 \cos 2\theta_{13}}{2\sqrt{2}G_F N_e}. \quad (101)$$

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

For $\bar{\nu}$ the dependence on N_e of the first eigenstate has a negative slope and therefore only the second resonance appears for IO.

The existence of the atmospheric resonance implies a large enhancement of the oscillation probability $P(\nu_e \leftrightarrow \nu_\mu)$ for NO for energies near the resonant energy and at sufficiently long baseline. For IO the enhancement occurs in $P(\bar{\nu}_e \leftrightarrow \bar{\nu}_\mu)$ instead. For the typical matter densities of the Earth crust and mantle and the value of the atmospheric mass splitting, the resonant energy for neutrinos travelling through Earth is $\simeq 6$ GeV, an energy that can be reached in accelerator neutrino beams. The measurement of the neutrino ordering becomes almost a digital measurement sending a conventional ν beam sufficiently far as shown in Fig. 38, which shows the oscillation probability $P(\nu_\mu \rightarrow \nu_e)$ as a function of the neutrino energy at a distance corresponding to the baseline from CERN-Kamioka (8770 km).

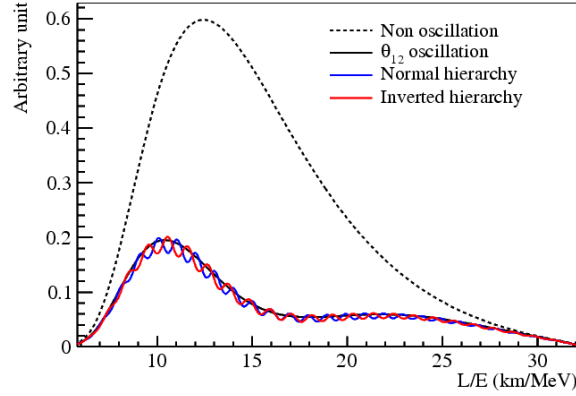


Fig. 39: Reactor neutrino spectrum in JUNO for NO/IO (from Ref. [48]).

The first experiment that will be sensitive to this effect is the NOvA experiment, optimized like T2K to see the ν_e appearance signal, with a baseline of 810km, which is however a bit short to see a large enhancement. Nevertheless if lucky NOvA could discriminate the ordering at 3σ .

The atmospheric resonance must also affect atmospheric neutrinos at the appropriate energy and baseline. Unfortunately the atmospheric flux contains both neutrinos and antineutrinos in similar numbers, and the corresponding events cannot be tell apart, because present atmospheric neutrino detectors cannot measure the lepton charge. If we superimpose the neutrino and antineutrino signals, both orderings will give rise to an enhancement in the resonance region, since no matter what the ordering is, either the neutrino or antineutrino channel will have a resonance. Nevertheless with sufficient statistics, there is discrimination power and in fact the biggest neutrino telescopes, ICECUBE and KM3NeT have proposed to instrument more finely some part of their detectors (PINGU and ORCA projects) to perform this measurement. Also the next generation of atmospheric neutrino detectors, such as HyperKamiokande, with a factor $\mathcal{O}(20)$ more mass than the present SuperKamiokande, or the INO detector that is designed to measure the muon charge in atmospheric events, could discriminate between the two orderings.

A very different strategy has been proposed for reactor neutrino experiments (e.g JUNO project). The idea is to measure very precisely the reactor neutrinos at a baseline of roughly 50 km, where the depletion of the flux due to the solar oscillation is maximal. At this optimal distance, one can get a superb measurement of the solar oscillation parameters, $(\theta_{12}, \Delta m_{12}^2)$, and, with sufficient energy resolution, one could detect the modulation of the signal due to the atmospheric oscillation [46,47]. Fig. 39 shows how this modulation is sensitive to the neutrino ordering. A leap ahead is needed to reach the required energy resolution.

9.1 Leptonic CP violation

As we have seen, the CP phase, δ , in the mixing matrix induces CP violation in neutrino oscillations, that is a difference between $P(\nu_\alpha \rightarrow \nu_\beta)$ and $P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta)$, for $\alpha \neq \beta$. As we saw in the general expression of Eq. (51), CP violation is possible if there are imaginary entries in the mixing matrix that make $\text{Im}[W_{\alpha\beta}^{jk}] \neq 0$. By CPT, disappearance probabilities cannot violate CP however, because under CPT

$$P(\nu_\alpha \rightarrow \nu_\beta) = P(\bar{\nu}_\beta \rightarrow \bar{\nu}_\alpha), \quad (102)$$

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

$$A_{\alpha\beta}^{CP} \equiv \frac{P(\nu_\alpha \rightarrow \nu_\beta) - P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta)}{P(\nu_\alpha \rightarrow \nu_\beta) + P(\bar{\nu}_\alpha \rightarrow \bar{\nu}_\beta)}, \quad A_{\alpha\beta}^T \equiv \frac{P(\nu_\alpha \rightarrow \nu_\beta) - P(\nu_\beta \rightarrow \nu_\alpha)}{P(\nu_\alpha \rightarrow \nu_\beta) + P(\nu_\beta \rightarrow \nu_\alpha)}. \quad (103)$$

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

$$A_{\nu_\alpha \nu_\beta}^{\text{CP(T)-odd}} = \frac{\overbrace{\sin \delta c_{13} \sin 2\theta_{13} \sin 2\theta_{12} \frac{\Delta m_{12}^2 L}{4E_\nu}}^{\text{solar}} \overbrace{\sin 2\theta_{23} \sin^2 \frac{\Delta m_{13}^2 L}{4E_\nu}}^{\text{atmos}}}{P_{\nu_\alpha \nu_\beta}^{\text{CP-even}}}. \quad (104)$$

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

It can be shown that including only statistical errors, the signal-to-noise for this asymmetry is maximized for $\langle E_\nu \rangle / L \sim |\Delta m_{\text{atmos}}^2|$. In this case, only two small parameters remain in the CP-odd terms: the solar splitting, $\Delta m_{\text{solar}}^2$ (i.e., small compared to the other scales, $\Delta m_{\text{atmos}}^2$ and $\langle E_\nu \rangle / L$), and the angle θ_{13} . The asymmetry is then larger in the subleading transitions: $\nu_e \rightarrow \nu_\mu (\nu_\tau)$, because the CP-even terms in the denominator are also suppressed by the same small parameters. A convenient approximation for the $\nu_e \leftrightarrow \nu_\mu$ transitions is obtained expanding to second order in both small parameters [49]:

$$\begin{aligned} P_{\nu_e \nu_\mu (\bar{\nu}_e \bar{\nu}_\mu)} &= s_{23}^2 \sin^2 2\theta_{13} \sin^2 \left(\frac{\Delta m_{13}^2 L}{4E_\nu} \right) \equiv P^{\text{atmos}} \\ &+ c_{23}^2 \sin^2 2\theta_{12} \sin^2 \left(\frac{\Delta m_{12}^2 L}{4E_\nu} \right) \equiv P^{\text{solar}} \\ &+ \tilde{J} \cos \left(\pm \delta - \frac{\Delta m_{13}^2 L}{4E_\nu} \right) \frac{\Delta m_{12}^2 L}{4E_\nu} \sin \left(\frac{\Delta m_{13}^2 L}{4E_\nu} \right) \equiv P^{\text{inter}}, \end{aligned} \quad (105)$$

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

These results correspond to vacuum propagation, but usually these experiments require the propagation of neutrinos in the Earth matter. The oscillation probabilities in matter can also be approximated by a similar series expansion [49]. The result has the same structure as in vacuum:

$$\begin{aligned} P_{\nu_e \nu_\mu (\bar{\nu}_e \bar{\nu}_\mu)} &= s_{23}^2 \sin^2 2\theta_{13} \left(\frac{\Delta_{13}}{B_\pm} \right)^2 \sin^2 \left(\frac{B_\pm L}{2} \right) \\ &+ c_{23}^2 \sin^2 2\theta_{12} \left(\frac{\Delta_{12}}{A} \right)^2 \sin^2 \left(\frac{AL}{2} \right) \\ &+ \tilde{J} \frac{\Delta_{12}}{A} \sin \left(\frac{AL}{2} \right) \frac{\Delta_{13}}{B_\pm} \sin \left(\frac{B_\pm L}{2} \right) \cos \left(\pm \delta - \frac{\Delta_{13} L}{2} \right), \end{aligned} \quad (106)$$

where

$$B_\pm = |A \pm \Delta_{13}| \quad \Delta_{ij} = \frac{\Delta m_{ij}^2}{2E_\nu} \quad A = \sqrt{2} G_F N_e. \quad (107)$$

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

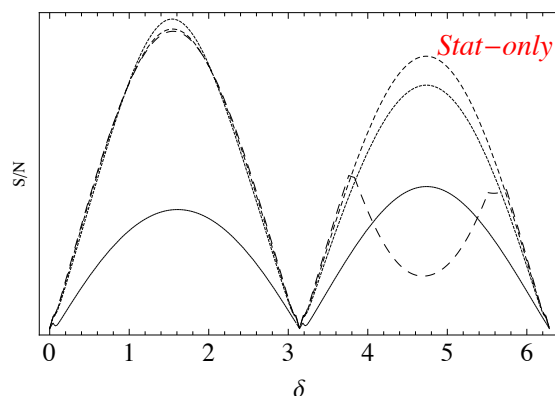


Fig. 40: Signal-to-noise for the discovery of CP violation at fixed $E/L \sim |\Delta m_{\text{atm}}^2|$ as a function of the true value of δ for $L = 295\text{km}$ (long-dashed), $L = 650\text{km}$ (short-dashed), $L = 1300\text{km}$ (dotted), $L = 2300\text{km}$ (solid). The ordering is assumed to be unknown.

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

A rough optimization of L for fixed E/L for discovering CP violation is shown in Fig. 40. It shows the signal-to-noise as a function of the true value of δ , assuming only statistical errors, but including the expected dependence of the cross sections and fluxes. At very short baselines, the sensitivity is compromised due to the lack of knowledge of the neutrino ordering. In a wide intermediate region around $\mathcal{O}(1000)\text{km}$ the sensitivity is optimal, and at much larger baselines the sensitivity deteriorates because the matter effects completely hide CP-violation.

Several projects have been proposed to search for leptonic CP violation, including conventional beams, but also novel neutrino beams from muon decays (neutrino factories), from radioactive ion decays (β -beams) or from spallation sources (ESS). The relatively large value of θ_{13} has refocused the interest in using the less challenging conventional beams and two projects are presently being developed: the HyperKamiokande detector, an upscaled version of SuperKamiokande that will measure atmospheric neutrinos with unprecedented precision, and also intercept a neutrino beam from JPARC at a relatively short baseline $L = 295\text{km}$, and the DUNE project that involves a ~ 30 kton liquid argon neutrino detector and a neutrino beam from Fermilab to the Soudan mine at a baseline of $L = 1500\text{km}$. The expected sensitivities to the neutrino ordering and to CP violation of both projects are shown in Figs. 41., 42.

10 Outliers: the LSND anomaly

The long-standing puzzle brought by the LSND experiment is still unresolved. This experiment [52] observed a surplus of electron events in a muon neutrino beam from π^+ decaying in flight (DIF) and a surplus of positron events in a neutrino beam from μ^+ decaying at rest (DAR). The interpretation of this data in terms of neutrino oscillations, that is a non-vanishing $P(\nu_\mu \rightarrow \nu_e)$, gives the range shown by a coloured band in Fig. 44.

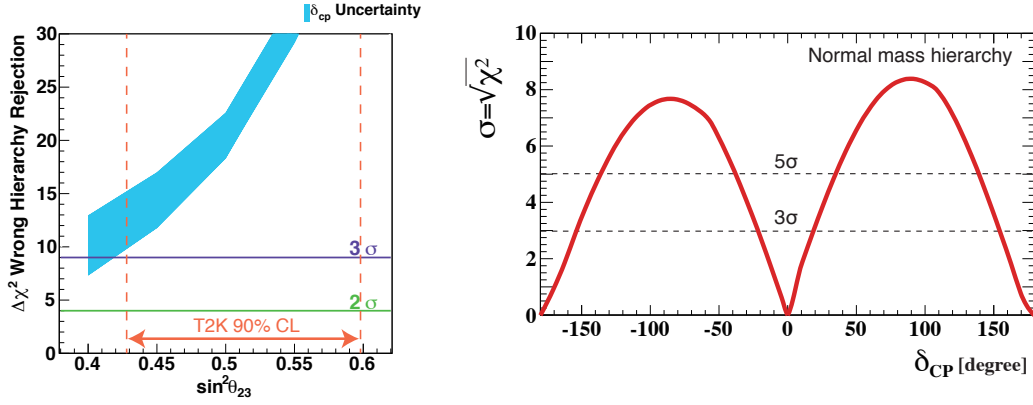


Fig. 41: Prospects for determining the ordering (left) and discovering CP violation (right) in HyperKamiokande (from Ref. [50]).

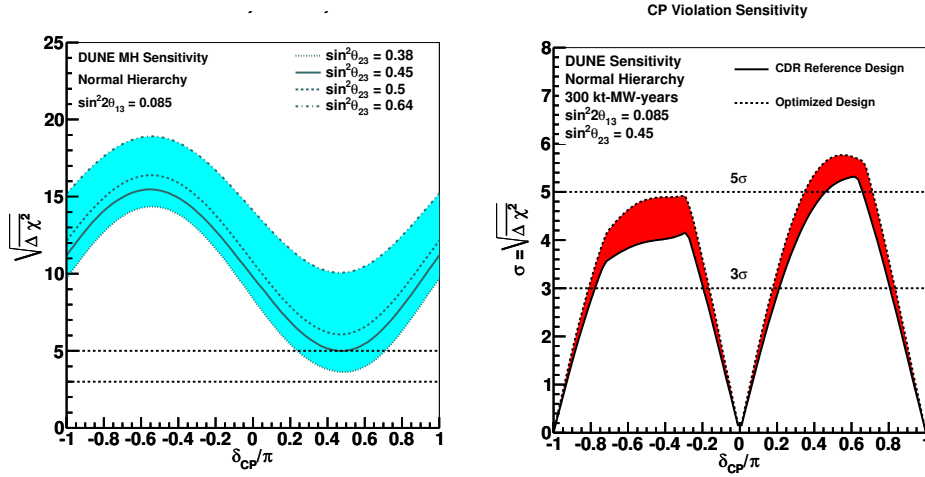


Fig. 42: Prospects for determining the ordering (left) and discovering CP violation (right) in DUNE (from Ref. [51]).

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

A significant fraction of this region was already excluded by the experiment KARMEN [53] that has unsuccessfully searched for $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ in a similar range.

The experiment MiniBOONE was designed to further investigate the LSND signal, with inconclusive results [54]. They did not confirm the anomaly, but found some anomaly at lower energies.

On the other hand, recently the results of various short baseline (tens of meters) reactor neutrino experiments were revised, after an update on the reactor neutrino flux predictions [55–57], which increased these fluxes by a few per cent. While the measured neutrino flux was found to be in agreement with predictions before, after this revision some reactor neutrinos seem to disappear before reaching near detectors, $L = \mathcal{O}(10)\text{m}$. This is the so-called reactor anomaly shown in Fig. 43. This result brought some excitement because if this disappearance is due to oscillations, it might reinforce the oscillation interpretation of the LSND anomaly.

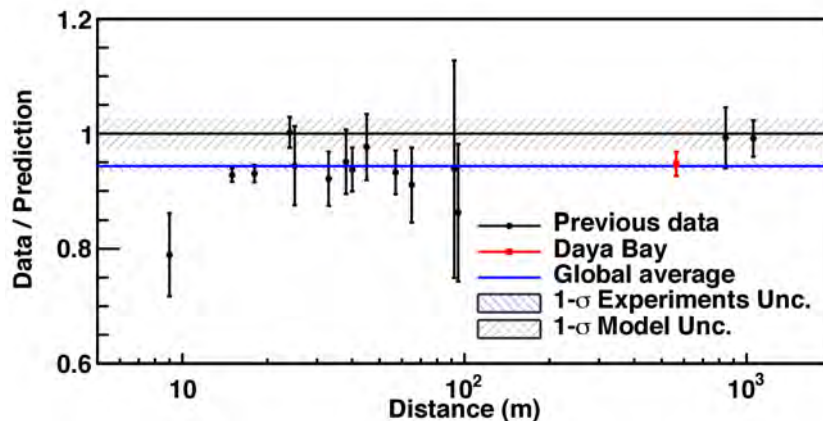


Fig. 43: Reactor neutrino flux measured by various near detectors compared with the recent flux predictions (from Ref. [58]).

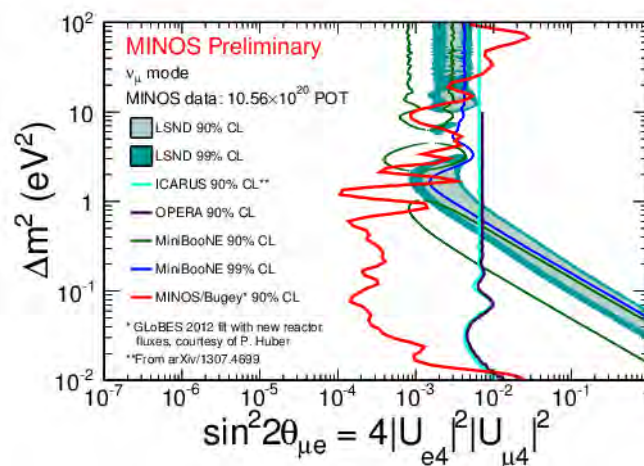


Fig. 44: Sterile neutrino search in the disappearance of ν_μ 's in MINOS (from Ref. [59]).

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

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

The strongest constraint on the disappearance of ν_μ in the LSND range has been recently set by MINOS+ [59] and is shown in Fig. 44 together with the region favoured by the LSND anomaly. An improvement of this sensitivity is expected also from the measurement of atmospheric neutrinos in ICECUBE.

A number of experiments are being constructed to clarify the reactor anomaly. Hopefully in the near future they will settle this long-standing puzzle.

11 Neutrinos and BSM Physics

The new lepton flavour sector of the SM has opened new perspectives into the flavour puzzle. As we have seen neutrinos are massive but significantly lighter than the remaining charged fermions. Clearly the gap of Fig. 11 calls for an explanation. The leptonic mixing matrix is also very different to that in the quark sector. The neutrino mixing matrix is approximately given by [43]

$$|U_{\text{PMNS}}|_{3\sigma} \simeq \begin{pmatrix} 0.80 - 0.84 & 0.51 - 0.58 & 0.137 - 0.158 \\ 0.22 - 0.52 & 0.44 - 0.70 & 0.61 - 0.79 \\ 0.25 - 0.53 & 0.46 - 0.71 & 0.59 - 0.78 \end{pmatrix}. \quad (108)$$

The CKM matrix is presently constrained [7] to be:

$$|V_{\text{CKM}}| \simeq \begin{pmatrix} 0.97427(14) & 0.22536(61) & 0.00355(15) \\ 0.22522(61) & 0.97343(15) & 0.0414(12) \\ 0.00886(33) & 0.0405(12) & 0.99914(5) \end{pmatrix}. \quad (109)$$

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

$$V_{\text{CKM}} \simeq \begin{pmatrix} 1 & O(\lambda) & O(\lambda^3) \\ O(\lambda) & 1 & O(\lambda^2) \\ O(\lambda^3) & O(\lambda^2) & 1 \end{pmatrix}, \quad \lambda \sim 0.2, \quad (110)$$

while the leptonic one has large off-diagonal entries. With a similar level of precision, it is close to the tri-bimaximal mixing pattern [60]

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

Discrete flavour symmetries have been extensively studied as the possible origin of this pattern. For recent review see [61].

While we do not have yet a compelling explanation of the different mixing patterns, we do have one for the gap between neutrino and other fermion masses. We saw that if the light neutrinos are Majorana particles and get their mass via the Weinberg interaction of Fig. 12, they are signalling BSM physics. Neutrino masses are suppressed because they arise from a new scale of physics that could be $\Lambda \gg v$. Generically such BSM would induce not only neutrino masses but also other effects represented at low-energies by the $d = 6$ effective operators of eq. (23). Unfortunately the list of $d = 6$ operators is too long to be of guidance: which one might be more relevant is to a large extent model dependent.

We could argue that there is not better motivated BSM physics than the one that gives rise to the Weinberg operator. The simplest possibility is that Weinberg's operator, like the Fermi one in Fig.2, arises from the exchange of a massive particle at tree level. The classification of what particles can induce the Weinberg operator at tree level has been done, and reproduces the three types of seesaw models, as depicted in Fig. 45:

- type I see-saw: SM+ heavy singlet fermions [62–65],
- type II see-saw: SM + heavy triplet scalar [66–70],
- type III see-saw: SM + heavy triple fermions [71, 72],

or combinations. The masses of the extra states define the scale Λ .

It is also possible that Weinberg's interaction is generated by new physics at higher orders, such as in the famous Zee model [73] and related ones [74, 75]. In this case, neutrino masses have an additional suppression by loop factors $1/(16\pi^2)$.

The $d = 6$ operators induced at tree level in see-saw models of Types I to III have been worked out [76]. They give rise to a rich phenomenology that could help discriminate between the models. In particular, they could induce beyond-the-standard-model signals in Z and W decays, deviations in the ρ parameter or the W mass, and mediate rare lepton decays, as well as violations of universality and unitarity of the neutrino mass matrix. It would therefore be extremely important to search for these effects. Whether they are large enough to be observed or not depends strongly on how high the scale Λ is, since all these effects are suppressed by two powers of Λ .

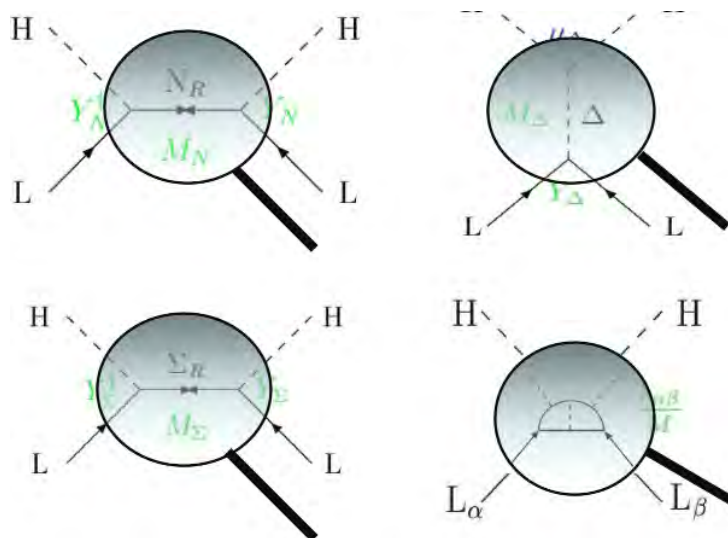


Fig. 45: Magnifying-glass view of Weinberg operator in see-saws Type I (top left), Type II (top right), Type III (bottom left) and Zee–Babu model (bottom right).

Unfortunately, the measurement of Weinberg's interaction leaves behind an unresolved $\lambda \leftrightarrow \Lambda$ degeneracy that makes it impossible to know what the scale of the new physics is, even if we were to know the absolute value of neutrino masses.

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

The most popular choice for Λ has traditionally been a grand-unification scale, given the intriguing fact that the seesaw-type ratio $\frac{v^2}{M_{\text{GUT}}} \sim 0.01\text{--}0.1$ eV, in the right ballpark of the neutrino mass scale. However, in the absence of any stabilizing mechanism such as supersymmetry, however, the electroweak scale needs to be fine-tuned [79, 80] since the Higgs mass receives quadratic loop corrections in Λ . A naturalness argument would then imply that $\Lambda < 10^7$ GeV. The opposite is not true however, the scale Λ would not get corrections from the electroweak scale: any value of $\Lambda \leq v$ is technically natural.

The possibility that the scale Λ might be of the order of the EW scale or lower has recently been studied in more detail, with special emphasis on establishing the existing experimental constraints, and the possibility that this new physics could explain other open problems in the SM such as: the LSND and reactor anomalies, dark matter, leptogenesis, etc. The type I seesaw model is the case better studied so we will concentrate on pinning down the scale Λ in this context.

11.1 Type I seesaw model

It is arguably the minimal extension of the SM allowing for neutrino masses [62–65]. It involves the addition of $n_R \geq 2$ singlet Weyl fermions, ν_R , to the SM. The most general renormalizable Lagrangian which satisfies Lorentz and the gauge symmetries is given by:

$$\mathcal{L}_{\text{TypeI}} = \mathcal{L}_{\text{SM}} - \sum_{\alpha,i} \bar{L}^\alpha Y_\nu^{\alpha i} \tilde{\Phi} \nu_R^i - \sum_{i,j} \frac{1}{2} \bar{\nu}_R^{ic} M_N^{ij} \nu_R^j + \text{h.c.}, \quad (111)$$

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

$$\mathcal{L}_{\text{TypeI}} \rightarrow \mathcal{L}_{\text{SM}} - \frac{1}{2} \begin{pmatrix} \bar{\nu}_L & \bar{\nu}_R^c \end{pmatrix} \begin{pmatrix} 0 & m_D \\ m_D^T & M_N \end{pmatrix} \begin{pmatrix} \nu_L^c \\ \nu_R \end{pmatrix} + \text{h.c.} + \dots \quad (112)$$

where

$$m_D = Y_\nu \frac{v}{\sqrt{2}}. \quad (113)$$

Note that Dirac neutrinos are a particular case of the model. If we invoke a global lepton number symmetry to force $M_N = 0$, the singlets are exactly equivalent to the right-handed neutrinos in the Dirac case described in sec. 3.1. In the opposite limit $M_N \gg v$, the singlets can be integrated out and give rise to the Weinberg interaction as well as others at $d = 6$, etc. For finite M_N , the spectrum of this theory contains in general $3 + n_R$ Majorana neutrinos, which are admixtures of the active ones and the extra singlets. It is easy to diagonalize the mass matrix in eq. (112) in an expansion in m_D/M_N . The result to leading order in this expansion is

$$U^T \begin{pmatrix} 0 & m_D \\ m_D^T & M_N \end{pmatrix} U \simeq \begin{pmatrix} -m_D \frac{1}{M_N} m_D^T & 0 \\ 0 & M_N \end{pmatrix} + \mathcal{O}(\theta^2), \quad U = \begin{pmatrix} 1 & \theta \\ -\theta^\dagger & 1 \end{pmatrix}, \quad (114)$$

where

$$\theta = m_D^* \frac{1}{M_N}. \quad (115)$$

To this order therefore the light neutrino and heavy neutrino masses are given by

$$m_l = \text{Diag} \left[-m_D \frac{1}{M_N} m_D^T \right], \quad M_h = \text{Diag}[M_N]. \quad (116)$$

Fig. (46) depicts the spectrum for the case of $n_R = 3$ as a function of a common M_N . In the limit $M_N \rightarrow 0$ the states degenerate in pairs to form Dirac fermions. As M_N increases three states get more massive proportional to M_N , while three get lighter proportional to M_N^{-1} . This is why the model is called seesaw. The number of new free parameters is large. For the case $n_R = 3$ there are 18 fundamental parameters in the lepton sector: six of them are masses, six mixing angles and six phases. The counting of parameters for general n_R is shown in Table 4. Out of these 18 parameters we have determined only 5: two mass differences and three neutrino mixing angles.

A very convenient parametrization in this model is the so-called Casas-Ibarra [81] parametrization, which allows to write in all generality (up to corrections of $\mathcal{O}(\theta^2)$) the Lagrangian parameters in terms of the parameters that can be measured at low energies: light neutrino masses and mixings, and others that cannot. In particular the phenomenology of this model depends on the spectrum of neutrino mass

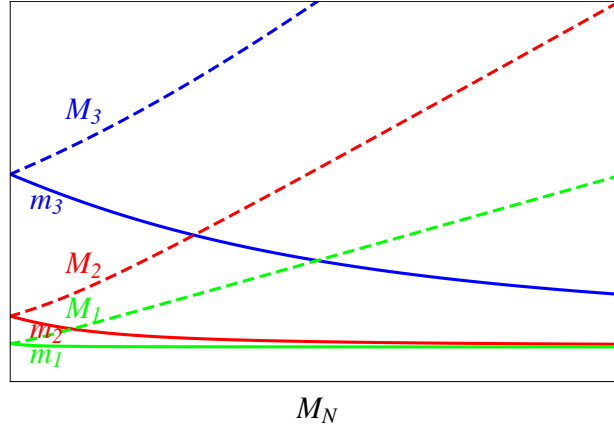


Fig. 46: Spectrum of the type I seesaw model for $n_R = 3$ as a function of a common M_N .

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

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

eigenstates, that we denote by $(\nu_1, \nu_2, \nu_3, N_1, N_2, \dots, N_{n_R})$, and their admixture in the flavour neutrino states which is given by:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U_{ll} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} + U_{lh} \begin{pmatrix} N_1 \\ N_2 \\ \dots \\ N_{n_R} \end{pmatrix}. \quad (117)$$

In the Casas-Ibarra parametrization we have

$$\begin{aligned} U_{ll} &= U_{\text{PMNS}} + \mathcal{O}(\theta^2), \\ U_{lh} &= iU_{\text{PMNS}} \sqrt{m_l} R \frac{1}{\sqrt{M_h}} + \mathcal{O}(\theta^2). \end{aligned} \quad (118)$$

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

The Casas-Ibarra parametrization needs to be modified in the presence of large unitarity violations. A similar parametrization valid to all orders in θ is given in [82].

Eqs. (118) indicate that in this model there is a strong correlation between flavour mixings of the heavy states, U_{lh} , and the ratio of light-to-heavy neutrino masses. However the presence of the unknown matrix R , which is not bounded, implies that the naive seesaw scaling, $|U_{lh}|^2 \sim m_l/M_h$, that would hold exactly for one neutrino family, is far too naive for $n_R > 1$. In fact there are regions of parameter space where these mixings can be much larger than suggested by the naive scaling, and these are precisely the regions with more phenomenological interest, as we will see below.

In this model we can ask the question. What is the value of the M_N scale to avoid the hierarchy between neutrinos and the remaining fermions. If we plot the distribution of Yukawa couplings instead of the masses, we find that neutrino masses can be explained with a scale $M_N \simeq \text{GUT}$, if the neutrino yukawa couplings are of $O(1)$ like the top. However if the yukawas are of the order of the electron yukawa, a scale $M_N \sim \text{TeV}$ can also explain neutrino masses. Clearly, in both cases we have avoided making neutrinos especial, and the flavour puzzle is no worse than in the charged fermion sector. Note that this wide range of scales between TeV-GUT is the result of the quadratic dependence of the light neutrino masses on the yukawas, as opposed to the linear dependence in the Dirac case.

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

Neutrinoless double-beta decay

For $M_N \geq 100 \text{ MeV}$, the model implies the presence of neutrinoless double beta decay at some level. The amplitude for this process gets contribution from the light and heavy states:

$$m_{\beta\beta} \equiv \sum_{i=1}^3 (U_{\text{PMNS}})_{ei}^2 m_i + \sum_{j=1}^{n_R} (U_{lh})_{ej}^2 M_j \frac{\mathcal{M}^{\beta\beta 0\nu}(M_j)}{\mathcal{M}^{\beta\beta 0\nu}(0)}, \quad (119)$$

where the ratio of matrix elements $\mathcal{M}^{\beta\beta 0\nu}$ for heavy and light mediators satisfy [83]:

$$\frac{\mathcal{M}^{\beta\beta 0\nu}(M_j)}{\mathcal{M}^{\beta\beta 0\nu}(0)} \propto \left(\frac{100 \text{ MeV}}{M_j} \right)^2, \quad M_j \rightarrow \infty. \quad (120)$$

If all the heavy state masses $\gg 100 \text{ MeV}$, the second term is suppressed and the amplitude contains only the light neutrino masses and mixings:

$$m_{\beta\beta} \simeq |c_{13}^2 (m_1 c_{12}^2 + m_2 e^{i\alpha_1} s_{12}^2) + m_3 e^{i\alpha_2} s_{13}^2|, \quad (121)$$

and is quite well constrained from neutrino oscillation experiments. Figure 47 shows the present allowed regions for $m_{\beta\beta}$ neglecting the heavy state contributions as a function of the sum of the light neutrino masses, that can be constrained from cosmology:

$$\Sigma \equiv m_1 + m_2 + m_3. \quad (122)$$

If the neutrino ordering would be inverted or Σ not much smaller than 0.1 eV , there is a good chance that the next generation of $\beta\beta 0\nu$ experiments will see a signal. A plethora of experiments using different technologies have been proposed to reach a sensitivity in $m_{\beta\beta}$ in the range of 10^{-2} eV , which could be sufficient to explore the full parameter space in the case of the IO. The importance of this measurement can hardly be overstated. A non-zero $m_{\beta\beta}$ will imply that neutrinos are Majorana and therefore a new physics scale must exist, that lepton number is violated, and might give very valuable information on the lightest neutrino mass, and even help establishing the neutrino mass ordering. If the heavy states are not too heavy, within 100 MeV -few GeV , they could also contribute to the process significantly and even dominate over the light neutrino contribution for both orderings [85–87].

Mini-seesaw and oscillations

If the scale $M_N \leq \text{eV}$ (mini-seesaw models [88]), the extra heavy states could affect neutrino oscillations significantly. Strong constraints can be derived therefore from neutrino oscillation experiments

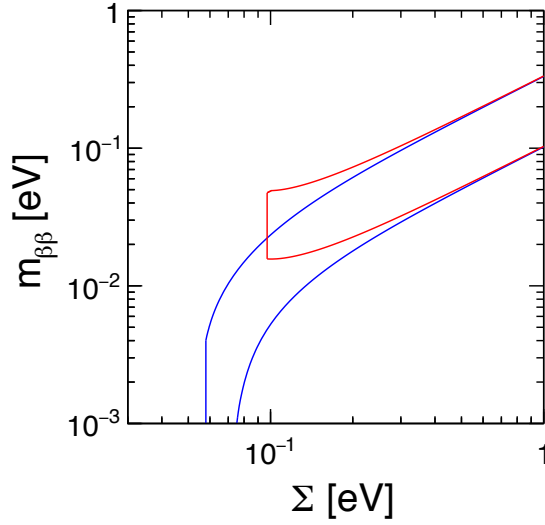


Fig. 47: Allowed region for $m_{\beta\beta}$ for NO (blue contour) and IO (red contour) from a global analysis of neutrino data (from Ref. [84]).

which essentially exclude the possibility that $M_N \in [10^{-9}(10^{-11}) \text{ eV}, 1 \text{ eV}]$ for NO(IO) [89–91]. The possibility that $M_N \sim 1 \text{ eV}$ could explain the LSND and reactor anomalies has also been studied. It is intriguing that the best fit to the LSND and reactor anomalies [92, 93] point to mixings and masses of the extra neutrino(s) that nicely match the naive seesaw scaling. In fact mini-seesaw models provide similar fits to data [82], with much less parameters, than general phenomenological models with $3 + N$ neutrino mixing. Both are affected however by the tension in data between the anomalies and the non-observation of ν_μ disappearance.

Cosmology and the seesaw scale

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

The singlet states in this mass range are produced at T below the electroweak phase transition via mixing. A simple estimate of their production rate is

$$\Gamma_{s_i}(T) \simeq \sum_{\alpha} |(\tilde{U}_{lh})_{\alpha i}|^2 \times \Gamma_{\nu_{\alpha}}(T), \quad (123)$$

where $\Gamma_{\nu_{\alpha}}$ is the interaction rate of the active neutrinos and the \tilde{U}_{lh} is the light-heavy mixing at T , strongly modified by forward scattering on the plasma particles [94]. The state i will reach thermal equilibrium if $\Gamma_{s_i}(T)$ is larger than the Hubble parameter at some T . If this is the case, the extra species will contribute like one extra neutrino for $T > M_i$ or like an extra component of dark matter for $T < M_i$. The latest results from Planck strongly constrain an extra radiation component at CMB:

$$N_{\text{eff}}(\text{CMB}) = 3.2 \pm 0.5. \quad (124)$$

and also measures the dark matter component to be $\Omega_m = 0.308 \pm 0.012$. Similar bounds are obtained from the abundance of light elements, BBN. These bounds exclude the possibility of having essentially any extra fully thermalized neutrino that is sufficiently long-lived to survive BBN. It can be shown that

the ratio $\frac{\Gamma_{s_i}(T)}{H(T)}$ reaches a maximum at T_{max} [95, 96] and

$$\frac{\Gamma_{s_i}(T_{max})}{H(T_{max})} \sim \frac{\sum_{\alpha} |(U_{lh})_{\alpha i}|^2 M_i}{\sqrt{g_*(T_{max})}}. \quad (125)$$

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

For $n_R = 2$, the heavy states must be $M_i \geq 100$ MeV [97], so that they might decay before BBN. For $n_R = 3$ two things can happen [98]. If the lightest neutrino mass $m_{\text{lightest}} \geq 3 \times 10^{-3}$ eV, all the three heavy states thermalize so $M_i \geq 100$ MeV. If $m_{\text{lightest}} \leq 3 \times 10^{-3}$ eV two states must be above this limit, but one of the states with mass M_1 might not thermalize and therefore be sufficiently diluted. M_1 may take any value provided m_{lightest} , which is presently unconstrained, is tuned accordingly.

The states that could explain the LSND and reactor anomalies will imply $\Delta N_{\text{eff}} \geq 1$ which is essentially excluded by cosmology. For a recent detailed analysis see [99]. Exotic extensions involving hidden interactions of the extra singlet states would be needed to make them compatible.

Warm dark matter

For $m_{\text{lightest}} \leq 10^{-5}$ eV, M_1 might be $\mathcal{O}(\text{keV})$, and a viable warm dark matter candidate [100, 101]. This scenario is the so-called νMSM model [101]. The most spectacular signal of this type of Dark Matter is a monochromatic X-ray line. Two recent analyses [102, 103] have recently shown some evidence for an unexplained X-ray line in galaxy clusters that might be compatible with a 7 keV neutrino. These results are under intense debate. If interpreted in terms of a keV neutrino, the mixing however is too small and some extra mechanism is needed to enhance the production so that it matches the required dark matter density, such as the presence of large primordial lepton asymmetries [104].

Direct searches

In summary, cosmology and neutrino oscillations restrict a huge range of $M_N \in [10^{-17} - 10^2]$ MeV. Naturalness arguments on the other hand point to a scale $M_N \leq 10^{10}$ MeV, suggesting that maybe the scale of M_N is not far from the electroweak scale. States with masses in this range could be produced on the lab and searched for as peaks in meson decays, in beam dump experiments, colliders, etc. [105]. The present experimental bounds on the e, μ mixings of these heavy states are shown in Figs. 48. The shaded regions correspond to existing constraints and the unshaded ones to prospects of various new experiments. For masses below a few GeV, the best constraints come from peak searches in meson decays. In particular the new beam dump experiment SHiP [106] can improve considerably the sensitivity in the region above kaon decays. For the lighter hadrons, improvements can be achieved with the more intense beams expected in long-baseline accelerator neutrino projects such as DUNE [107]. For

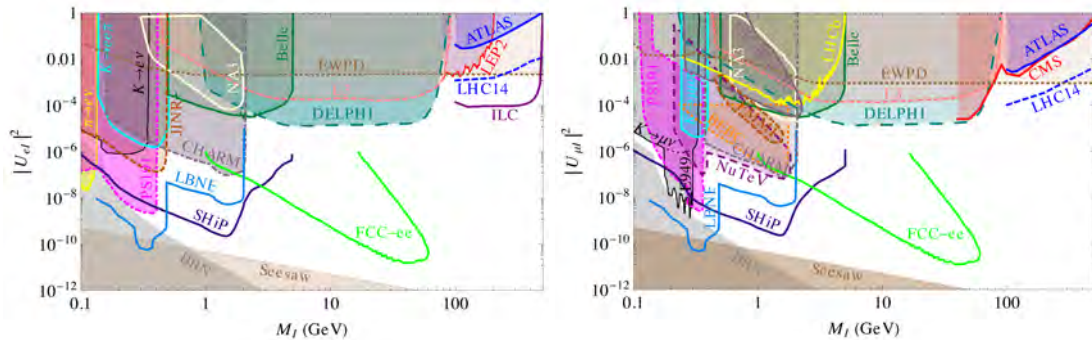


Fig. 48: Constraints from present and future experiments on a heavy neutrino with mixing to the electron (left) and muon (right). Shaded regions are existing bounds and the empty ones are prospects (from Ref. [106]).

masses above the W, Z masses the best constraints are presently coming from LHC searches. Processes with three leptons in the final state as in Fig. 49 seem most promising [108], although other production mechanisms like $W\gamma$ fusion can dominate at higher masses [109]. For a recent review and further references see [106]. For masses below the W mass, it has been pointed out recently that LHC might also improve the present constraints by looking for displaced vertices in the range 1mm-1m [110–112].

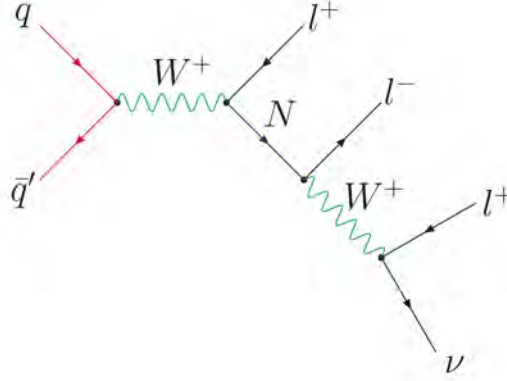


Fig. 49: Process to search for heavy Majoranas at LHC.

Note that present sensitivities are very far from the naive seesaw scaling $|U_{lh}|^2 \sim m_l/M_h$, so they are only exploring a relatively small corner of parameter space.

Lepton flavour violating processes

Massive neutrinos imply that lepton flavour violating processes, such as $\mu \rightarrow e\gamma$, eee or $\mu - e$ conversion in atoms, must exist at some level. Heavy Majorana neutrinos around the EW scale can significantly enhance these rates. The constraints on the mixing and mass coming from these searches cannot be included in Figs. 48 without further assumptions, since they depend on the different combination, $|\sum_i U_{ei}U_{\mu i}^*|$. They are shown in Fig. 50 and compared with other present constraints. Future searches will significantly improve present constraints for $M_N \in [1, 100]$ GeV.

11.2 Approximate Lepton Number

Type I seesaw models with a scale around the electroweak scale are very hard to test unless $|U_{lh}|^2 \gg m_l/M_h$. Although this is possible in some corners of parameter space for $n_R \geq 2$, being in such corners might be enforced and technically natural by an approximate lepton number symmetry [113, 114].

Let us consider the simplest case $n_R = 2$ [115]. If the two singlet states have opposite lepton charges and we impose an exact $U(1)$ global symmetry, the 3×2 Yukawa matrix, and the Majorana mass matrix have the following structures:

$$Y_\nu = \begin{pmatrix} Y_{e1} & 0 \\ Y_{\mu 1} & 0 \\ Y_{\tau 1} & 0 \end{pmatrix}, \quad M_N = \begin{pmatrix} 0 & M \\ M & 0 \end{pmatrix}. \quad (126)$$

For this texture, the heavy states form a Dirac pair, while the light neutrino masses vanish identically. The global symmetry can be only approximate if the zero entries in these matrices are small compared to the non-zero ones, but non vanishing. For example if we lift the zero in the 22 element of the M_N matrix to be $\mu \ll M$, we get the type of texture found in the so-called inverse seesaw models³ [116, 117]. In

³In order to get at least two non-zero light neutrino masses by lifting the zeros of M_N only, it is necessary to have two pairs of singlets, each pair with +1 or -1 lepton charge, ie. $n_R = 4$. For $n_R = 2$, the zero's in the Yukawa matrix must be lifted aswell.

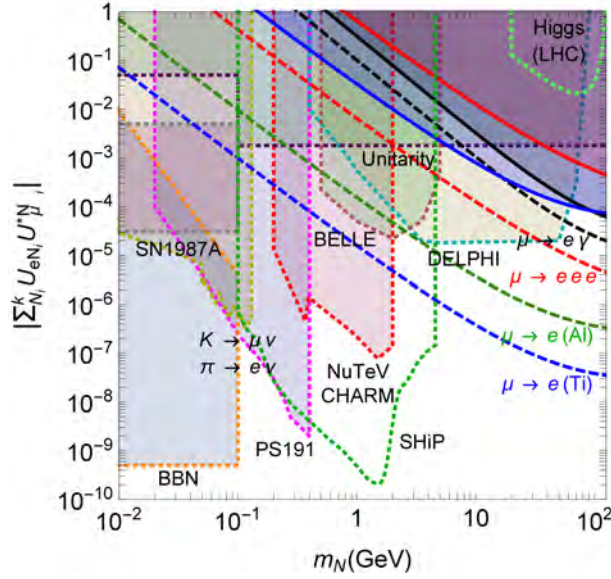


Fig. 50: Present bounds and prospects from lepton $\mu \rightarrow e\gamma$, $\mu \rightarrow eee$ and $\mu - e$ conversion searches (from Ref. [106]).

this case the light neutrino mass satisfies $m_l \propto \frac{\mu}{M^2}$. If we integrate out the scale M , the effective field theory describing this type of models is of the form

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \sum_i \frac{\alpha_i}{\Lambda_{\text{LN}}} \mathcal{O}_i^{d=5} + \sum_i \frac{\beta_i}{\Lambda_{\text{LFV}}^2} \mathcal{O}_i^{d=6} + \dots, \quad (127)$$

where the operators that break lepton number ($d = 5$) and those that preserve this symmetry ($d = 6$) are generically suppressed by different scales: $\Lambda_{\text{LN}} \simeq \frac{M^2}{\mu} \gg \Lambda_{\text{LFV}} \simeq M$. These models therefore have a richer phenomenology if M is at the EW scale, since yukawa's need not be suppressed. Future searches such as those mentioned in the previous section will be particularly important to constraint this subclass of seesaw models.

We have discussed the phenomenological implications of the minimal Type I seesaw model, which will be the hardest to test. The other types of models leading to the Weinberg operator have a richer phenomenology since the extra states couple to gauge fields (e.g the triplet scalar in type II or the fermion in type III), and therefore can be more copiously produced at colliders. In particular lepton number violation could give rise to spectacular signals at LHC, like same-sign lepton resonances in the type II seesaw model [118]:

$$pp \rightarrow H^{++} H^{--} \rightarrow l^+ l^+ l^- l^-. \quad (128)$$

Searches for triplet scalar and fermions are now standard LHC analyses.

12 Leptogenesis

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

$$\eta_B \equiv \frac{N_b - N_{\bar{b}}}{N_\gamma} \sim 6.21(16) \times 10^{-10}. \quad (129)$$

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

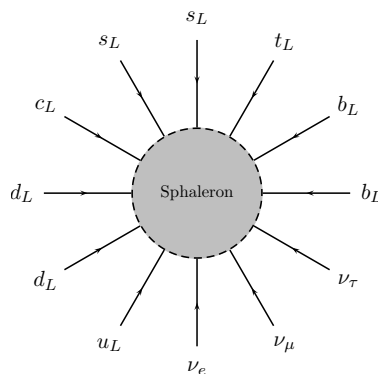


Fig. 51: Artistic view of a sphaleron.

It has been known for a long time that all the ingredients to generate such an asymmetry from a symmetric initial state are present in the laws of particle physics. These ingredients were first put forward by Sakharov [119]:

Baryon number violation

$B + L$ is anomalous in the SM [120] both with and without massive neutrinos. At high T in the early Universe, $B + L$ violating transitions are in thermal equilibrium [121] due to the thermal excitation of configurations with topological charge called sphalerons, see Fig. 51.

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

$$\Delta B = \Delta L. \quad (130)$$

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

C and CP violation

Any lepton or baryon asymmetry can only be generated if there is C and CP violation. Seesaw models generically include new sources of CP violation. As we have seen in type I seesaw model with $n_R = 3$ there are six new CP phases in the lepton sector. They can be absorbed in the Yukawa matrix, Y_ν of eq. (111). For example, in the Casas-Ibarra parametrization, this matrix is written as

$$Y_\nu = U_{\text{PMNS}}^* \sqrt{m_l} R \sqrt{M_h} \frac{\sqrt{2}}{v}. \quad (131)$$

Three phases can be chosen as those in the PMNS matrix, and therefore accessible via neutrino oscillations and neutrinoless double-beta decay. The other three are the parameters of the general complex matrix R , that we cannot access at low-energies. Note that the combination $Y_\nu^\dagger Y_\nu$ only depends on the latter.

Departure from thermal equilibrium

In seesaw models, $B - L$ violating processes can be out-of-equilibrium at $T \gg T_{\text{EW}}$ where the sphalerons are still in thermal equilibrium. In the type I seesaw model two possibilities of non-equilibrium L violation can be realised. In the high scale scenario $M_i \gg v$, the non-equilibrium condition is met at freeze out. The heavy states are thermally produced and freeze out at a temperatures similar to their masses [122]. A net lepton asymmetry can be produced if the decay rate is slower than the expansion of the Universe close to the decoupling temperature, so that the distribution functions of these states differ slightly from the thermal ones, as shown in Fig. 52. It is necessary however that CP

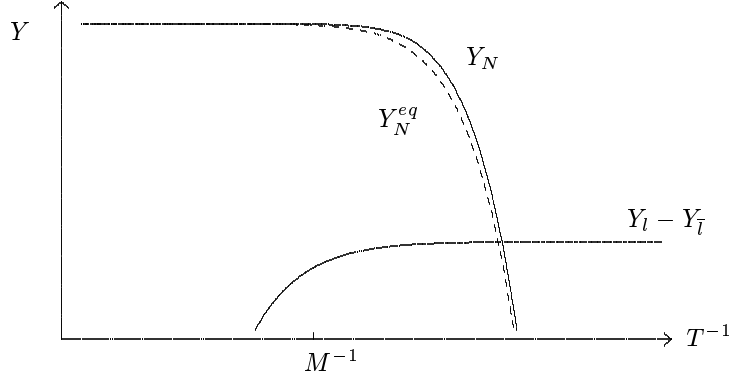


Fig. 52: Abundance of the heavy Majorana singlets at the decoupling temperature and the lepton number generated in the decay.

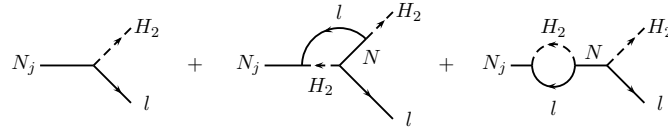


Fig. 53: Tree-level and one-loop diagrams contributing to heavy neutrino decays.

and C be violated in the out-of-equilibrium decays:

$$\epsilon_1 = \frac{\Gamma(N \rightarrow \Phi l) - \Gamma(N \rightarrow \Phi \bar{l})}{\Gamma(N \rightarrow \Phi l) + \Gamma(N \rightarrow \Phi \bar{l})} \neq 0. \quad (132)$$

The new CP phases in the Yukawa matrix induce an asymmetry, ϵ_1 , at the one-loop level (see Fig. 53). The final asymmetry is given by

$$Y_B = 10^{-2} \underbrace{\epsilon_1}_{\text{CP-asym}} \underbrace{\kappa}_{\text{eff. factor}}, \quad (133)$$

where κ is an efficiency factor which depends on the non-equilibrium dynamics. Therefore a relation between the baryon number of the Universe and the neutrino flavour parameters in ϵ_1 exists.

In the low-scale scenario, for $M_i < v$, the out-of-equilibrium condition is met at freeze-in [123] [101, 124]. It is possible that not all the massive states reach thermal equilibrium before T_{EW} . A non-vanishing lepton and baryon asymmetry can survive at T_{EW} and, if this is the case, sphaleron transitions that decouple at this point, can no longer wash it out. It turns out that these conditions can be met naturally in type I seesaw model for masses in the range [0.1, 100] GeV. The relevant CP asymmetries arise in the production of the heavy seesaw states via the interference of CP-odd phases from the Yukawa's with CP-even phases from propagation. A quantum treatment of the corresponding kinetic equations is mandatory in this case.

An interesting question is whether the baryon asymmetry can be predicted quantitatively from the measurements at low energies of the neutrino mass matrix. Unfortunately this is not the case generically, because the asymmetry depends on more parameters than those that are observable at low energies.

For example, in the high-scale scenario, ϵ_1 can be approximated by [125]

$$\epsilon_1 = -\frac{3}{16\pi} \sum_i \frac{\text{Im}[(Y_\nu^\dagger Y_\nu)_{i1}^2]}{(Y_\nu^\dagger Y_\nu)_{11}} \frac{M_1}{M_i}, \quad (134)$$

in the minimal model with $M_{2,3} \gg M_1$. It depends only on the CP phases of R , but not those in U_{PMNS} .

If the prediction of the lepton asymmetry is not possible, it is possible to constrain the neutrino mass matrix, assuming that the lepton asymmetry explains the measured baryon asymmetry. Indeed, various upper bounds can be derived on the generated asymmetry. In particular ϵ_1 has been shown [126] to satisfy

$$|\epsilon_1| \leq \frac{8}{16\pi} \frac{M_1}{v^2} |\Delta m_{\text{atm}}^2|^{1/2}, \quad (135)$$

and therefore leptogenesis in this model requires that the lightest heavy neutrino is rather heavy:

$$M_1 \geq \mathcal{O}(10^9 \text{ GeV}). \quad (136)$$

For further details and references see Ref. [125].

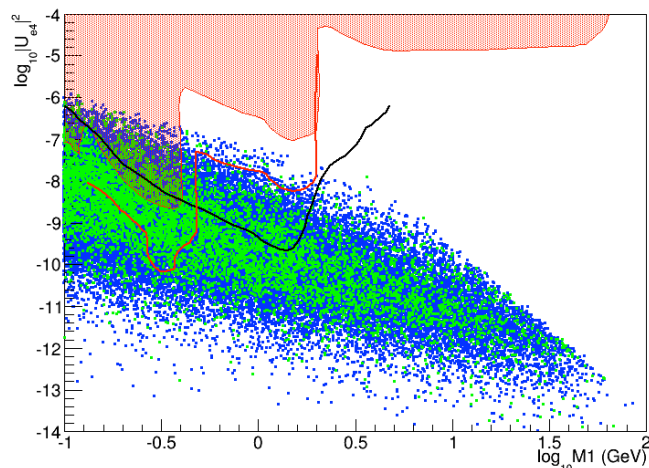


Fig. 54: Points on the plane $|U_{e4}|^2$ vs M_1 for which the baryon asymmetry, Y_B , is in the range $[1/5 - 1] \times Y_B^{\text{exp}}$ (blue) and $[1 - 5] \times Y_B^{\text{exp}}$ (green) for IO (from Ref. [127]). The red band are the present constraints [105], the solid black line shows the reach of the SHiP experiment [106] and the solid red line is the reach of LBNE near detector [107].

Interestingly, in the low-scale scenario, the states responsible for generating the baryon asymmetry might be accessible experimentally. For example, Fig. 54 shows the values of the mixing $|(U_{lh})_{e1}|^2$ and mass M_1 for which the baryon asymmetry can be explained within the type I seesaw model, compared to the sensitivity of future experiments such as SHiP and DUNE.

13 Conclusions

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

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

Many fundamental questions remain to be answered in future neutrino experiments, and these can have very important implications for our understanding of the Standard Model and of what lies beyond:

Are neutrinos Majorana particles? Are neutrino masses the result of a new physics scale? Is CP violated in the lepton sector? Could neutrinos be the seed of the matter–antimatter asymmetry in the Universe?

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

References

- [1] E. Fermi, “Trends to a Theory of beta Radiation. (In Italian),” *Nuovo Cim.*, vol. 11, pp. 1–19, 1934. [535(1934)].
- [2] H. Bethe and R. Peierls, “The ‘neutrino’,” *Nature*, vol. 133, p. 532, 1934.
- [3] B. Pontecorvo, “Inverse beta process,” *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, vol. 1, pp. 25–31, 1991.
- [4] F. Reines and C. L. Cowan, “The neutrino,” *Nature*, vol. 178, pp. 446–449, 1956.
- [5] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, “Detection of the free neutrino: A Confirmation,” *Science*, vol. 124, pp. 103–104, 1956.
- [6] G. Danby, J. M. Gaillard, K. A. Goulianos, L. M. Lederman, N. B. Mistry, M. Schwartz, and J. Steinberger, “Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos,” *Phys. Rev. Lett.*, vol. 9, pp. 36–44, 1962.
- [7] K. A. Olive *et al.*, “Review of Particle Physics,” *Chin. Phys.*, vol. C38, p. 090001, 2014.
- [8] K. Assamagan *et al.*, “Upper limit of the muon-neutrino mass and charged pion mass from momentum analysis of a surface muon beam,” *Phys. Rev.*, vol. D53, pp. 6065–6077, 1996.
- [9] R. Barate *et al.*, “An Upper limit on the tau-neutrino mass from three-prong and five-prong tau decays,” *Eur. Phys. J.*, vol. C2, pp. 395–406, 1998.
- [10] S. Weinberg, “Baryon and Lepton Nonconserving Processes,” *Phys. Rev. Lett.*, vol. 43, pp. 1566–1570, 1979.
- [11] S. Weinberg, “Phenomenological Lagrangians,” *Physica*, vol. A96, pp. 327–340, 1979.
- [12] W. Buchmuller and D. Wyler, “Effective Lagrangian Analysis of New Interactions and Flavor Conservation,” *Nucl. Phys.*, vol. B268, pp. 621–653, 1986.
- [13] Z. Maki, M. Nakagawa, and S. Sakata, “Remarks on the unified model of elementary particles,” *Prog. Theor. Phys.*, vol. 28, pp. 870–880, 1962.
- [14] B. Pontecorvo, “Neutrino Experiments and the Problem of Conservation of Leptonic Charge,” *Sov. Phys. JETP*, vol. 26, pp. 984–988, 1968. [Zh. Eksp. Teor. Fiz.53,1717(1967)].
- [15] B. Pontecorvo, “Mesonium and anti-mesonium,” *Sov. Phys. JETP*, vol. 6, p. 429, 1957. [Zh. Eksp. Teor. Fiz.33,549(1957)].
- [16] E. K. Akhmedov and A. Yu. Smirnov, “Paradoxes of neutrino oscillations,” *Phys. Atom. Nucl.*, vol. 72, pp. 1363–1381, 2009, 0905.1903.
- [17] E. K. Akhmedov and J. Kopp, “Neutrino oscillations: Quantum mechanics vs. quantum field theory,” *JHEP*, vol. 04, p. 008, 2010, 1001.4815. [Erratum: JHEP10,052(2013)].
- [18] M. Cerdá, “,” *Master Thesis University of Valencia*, 2011, 0905.1903.
- [19] L. Wolfenstein, “Neutrino Oscillations in Matter,” *Phys. Rev.*, vol. D17, pp. 2369–2374, 1978.
- [20] S. P. Mikheev and A. Yu. Smirnov, “Resonance Amplification of Oscillations in Matter and Spectroscopy of Solar Neutrinos,” *Sov. J. Nucl. Phys.*, vol. 42, pp. 913–917, 1985. [Yad. Fiz.42,1441(1985)].
- [21] H. A. Bethe, “Energy production in stars,” *Phys. Rev.*, vol. 55, pp. 434–456, 1939.
- [22] J. N. Bahcall, M. H. Pinsonneault, and S. Basu, “Solar models: Current epoch and time dependences, neutrinos, and helioseismological properties,” *Astrophys. J.*, vol. 555, pp. 990–1012, 2001,

- astro-ph/0010346.
- [23] B. T. Cleveland, T. Daily, R. Davis, Jr., J. R. Distel, K. Lande, C. K. Lee, P. S. Wildenhain, and J. Ullman, “Measurement of the solar electron neutrino flux with the Homestake chlorine detector,” *Astrophys. J.*, vol. 496, pp. 505–526, 1998.
 - [24] W. Hampel *et al.*, “GALLEX solar neutrino observations: Results for GALLEX IV,” *Phys. Lett.*, vol. B447, pp. 127–133, 1999.
 - [25] 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.*, vol. 95, pp. 181–193, 2002, astro-ph/0204245. [*Zh. Eksp. Teor. Fiz.*122,211(2002)].
 - [26] Y. Fukuda *et al.*, “Solar neutrino data covering solar cycle 22,” *Phys. Rev. Lett.*, vol. 77, pp. 1683–1686, 1996.
 - [27] Y. Fukuda *et al.*, “Measurement of the solar neutrino energy spectrum using neutrino electron scattering,” *Phys. Rev. Lett.*, vol. 82, pp. 2430–2434, 1999, hep-ex/9812011.
 - [28] J. Hosaka *et al.*, “Solar neutrino measurements in super-Kamiokande-I,” *Phys. Rev.*, vol. D73, p. 112001, 2006, hep-ex/0508053.
 - [29] Q. R. Ahmad *et al.*, “Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by 8B solar neutrinos at the Sudbury Neutrino Observatory,” *Phys. Rev. Lett.*, vol. 87, p. 071301, 2001, nucl-ex/0106015.
 - [30] Q. R. Ahmad *et al.*, “Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory,” *Phys. Rev. Lett.*, vol. 89, p. 011301, 2002, nucl-ex/0204008.
 - [31] K. Eguchi *et al.*, “First results from KamLAND: Evidence for reactor anti-neutrino disappearance,” *Phys. Rev. Lett.*, vol. 90, p. 021802, 2003, hep-ex/0212021.
 - [32] S. Abe *et al.*, “Precision Measurement of Neutrino Oscillation Parameters with KamLAND,” *Phys. Rev. Lett.*, vol. 100, p. 221803, 2008, 0801.4589.
 - [33] G. Bellini *et al.*, “Final results of Borexino Phase-I on low energy solar neutrino spectroscopy,” *Phys. Rev.*, vol. D89, no. 11, p. 112007, 2014, 1308.0443.
 - [34] M. Honda, T. Kajita, K. Kasahara, and S. Midorikawa, “A New calculation of the atmospheric neutrino flux in a 3-dimensional scheme,” *Phys. Rev.*, vol. D70, p. 043008, 2004, astro-ph/0404457.
 - [35] Y. Fukuda *et al.*, “Evidence for oscillation of atmospheric neutrinos,” *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, 1998, hep-ex/9807003.
 - [36] Y. Ashie *et al.*, “A Measurement of atmospheric neutrino oscillation parameters by SUPER-KAMIOKANDE I,” *Phys. Rev.*, vol. D71, p. 112005, 2005, hep-ex/0501064.
 - [37] Y. Ashie *et al.*, “Evidence for an oscillatory signature in atmospheric neutrino oscillation,” *Phys. Rev. Lett.*, vol. 93, p. 101801, 2004, hep-ex/0404034.
 - [38] A. Holin, “Results from the MINOS Experiment and New MINOS+ Data,” *PoS*, vol. NU-FACT2014, p. 028, 2014, 1507.08564.
 - [39] F. P. An *et al.*, “Observation of electron-antineutrino disappearance at Daya Bay,” *Phys. Rev. Lett.*, vol. 108, p. 171803, 2012, 1203.1669.
 - [40] J. K. Ahn *et al.*, “Observation of Reactor Electron Antineutrino Disappearance in the RENO Experiment,” *Phys. Rev. Lett.*, vol. 108, p. 191802, 2012, 1204.0626.
 - [41] Y. Abe *et al.*, “Indication for the disappearance of reactor electron antineutrinos in the Double Chooz experiment,” *Phys. Rev. Lett.*, vol. 108, p. 131801, 2012, 1112.6353.
 - [42] K. Abe *et al.*, “Observation of Electron Neutrino Appearance in a Muon Neutrino Beam,” *Phys. Rev. Lett.*, vol. 112, p. 061802, 2014, 1311.4750.
 - [43] M. C. Gonzalez-Garcia, M. Maltoni, and T. Schwetz, “Updated fit to three neutrino mixing: status of leptonic CP violation,” *JHEP*, vol. 11, p. 052, 2014, 1409.5439.

- [44] P. A. R. Ade *et al.*, “Planck 2015 results. XIII. Cosmological parameters,” *arXiv:1502.01589[astro-ph.CO]*, 2015, 1502.01589.
- [45] S. K. Agarwalla and P. Hernandez, “Probing the Neutrino Mass Hierarchy with Super-Kamiokande,” *JHEP*, vol. 10, p. 086, 2012, 1204.4217.
- [46] S. T. Petcov and M. Piai, “The LMA MSW solution of the solar neutrino problem, inverted neutrino mass hierarchy and reactor neutrino experiments,” *Phys. Lett.*, vol. B533, pp. 94–106, 2002, hep-ph/0112074.
- [47] S. Choubey, S. T. Petcov, and M. Piai, “Precision neutrino oscillation physics with an intermediate baseline reactor neutrino experiment,” *Phys. Rev.*, vol. D68, p. 113006, 2003, hep-ph/0306017.
- [48] F. An *et al.*, “Neutrino Physics with JUNO,” *J. Phys.*, vol. G43, no. 3, p. 030401, 2016, 1507.05613.
- [49] A. Cervera, A. Donini, M. B. Gavela, J. J. Gomez Cadenas, P. Hernandez, O. Mena, and S. Rigolin, “Golden measurements at a neutrino factory,” *Nucl. Phys.*, vol. B579, pp. 17–55, 2000, hep-ph/0002108. [Erratum: *Nucl. Phys.*B593,731(2001)].
- [50] K. Abe *et al.*, “Physics potential of a long-baseline neutrino oscillation experiment using a J-PARC neutrino beam and Hyper-Kamiokande,” *PTEP*, vol. 2015, p. 053C02, 2015, 1502.05199.
- [51] R. Acciarri *et al.*, “Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE) Conceptual Design Report Volume 2: The Physics Program for DUNE at LBNF,” *arXiv:1512.06148 [hep-ex]*, 2015, 1512.06148.
- [52] A. Aguilar-Arevalo *et al.*, “Evidence for neutrino oscillations from the observation of anti-neutrino(electron) appearance in a anti-neutrino(muon) beam,” *Phys. Rev.*, vol. D64, p. 112007, 2001, hep-ex/0104049.
- [53] B. Armbruster *et al.*, “Upper limits for neutrino oscillations muon-anti-neutrino \rightarrow electron-anti-neutrino from muon decay at rest,” *Phys. Rev.*, vol. D65, p. 112001, 2002, hep-ex/0203021.
- [54] A. A. Aguilar-Arevalo *et al.*, “Improved Search for $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ Oscillations in the MiniBooNE Experiment,” *Phys. Rev. Lett.*, vol. 110, p. 161801, 2013, 1207.4809.
- [55] T. A. Mueller *et al.*, “Improved Predictions of Reactor Antineutrino Spectra,” *Phys. Rev.*, vol. C83, p. 054615, 2011, 1101.2663.
- [56] G. Mention, M. Fechner, T. Lasserre, T. A. Mueller, D. Lhuillier, M. Cribier, and A. Letourneau, “The Reactor Antineutrino Anomaly,” *Phys. Rev.*, vol. D83, p. 073006, 2011, 1101.2755.
- [57] P. Huber, “On the determination of anti-neutrino spectra from nuclear reactors,” *Phys. Rev.*, vol. C84, p. 024617, 2011, 1106.0687. [Erratum: *Phys. Rev.*C85,029901(2012)].
- [58] F. P. An *et al.*, “Measurement of the Reactor Antineutrino Flux and Spectrum at Daya Bay,” *Phys. Rev. Lett.*, vol. 116, no. 6, p. 061801, 2016, 1508.04233.
- [59] A. B. Sousa, “First MINOS+ Data and New Results from MINOS,” *AIP Conf. Proc.*, vol. 1666, p. 110004, 2015, 1502.07715.
- [60] P. F. Harrison, D. H. Perkins, and W. G. Scott, “Tri-bimaximal mixing and the neutrino oscillation data,” *Phys. Lett.*, vol. B530, p. 167, 2002, hep-ph/0202074.
- [61] S. F. King and C. Luhn, “Neutrino Mass and Mixing with Discrete Symmetry,” *Rept. Prog. Phys.*, vol. 76, p. 056201, 2013, 1301.1340.
- [62] P. Minkowski, “ $\mu \rightarrow e\gamma$ at a Rate of One Out of 10^9 Muon Decays?,” *Phys. Lett.*, vol. B67, pp. 421–428, 1977.
- [63] M. Gell-Mann, P. Ramond, and R. Slansky, “Complex Spinors and Unified Theories,” *Conf. Proc.*, vol. C790927, pp. 315–321, 1979, 1306.4669.
- [64] T. Yanagida, “HORIZONTAL SYMMETRY AND MASSES OF NEUTRINOS,” *Conf. Proc.*, vol. C7902131, pp. 95–99, 1979.
- [65] R. N. Mohapatra and G. Senjanovic, “Neutrino Mass and Spontaneous Parity Violation,” *Phys.*

- Rev. Lett.*, vol. 44, p. 912, 1980.
- [66] M. Magg and C. Wetterich, “Neutrino Mass Problem and Gauge Hierarchy,” *Phys. Lett.*, vol. B94, p. 61, 1980.
- [67] J. Schechter and J. W. F. Valle, “Neutrino Masses in SU(2) x U(1) Theories,” *Phys. Rev.*, vol. D22, p. 2227, 1980.
- [68] C. Wetterich, “Neutrino Masses and the Scale of B-L Violation,” *Nucl. Phys.*, vol. B187, p. 343, 1981.
- [69] G. Lazarides, Q. Shafi, and C. Wetterich, “Proton Lifetime and Fermion Masses in an SO(10) Model,” *Nucl. Phys.*, vol. B181, pp. 287–300, 1981.
- [70] R. N. Mohapatra and G. Senjanovic, “Neutrino Masses and Mixings in Gauge Models with Spontaneous Parity Violation,” *Phys. Rev.*, vol. D23, p. 165, 1981.
- [71] R. Foot, H. Lew, X. G. He, and G. C. Joshi, “Seesaw Neutrino Masses Induced by a Triplet of Leptons,” *Z. Phys.*, vol. C44, p. 441, 1989.
- [72] E. Ma, “Pathways to naturally small neutrino masses,” *Phys. Rev. Lett.*, vol. 81, pp. 1171–1174, 1998, hep-ph/9805219.
- [73] A. Zee, “A Theory of Lepton Number Violation, Neutrino Majorana Mass, and Oscillation,” *Phys. Lett.*, vol. B93, p. 389, 1980. [Erratum: *Phys. Lett.* B95, 461 (1980)].
- [74] A. Zee, “Charged Scalar Field and Quantum Number Violations,” *Phys. Lett.*, vol. B161, p. 141, 1985.
- [75] K. S. Babu, “Model of ‘Calculable’ Majorana Neutrino Masses,” *Phys. Lett.*, vol. B203, p. 132, 1988.
- [76] A. Abada, C. Biggio, F. Bonnet, M. B. Gavela, and T. Hambye, “Low energy effects of neutrino masses,” *JHEP*, vol. 12, p. 061, 2007, 0707.4058.
- [77] G. Aad *et al.*, “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments,” *Phys. Rev. Lett.*, vol. 114, p. 191803, 2015, 1503.07589.
- [78] G. Degrossi, S. Di Vita, J. Elias-Miro, J. R. Espinosa, G. F. Giudice, G. Isidori, and A. Strumia, “Higgs mass and vacuum stability in the Standard Model at NNLO,” *JHEP*, vol. 08, p. 098, 2012, 1205.6497.
- [79] F. Vissani, “Do experiments suggest a hierarchy problem?,” *Phys. Rev.*, vol. D57, pp. 7027–7030, 1998, hep-ph/9709409.
- [80] J. A. Casas, J. R. Espinosa, and I. Hidalgo, “Implications for new physics from fine-tuning arguments. 1. Application to SUSY and seesaw cases,” *JHEP*, vol. 11, p. 057, 2004, hep-ph/0410298.
- [81] J. A. Casas and A. Ibarra, “Oscillating neutrinos and $\mu \rightarrow e, \gamma$,” *Nucl. Phys.*, vol. B618, pp. 171–204, 2001, hep-ph/0103065.
- [82] A. Donini, P. Hernandez, J. Lopez-Pavon, M. Maltoni, and T. Schwetz, “The minimal 3+2 neutrino model versus oscillation anomalies,” *JHEP*, vol. 07, p. 161, 2012, 1205.5230.
- [83] M. Blennow, E. Fernandez-Martinez, J. Lopez-Pavon, and J. Menendez, “Neutrinoless double beta decay in seesaw models,” *JHEP*, vol. 07, p. 096, 2010, 1005.3240.
- [84] F. Capozzi, E. Lisi, A. Marrone, D. Montanino, and A. Palazzo, “Neutrino masses and mixings: Status of known and unknown 3ν parameters,” *arXiv:1601.07777 [hep-ph]*, 2016, 1601.07777.
- [85] A. Ibarra, E. Molinaro, and S. T. Petcov, “TeV Scale See-Saw Mechanisms of Neutrino Mass Generation, the Majorana Nature of the Heavy Singlet Neutrinos and $(\beta\beta)_{0\nu}$ -Decay,” *JHEP*, vol. 09, p. 108, 2010, 1007.2378.
- [86] M. Mitra, G. Senjanovic, and F. Vissani, “Neutrinoless Double Beta Decay and Heavy Sterile Neutrinos,” *Nucl. Phys.*, vol. B856, pp. 26–73, 2012, 1108.0004.
- [87] J. Lopez-Pavon, E. Molinaro, and S. T. Petcov, “Radiative Corrections to Light Neutrino Masses

- in Low Scale Type I Seesaw Scenarios and Neutrinoless Double Beta Decay,” *JHEP*, vol. 11, p. 030, 2015, 1506.05296.
- [88] A. de Gouvea, “See-saw energy scale and the LSND anomaly,” *Phys. Rev.*, vol. D72, p. 033005, 2005, hep-ph/0501039.
- [89] A. de Gouvea, W.-C. Huang, and J. Jenkins, “Pseudo-Dirac Neutrinos in the New Standard Model,” *Phys.Rev.*, vol. D80, p. 073007, 2009, 0906.1611.
- [90] A. de Gouvea and W.-C. Huang, “Constraining the (Low-Energy) Type-I Seesaw,” *Phys.Rev.*, vol. D85, p. 053006, 2012, 1110.6122.
- [91] A. Donini, P. Hernandez, J. Lopez-Pavon, and M. Maltoni, “Minimal models with light sterile neutrinos,” *JHEP*, vol. 07, p. 105, 2011, 1106.0064.
- [92] C. Giunti and M. Laveder, “3+1 and 3+2 Sterile Neutrino Fits,” *Phys. Rev.*, vol. D84, p. 073008, 2011, 1107.1452.
- [93] J. Kopp, P. A. N. Machado, M. Maltoni, and T. Schwetz, “Sterile Neutrino Oscillations: The Global Picture,” *JHEP*, vol. 05, p. 050, 2013, 1303.3011.
- [94] D. Notzold and G. Raffelt, “Neutrino Dispersion at Finite Temperature and Density,” *Nucl. Phys.*, vol. B307, p. 924, 1988.
- [95] R. Barbieri and A. Dolgov, “Bounds on Sterile-neutrinos from Nucleosynthesis,” *Phys. Lett.*, vol. B237, p. 440, 1990.
- [96] K. Kainulainen, “Light Singlet Neutrinos and the Primordial Nucleosynthesis,” *Phys. Lett.*, vol. B244, pp. 191–195, 1990.
- [97] P. Hernandez, M. Kekic, and J. Lopez-Pavon, “Low-scale seesaw models versus N_{eff} ,” *Phys.Rev.*, vol. D89, no. 7, p. 073009, 2014, 1311.2614.
- [98] P. Hernandez, M. Kekic, and J. Lopez-Pavon, “ N_{eff} in low-scale seesaw models versus the lightest neutrino mass,” *Phys.Rev.*, vol. D90, no. 6, p. 065033, 2014, 1406.2961.
- [99] S. Hannestad, R. S. Hansen, T. Tram, and Y. Y. Y. Wong, “Active-sterile neutrino oscillations in the early Universe with full collision terms,” *JCAP*, vol. 1508, no. 08, p. 019, 2015, 1506.05266.
- [100] S. Dodelson and L. M. Widrow, “Sterile-neutrinos as dark matter,” *Phys. Rev. Lett.*, vol. 72, pp. 17–20, 1994, hep-ph/9303287.
- [101] T. Asaka and M. Shaposhnikov, “The nuMSM, dark matter and baryon asymmetry of the universe,” *Phys.Lett.*, vol. B620, pp. 17–26, 2005, hep-ph/0505013.
- [102] E. Bulbul, M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein, and S. W. Randall, “Detection of An Unidentified Emission Line in the Stacked X-ray spectrum of Galaxy Clusters,” *Astrophys. J.*, vol. 789, p. 13, 2014, 1402.2301.
- [103] A. Boyarsky, O. Ruchayskiy, D. Iakubovskiy, and J. Franse, “Unidentified Line in X-Ray Spectra of the Andromeda Galaxy and Perseus Galaxy Cluster,” *Phys. Rev. Lett.*, vol. 113, p. 251301, 2014, 1402.4119.
- [104] X.-D. Shi and G. M. Fuller, “A New dark matter candidate: Nonthermal sterile neutrinos,” *Phys. Rev. Lett.*, vol. 82, pp. 2832–2835, 1999, astro-ph/9810076.
- [105] A. Atre, T. Han, S. Pascoli, and B. Zhang, “The Search for Heavy Majorana Neutrinos,” *JHEP*, vol. 05, p. 030, 2009, 0901.3589.
- [106] S. Alekhin *et al.*, “A facility to Search for Hidden Particles at the CERN SPS: the SHiP physics case,” *arXiv:1504.04855[hep-ph]*, 2015, 1504.04855.
- [107] C. Adams *et al.*, “The Long-Baseline Neutrino Experiment: Exploring Fundamental Symmetries of the Universe,” 2013, 1307.7335.
- [108] F. del Aguila and J. A. Aguilar-Saavedra, “Distinguishing seesaw models at LHC with multi-lepton signals,” *Nucl. Phys.*, vol. B813, pp. 22–90, 2009, 0808.2468.
- [109] P. S. B. Dev, A. Pilaftsis, and U.-k. Yang, “New Production Mechanism for Heavy Neutrinos at

- the LHC,” *Phys. Rev. Lett.*, vol. 112, no. 8, p. 081801, 2014, 1308.2209.
- [110] J. C. Helo, M. Hirsch, and S. Kovalenko, “Heavy neutrino searches at the LHC with displaced vertices,” *Phys. Rev.*, vol. D89, p. 073005, 2014, 1312.2900.
- [111] E. Izaguirre and B. Shuve, “Multilepton and Lepton Jet Probes of Sub-Weak-Scale Right-Handed Neutrinos,” *Phys. Rev.*, vol. D91, no. 9, p. 093010, 2015, 1504.02470.
- [112] A. M. Gago, P. Hernandez, J. Jones-Perez, M. Losada, and A. Moreno Briceño, “Probing the Type I Seesaw Mechanism with Displaced Vertices at the LHC,” *Eur. Phys. J.*, vol. C75, no. 10, p. 470, 2015, 1505.05880.
- [113] G. C. Branco, W. Grimus, and L. Lavoura, “The Seesaw Mechanism in the Presence of a Conserved Lepton Number,” *Nucl. Phys.*, vol. B312, p. 492, 1989.
- [114] J. Kersten and A. Yu. Smirnov, “Right-Handed Neutrinos at CERN LHC and the Mechanism of Neutrino Mass Generation,” *Phys. Rev.*, vol. D76, p. 073005, 2007, 0705.3221.
- [115] M. B. Gavela, T. Hambye, D. Hernandez, and P. Hernandez, “Minimal Flavour Seesaw Models,” *JHEP*, vol. 09, p. 038, 2009, 0906.1461.
- [116] D. Wyler and L. Wolfenstein, “Massless Neutrinos in Left-Right Symmetric Models,” *Nucl. Phys.*, vol. B218, p. 205, 1983.
- [117] R. N. Mohapatra and J. W. F. Valle, “Neutrino Mass and Baryon Number Nonconservation in Superstring Models,” *Phys. Rev.*, vol. D34, p. 1642, 1986.
- [118] W.-Y. Keung and G. Senjanovic, “Majorana Neutrinos and the Production of the Right-handed Charged Gauge Boson,” *Phys. Rev. Lett.*, vol. 50, p. 1427, 1983.
- [119] A. D. Sakharov, “Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe,” *Pisma Zh. Eksp. Teor. Fiz.*, vol. 5, pp. 32–35, 1967. [Usp. Fiz. Nauk161,61(1991)].
- [120] G. ’t Hooft, “Symmetry Breaking Through Bell-Jackiw Anomalies,” *Phys. Rev. Lett.*, vol. 37, pp. 8–11, 1976.
- [121] V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov, “On the Anomalous Electroweak Baryon Number Nonconservation in the Early Universe,” *Phys. Lett.*, vol. B155, p. 36, 1985.
- [122] M. Fukugita and T. Yanagida, “Baryogenesis Without Grand Unification,” *Phys. Lett.*, vol. B174, p. 45, 1986.
- [123] E. K. Akhmedov, V. Rubakov, and A. Y. Smirnov, “Baryogenesis via neutrino oscillations,” *Phys.Rev.Lett.*, vol. 81, pp. 1359–1362, 1998, hep-ph/9803255.
- [124] L. Canetti, M. Drewes, T. Frossard, and M. Shaposhnikov, “Dark Matter, Baryogenesis and Neutrino Oscillations from Right Handed Neutrinos,” *Phys. Rev.*, vol. D87, p. 093006, 2013, 1208.4607.
- [125] S. Davidson, E. Nardi, and Y. Nir, “Leptogenesis,” *Phys. Rept.*, vol. 466, pp. 105–177, 2008, 0802.2962.
- [126] S. Davidson and A. Ibarra, “A Lower bound on the right-handed neutrino mass from leptogenesis,” *Phys. Lett.*, vol. B535, pp. 25–32, 2002, hep-ph/0202239.
- [127] P. Hernandez, M. Kekic, J. Lopez-Pavon, J. Racker, and N. Rius, “Leptogenesis in GeV scale seesaw models,” *JHEP*, vol. 10, p. 067, 2015, 1508.03676.

Higgs Physics

C. Grojean

DESY, 22607 Hamburg, Germany

Institut für Physik, Humboldt-Universität zu Berlin, 12489 Berlin, Germany

ICREA, 08010 Barcelona, Spain (on leave)

IFAE, BIST, 08193 Bellaterra, Barcelona, Spain (on leave)

Abstract

The cause of the screening of the weak interactions at long distances puzzled the high-energy community for nearly half a century. With the discovery of the Higgs boson a new era started with direct experimental information on the physics behind the breaking of the electroweak symmetry. This breaking plays a fundamental role in our understanding of particle physics and sits at the high-energy frontier beyond which we expect new physics that supersedes the Standard Model. The Higgs boson (inclusive and differential) production and decay rates offer a new way to probe this frontier.

Keywords

Higgs particle; large hadron collider; lectures; standard model; electroweak symmetry breaking; new physics.

Exhaustive reviews on electroweak symmetry breaking (EWSB) and Higgs physics are easily accessible online [1–7] and the purpose of these notes is not to duplicate them but rather to serve as a concise introduction to the topic and to present (a personalized hence biased selection of) recent developments in the field.

1 Introduction

The Standard Model (SM) is a triumph of the combination of the two pillars of twentieth-century physics, namely quantum mechanics and special relativity. Particles are defined as representations of the Poincaré group. Mathematically, these representations are labelled by two quantities: the spin that is quantized and takes only discrete values, and the mass, which a priori is a continuous parameter. However, the transformation laws for the various elementary particles under the gauge symmetries associated to the fundamental interactions force the masses of these particles to vanish. This would be in flagrant contradiction with the experimental measurements.

The Brout–Englert–Higgs mechanism (BEH) [8–11] provides the solution to this mass conundrum. The discovery of a Higgs boson in July 2012 and the experimental confirmation of the BEH mechanism by the ATLAS and CMS collaborations [12, 13] has been a historical step in our understanding of nature: the masses of the elementary particles are not fundamental parameters defined at very high energy but rather emergent quantities appearing at low energy as a result of the particular structure of the vacuum.

2 The HEP landscape after the Higgs discovery

During its first run, the LHC certainly fulfilled its commitments: The machine and its detectors were mostly designed to find the Higgs boson and “[they] got it!” according to the words of R. Heuer, director general of CERN, on 4 July 2012. It was an important step in the understanding of the mechanism of electroweak symmetry breaking. But the journey is not over.

One can ask how the Higgs discovery reshaped the High Energy Physics (HEP) landscape. The days of theoretically guaranteed discoveries imposed on us by some no-lose theorems are over: indeed, with the addition of a light Higgs boson with a mass around 125 GeV, the Standard Model is theoretically consistent and can be extrapolated up to very high energy, maybe as high as $10^{14\div 16}$ GeV or even the Planck scale. But at the same time, the big questions of our field, or the ones that we have considered so far as the big questions, remain wildly open: the hierarchy problem, the origin of flavor, the issue of the neutrino masses and mixings, the question of the identity of Dark Matter, the source of dynamical preponderance of matter over antimatter during the cosmological evolution of our Universe... are left unanswered (see the BSM lectures [14] in these proceedings). In the next decades, future progress in HEP is in the hands of experimentalists whose discoveries will reveal the way Nature has solved these big questions, forcing the theorists to renounce/review/question deeply rooted bias/prejudice. The Higgs discovery sets a large part of the agenda for the theoretical and experimental HEP programs over the next couple of decades.

3 Open questions about the Higgs

The LHC accumulated striking evidence that the Higgs vacuum expectation value (vev) is the cause of the screening of the weak interaction at long distances and the source of the gauge boson masses.

However, this evidence only addresses the question of *how* the symmetry of the weak interaction is broken. It does not address the question of *why* the symmetry is broken or why the Higgs field acquires an expectation value. The situation is simply summarized in the following tautology

Why is electroweak symmetry broken?
Because the Higgs potential is unstable at the origin.
 Why is the Higgs potential unstable at the origin?
Because otherwise EW symmetry would not be broken.

The discovery of a Higgs boson allowed first glimpses into a new sector of the microscopic world. Now comes the time of the detailed exploration of this new Higgs sector. And some key questions about the Higgs boson emerge:

1. Is it the SM Higgs?
2. Is it an elementary or a composite particle?
3. Is it unique and solitary? Or are there additional states populating the Higgs sector?
4. Is it eternal or only temporarily living in a metastable vacuum?
5. Is its mass natural following the criteria of Dirac, Wilson or 't Hooft?
6. Is it the first superparticle ever observed?
7. Is it really responsible for the masses of all the elementary particles?
8. Is it mainly produced by top quarks or by new heavy vector-like particles?
9. Is it a portal to a hidden world forming the dark matter component of the Universe?
10. Is it at the origin of the matter-antimatter asymmetry?
11. Has it driven the primordial inflationary expansion of the Universe?

The answers to these questions will have profound implications on our understanding of the fundamental laws of physics. Establishing that the Higgs boson is weakly coupled, elementary and solitary, would surely be as shocking as unexpected, but it may well indicate the existence of a multiverse ruled by anthropic selection rules. If instead deviations from the SM emerge in the dynamics of the Higgs, we will have to use them as a diagnostic tool of the underlying dynamics. The pattern of these deviations will carry indirect information about the nature of the completion of the SM at higher energies. In supersymmetric models, and more generally in models with an extended electroweak symmetry breaking

sector, the largest deviations will be observed in the couplings to leptons and to the down-type quarks, as well as in the decay amplitudes to photons and gluons. In models of strong interactions, in which the Higgs boson is a bound state, the effects of compositeness uniformly suppress all the Higgs couplings while the self-interactions of the particles inside the Higgs sector, namely the Higgs particle and the longitudinal components of the W and Z bosons, will increase with the transferred energy. Moreover, the measurements of the Higgs couplings will also reveal the symmetry properties of the ‘‘Higgs boson’’ observed. For instance, it can be established whether the new scalar is indeed ‘‘a Higgs’’ fitting into a SU(2) doublet together with the degrees of freedom associated with the longitudinal W and Z and not some exotic impostor, like for instance a pseudo-dilaton. If the Higgs is found to have an internal structure, a detailed study of the Higgs couplings can also establish whether it is just an ordinary composite, like a σ particle, or whether it is a pseudo-Nambu–Goldstone boson endowed with additional symmetry properties, like the π ’s of QCD.

4 What is the SM Higgs the name of?

4.1 The SM Higgs boson as a UV regulator

The SM Higgs boson ensures the proper decoupling of the longitudinal polarizations of the massive EW gauge bosons at high energy. Indeed, these longitudinal modes of W^\pm and Z can be described by Nambu–Goldstone bosons associated to the coset $SU(2)_L \times SU(2)_R/SU(2)_{\text{isospin}}$. Their kinetic term corresponds to the gauge boson mass terms,

$$\frac{1}{2}m_Z^2 Z_\mu Z^\mu + m_W^2 W_\mu^+ W^{-\mu} = \frac{v^2}{4} \text{Tr}(D_\mu \Sigma^\dagger D^\mu \Sigma) \quad (1)$$

with $\Sigma = e^{i\sigma^a \pi^a/v}$, where σ^a ($a = 1, 2, 3$) are the usual Pauli matrices. Due to the Goldstone boson equivalence theorem the non-trivial scattering of the longitudinal gauge bosons V ($V = W^\pm, Z$) is controlled by the contact interactions among four pions from the expansion of the Lagrangian of Eq. (1), leading to amplitudes growing with the energy,

$$\mathcal{A}(V_L^a V_L^b \rightarrow V_L^c V_L^d) = \mathcal{A}(s)\delta^{ab}\delta^{cd} + \mathcal{A}(t)\delta^{ac}\delta^{bd} + \mathcal{A}(u)\delta^{ad}\delta^{bc} \quad \text{with} \quad \mathcal{A}(s) \approx \frac{s}{v^2}. \quad (2)$$

Here s, t, u denote the Mandelstam variables, and v represents the ‘‘Higgs vev’’, with $v = 246$ GeV. The amplitude grows with the center-of-mass (c.m.) energy squared s , and therefore perturbative unitarity will be lost around $4\pi v \sim 3$ TeV, unless there is a new weakly coupled elementary degree of freedom. The simplest realization of new dynamics restoring perturbative unitarity is given by a single scalar field h , which is singlet under $SU(2)_L \times SU(2)_R/SU(2)_{\text{isospin}}$ and couples to the longitudinal gauge bosons and fermions as [15, 16],

$$\begin{aligned} \mathcal{L}_{\text{EWSB}} = & \frac{1}{2}(\partial_\mu h)^2 - V(h) + \frac{v^2}{4} \text{Tr}(D_\mu \Sigma^\dagger D^\mu \Sigma) \left(1 + 2a\frac{h}{v} + \sum_{n \geq 2} b_n \frac{h^n}{v^n} + \dots \right) \\ & - \frac{v}{\sqrt{2}}(\bar{u}_L^i \bar{d}_L^i) \Sigma \left(1 + c\frac{h}{v} + \sum_{n \geq 2} c_n \frac{h^n}{v^n} + \dots \right) \begin{pmatrix} y_{ij}^u u_R^j \\ y_{ij}^d d_R^j \end{pmatrix} + \text{h.c.} \end{aligned}$$

with

$$V(h) = \frac{1}{2}m_h^2 h^2 + \frac{d_3}{6} \left(\frac{3m_h^2}{v} \right) h^3 + \frac{d_4}{24} \left(\frac{3m_h^2}{v^2} \right) h^4 + \dots \quad (3)$$

For $a = 1$ the scalar exchange cancels the piece growing with the energy in the $V_L V_L$ amplitude. If in addition $b_2 = a^2$ then also in the inelastic amplitude $V_L V_L \rightarrow hh$ perturbative unitarity is maintained,

while for $ac = 1$ the $V_L V_L \rightarrow f f'$ amplitude also remains finite. The SM Higgs boson is defined by the point $a = b_2 = c = 1$ and $d_3 = d_4 = 1$, $c_{n \geq 2} = b_{n \geq 3} = 0$. The scalar resonance and the pions then combine to form a doublet which transforms linearly under $SU(2)_L \times SU(2)_R$.

The requirement that the Higgs boson fully unitarizes the scattering amplitudes of massive particles therefore implies that the Higgs couplings to the various SM particles are directly proportional to their masses. This fundamental property is in remarkable agreement with the state-of-the-art fit of the current Higgs data collected at the LHC, see Fig. 1.

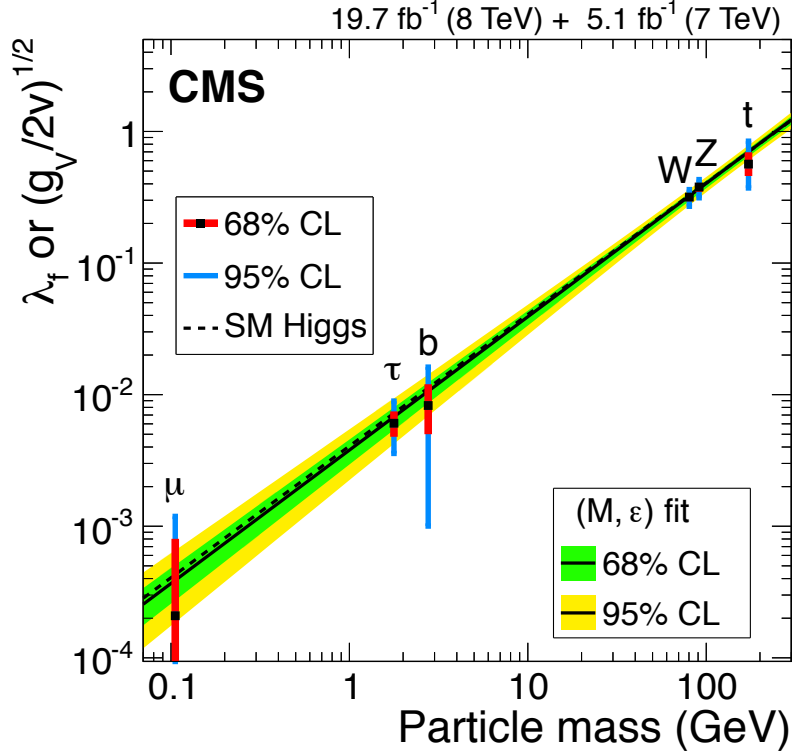


Fig. 1: Comparison of the SM predictions (black dashed line) and the fit of the LHC measurements of the Higgs couplings to various SM particles. From Ref. [17].

The couplings to the heaviest particles, namely W and Z bosons, the top quark and the τ lepton, are already established. The measurement of the couplings to other quarks and leptons, in particular the lightest ones, will require considerably more statistics at the LHC. Nonetheless, it is already established that the Higgs boson, contrary to all the gauge bosons, has non-universal couplings among the particles of the three different generations of quarks and leptons. The Higgs particle is not a Z' gauge boson! The Higgs boson mediates new fundamental forces different in nature than the electromagnetic, weak and strong interactions. Are other forces of this type going to be discovered? Models of DM and baryogenesis make use of new forces like the ones mediated by the Higgs boson.

4.2 The flavor preserving nature of the SM Higgs

In the SM, the Yukawa interactions are the only source of the fermion masses and they also generate linear interaction with the physical Higgs boson

$$Y_{ij} \bar{\psi}_i H \psi_j = \frac{Y_{ij} v}{\sqrt{2}} \bar{\psi}_i \psi_j + \frac{Y_{ij}}{\sqrt{2}} h \bar{\psi}_i \psi_j. \quad (4)$$

Clearly both matrices can be diagonalized simultaneously and this ensures the absence of flavor changing neutral currents induced by the Higgs boson exchange.

This nice property is no longer true if the SM fermions mix with vector-like partners or in the presence of generic higher dimension Yukawa interactions (see for instance Ref. [18] for a general phenomenological discussion):

$$Y_{ij} \left(1 + \frac{c_{ij}}{f^2} |H|^2 \right) \bar{\psi}_i H \psi_j = \frac{Y_{ij} v}{\sqrt{2}} \left(1 + \frac{c_{ij} v^2}{2f^2} \right) \bar{\psi}_i \psi_j + \frac{Y_{ij}}{\sqrt{2}} \left(1 + \frac{3c_{ij} v^2}{2f^2} \right) h \bar{\psi}_i \psi_j. \quad (5)$$

Therefore it is particularly important to probe the flavor structure of the Higgs interactions and to look for flavor-violating decays, e.g. $h \rightarrow \mu\tau$, or production, e.g. $t \rightarrow hc$. Limits from low-energy flavor-changing interactions indirectly constrain these processes especially in the quark sector but leave the possibility of sizeable effects in the lepton sector (see for instance Ref. [19] for an extensive discussion). The slight 2.5σ excess in the $h \rightarrow \mu\tau$ decay initially reported by CMS with the full run 1 dataset [20] is confirmed neither by the CMS run 2 data [21], nor by the ATLAS run 1 analysis [22]. Nonetheless, these analyses start probing the interesting region of parameter space where the off-diagonal Yukawa couplings are set by the mass of the leptons, $|Y_{\mu\tau} Y_{\tau\mu}| \sim m_\tau m_\mu / v^2$, one order of magnitude better than the indirect bounds set by $\tau \rightarrow \mu\gamma$ and $\tau \rightarrow 3\mu$.

5 The SM Higgs boson at the LHC

The main production mechanisms at the LHC are gluon fusion, weak-boson fusion, associated production with a gauge boson and associated production with a pair of top/antitop quarks. Figure 2 depicts representative diagrams for these dominant Higgs production processes.

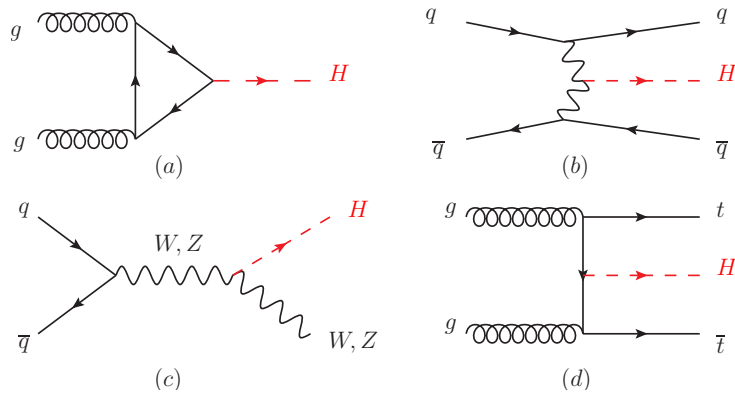


Fig. 2: Generic Feynman diagrams contributing to the Higgs production in (a) gluon fusion, (b) weak-boson fusion, (c) Higgs-strahlung (or associated production with a gauge boson) and (d) associated production with top quarks. From Ref. [6].

The cross sections for the production of a SM Higgs boson as a function of \sqrt{s} , the center of mass energy, for pp collisions, are summarized in Figure 3(left). A detailed discussion, including uncertainties in the theoretical calculations due to missing higher-order effects and experimental uncertainties on the determination of SM parameters involved in the calculations can be found in Refs. [7, 23–25]. These references also discuss the impact of PDF’s uncertainties, QCD scale uncertainties and uncertainties due to different matching procedures when including higher-order corrections matched to parton shower simulations as well as uncertainties due to hadronization and parton-shower events. Table 1 summarizes state-of-the-art of the theoretical calculations in the main different production channels.

Among other subdominant production channels of the Higgs boson at the LHC, the production in association with a single top quark, the production in association with a pair of bottom quarks and the production in association with a pair of charm quarks are particularly interesting and may become visible with the high statistics of the HL-LHC run.

ggF	VBF	VH	$t\bar{t}H$
Fixed order: NNLO QCD + NLO EW (HIGLU, iHixs, FeHiPro, HNNLO)	Fixed order: NNLO QCD (VBF@NNLO)	Fixed order: NLO QCD+EW (V2HV and HAWK)	Fixed order: NLO QCD (Powheg)
Resummed: NNLO + NNLL QCD (HRes)	Fixed order: NLO QCD + NLO EW (HAWK)	Fixed order: NNLO QCD (VH@NNLO)	(MG5_aMC@NLO)
Higgs p_T : NNLO+NNLL (HqT, HRes)			
Jet Veto: N3LO+NNLL			

Table 1: State-of-the-art of the theoretical calculations in the main different Higgs production channels in the SM, and main MC tools used in the simulations. From Ref. [6].

Figure 3 (right) reports the SM predictions for the decay fractions of the Higgs boson. A Higgs mass of about 125 GeV provides an excellent opportunity to explore the Higgs couplings to many SM particles. In particular the dominant decay modes are $H \rightarrow b\bar{b}$ and $H \rightarrow WW^*$, followed by $H \rightarrow gg$, $H \rightarrow \tau^+\tau^-$, $H \rightarrow c\bar{c}$ and $H \rightarrow ZZ^*$. With much smaller rates follow the Higgs decays into $H \rightarrow \gamma\gamma$, $H \rightarrow \gamma Z$ and $H \rightarrow \mu^+\mu^-$. Since the decays into gluons, diphotons and $Z\gamma$ are loop induced, they provide indirect information on the Higgs couplings to WW , ZZ and $t\bar{t}$ in different combinations. The uncertainties in the branching ratios include the missing higher-order corrections in the theoretical calculations as well as the errors in the SM input parameters, in particular fermion masses and the QCD gauge coupling, involved in the decay. The state-of-the-art calculations of the theoretical uncertainties is discussed in Ref. [7].

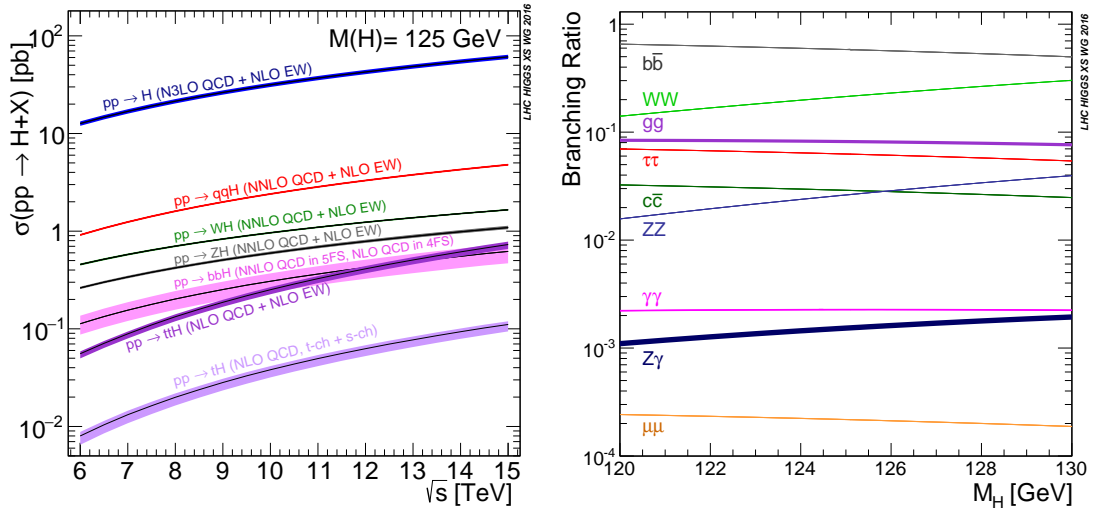


Fig. 3: (Left) The SM Higgs boson production cross sections as a function of the center of mass energy, \sqrt{s} , for pp collisions. (Right) The branching ratios for the main decays of the SM Higgs boson near $m_H = 125$ GeV. For both plots, the theoretical uncertainties are indicated as bands. From Ref. [7].

6 The Higgs mass as a model-discriminator

As indicated in the previous section, the value of the Higgs boson mass opens many decay modes at a rate accessible experimentally. Two channels are particularly accurate in accessing the Higgs mass: $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 2\ell^+2\ell^-$. Figure 4 summarizes the mass measurements in these two channels and their combination [26]. The ATLAS and CMS combined mass measurement:

$$m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.}) \text{ GeV}$$

reaches a precision of 0.2% and is dominated by statistical uncertainties.

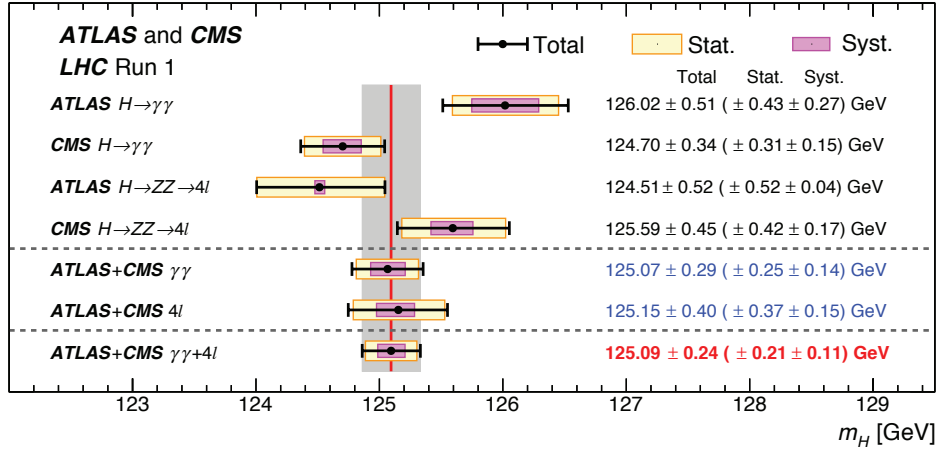


Fig. 4: Compilation of the CMS and ATLAS mass measurements in the $\gamma\gamma$ and ZZ channels, the combined result from each experiment and their combination. From Ref. [26].

Under the assumption that the SM laws govern Nature up to very high energy, the precise value of the Higgs mass has thrilling implications on the stability of the EW vacuum and hence the fate of our Universe (see for instance Ref. [27] for an extensive list of references).

The value of the Higgs mass also gives clues about the details of possible Ultra-Violet (UV) completions of the SM itself. This can be exemplified in the leading scenarios, namely the Minimal Supersymmetric Model (MSSM) and the Minimal Composite Higgs model (MCHM). In short, the Higgs mass is larger than what is typically expected in the MSSM and smaller than what is expected in the MCHM. At the classical/Born level, the mass of the lightest MSSM (SM-like) Higgs boson is bounded to be lower than the Z -boson mass since supersymmetry dictates the Higgs quartic to be fixed in terms of the gauge couplings. Some significant amount of radiative corrections, mostly from the top and stop sectors, are therefore called to raise the value of the Higgs mass. At one-loop, the Higgs mass can be approximated by

$$M_h^2 \simeq M_Z^2 \cos^2 2\beta + \frac{3\sqrt{2}G_F M_t^4}{16\pi^2} \left[\log \frac{M_{\tilde{t}}^2}{M_t^2} + \frac{X_t^2}{M_{\tilde{t}}^2} \left(1 - \frac{X_t^2}{12M_{\tilde{t}}^2} \right) \right], \quad (6)$$

where $M_{\tilde{t}}^2 = M_{Q_3} M_{U_3}$ is the geometric mean of the stop masses and X_t is the mixing between the two stops. Clearly, a Higgs boson as heavy as 125 GeV requires either heavy stops ($M_{\tilde{t}} > 800$ GeV) and/or maximally mixed stops ($X_t \simeq \sqrt{6}M_{\tilde{t}}$), which brings back some amount of irreducible fine-tuning or call for non-trivial boundary conditions like non-universal gaugino masses at high-energy. Going beyond the minimal model, for instance by adding an extra gauge singlet, can easily help increasing the Higgs mass with significantly less amount of tuning, see for instance Ref. [28] for a discussion.

In the Minimal Composite Higgs models, the Higgs boson emerges from the strong sector as a pseudo-Nambu-Goldstone boson. Therefore, the strong interactions themselves are not responsible for

generating a potential for the Higgs field, that is generated only at the one-loop level from the interactions between the strong sector and the SM. Computing the details of the potential from first principles remains out of reach but it is possible [29] to estimate the Higgs mass using general properties about the asymptotic behavior of correlators, i.e. imposing the saturation of the Weinberg sum rules with the first few light resonances, to obtain

$$M_h^2 \simeq \frac{3M_t^2 M_Q^2}{\pi^2 f^2}, \quad (7)$$

where f is the scale of the strong interactions (the decay constant of the Higgs boson, the equivalent of f_π for the QCD pions) and M_Q is the typical mass scale of the fermion resonances (aka the top partners). This estimate can read as

$$M_Q \simeq 700 \text{ GeV} \left(\frac{M_h}{125 \text{ GeV}} \right) \left(\frac{160 \text{ GeV}}{M_t} \right) \left(\frac{f}{500 \text{ GeV}} \right). \quad (8)$$

For a natural set-up ($v^2/f^2 \leq 0.2$), we therefore expect some light top partners below one TeV. The discovery of such fermionic top-partners would be a first evidence of a strong dynamics at the origin of the breaking of the electroweak symmetry.

7 The Higgs profile as a probe of new physics

A dedicated study of the Higgs boson properties and couplings offers a way to infer what the structure of physics beyond the Standard Model can be. Natural models trying to give a rationale for why/how the Higgs mass is screened from high energy corrections at the quantum level generically predict some deviations in the Higgs couplings compared to the SM predictions of the order 1% to 100%. The current Higgs data accumulated at the LHC by the ATLAS and CMS collaborations already constrain the Higgs couplings to massive gauge bosons and to fermions not to deviate by more than 20–30% from the SM predictions [30].

In general, new physics can deform the SM in many ways but most of these deformations are already severely constrained by electroweak precision measurements or flavor data. Assuming flavor universality among the couplings between the Higgs boson and the SM fermions, it was shown [4, 31] that eight directions among the leading CP-conserving deformations of the SM can be probed, at tree-level, only in processes with a physical Higgs boson. These deformation induce deviations in the Higgs couplings that respect the Lorentz structure of the SM interactions, or generate simple new interactions of the Higgs boson to the W and Z field strengths, or induce some contact interactions of the Higgs boson to photons (and to a photon and a Z boson) and gluons that take the form of the ones that are generated by integrating out the top quark. In other words, the Higgs couplings are described, in the unitary gauge, by the following effective Lagrangian [25, 32]

$$\begin{aligned} \mathcal{L} = & \kappa_3 \frac{m_H^2}{2v} H^3 + \kappa_Z \frac{m_Z^2}{v} Z_\mu Z^\mu H + \kappa_W \frac{2m_W^2}{v} W_\mu^+ W^{-\mu} H + \kappa_g \frac{\alpha_s}{12\pi v} G_{\mu\nu}^a G^{a\mu\nu} H \\ & + \kappa_\gamma \frac{\alpha}{2\pi v} A_{\mu\nu} A^{\mu\nu} H + \kappa_{Z\gamma} \frac{\alpha}{\pi v} A_{\mu\nu} Z^{\mu\nu} H + \kappa_{VV} \frac{\alpha}{2\pi v} (\cos^2 \theta_W Z_{\mu\nu} Z^{\mu\nu} + 2 W_{\mu\nu}^+ W^{-\mu\nu}) H \\ & - \left(\kappa_t \sum_{f=u,c,t} \frac{m_f}{v} \bar{f}_L f_R + \kappa_b \sum_{f=d,s,b} \frac{m_f}{v} \bar{f}_L f_R + \kappa_\tau \sum_{f=e,\mu,\tau} \frac{m_f}{v} \bar{f}_L f_R + h.c. \right) H. \end{aligned} \quad (9)$$

In the SM, the Higgs boson does not couple to massless gauge bosons at tree level, hence $\kappa_g = \kappa_\gamma = \kappa_{Z\gamma} = 0$. Nonetheless, the contact operators are generated radiatively by loops of SM particles. In particular, the top quark gives a contribution to the 3 coefficients $\kappa_g, \kappa_\gamma, \kappa_{Z\gamma}$ that does not decouple in the infinite top mass limit. For instance, in that limit $\kappa_\gamma = \kappa_g = 1$ [33, 34].

The coefficient for the contact interactions of the Higgs boson to the W and Z field strengths is not independent but obeys the relation

$$(1 - \cos^4 \theta_W) \kappa_{VV} = \sin 2\theta_W \kappa_{Z\gamma} + \sin^2 \theta_W \kappa_\gamma. \quad (10)$$

This relation is a general consequence of the so-called custodial symmetry [35]. When the Higgs boson is part of an $SU(2)_L$ doublet, the custodial symmetry could only be broken by a single operator at the level of dimension-6 operators and it is accidentally realized among the interactions with four derivatives, like the contact interactions considered. Custodial symmetry also implies $\kappa_Z = \kappa_W$, leaving exactly 8 free couplings [4, 31]. Out of these 8 coefficients, only κ_V can be indirectly constrained by EW precision data at a level comparable from the direct constraints from LHC Higgs data [36].

Table 2 reports the best measurements of the production cross section times branching ratio for the main Higgs channels. These measurements constitute a stress-test of the SM itself (any deviation from $\mu_i = 1$ being an indication of new physics) but they are also used as inputs to fit the κ coupling modifiers. Under several assumptions, for instance on the total width of the Higgs boson, a global fit, as the one reported on Fig. 5, can be performed.

	$\gamma\gamma$	ZZ (4 ℓ)	WW ($\ell\nu\ell\nu$)	$\tau^+\tau^-$	$b\bar{b}$	Comb.
ggF	$1.10^{+0.22+0.07}_{-0.21-0.05}$	$1.13^{+0.33+0.09}_{-0.30-0.07}$	$0.84^{+0.12+0.12}_{-0.12-0.11}$	$1.00^{+0.4+0.4}_{-0.4-0.4}$	—	$1.03^{+0.16}_{-0.14}$
VBF	$1.3 \pm 0.5^{+0.2}_{-0.1}$	$0.1^{+1.1+0.2}_{-0.6-0.2}$	$1.2^{+0.4+0.2}_{-0.3-0.2}$	$1.3^{+0.3+0.2}_{-0.3-0.2}$	—	$1.18^{+0.25}_{-0.23}$
WH	$0.5^{+1.3+0.2}_{-1.2-0.1}$	—	$1.6^{+1.0+0.6}_{-0.9-0.5}$	$-1.4^{+1.2+0.7}_{-1.1-0.8}$	$1.0^{+0.4+0.3}_{-0.4-0.3}$	$0.89^{+0.40}_{-0.38}$
ZH	$0.5^{3.0}_{-2.5}^{+0.5}_{-0.2}$	—	$5.9^{+2.3+1.1}_{-2.1-0.8}$	$2.2^{+2.2+0.8}_{-1.7-0.6}$	$0.4^{+0.3+0.2}_{-0.3-0.2}$	$0.79^{+0.38}_{-0.36}$
ttH	$2.2^{1.6}_{-1.3}^{+0.2}_{-0.1}$	—	$5.0^{+1.5+1.0}_{-1.5-0.9}$	$-1.9^{+3.2+1.9}_{-2.7-1.8}$	$1.1^{+0.5+0.8}_{-0.5-0.8}$	$2.3^{+0.7}_{-0.6}$
Comb.	$1.14^{+0.19}_{-0.18}$	$1.29^{+0.26}_{-0.23}$	$1.09^{+0.18}_{-0.16}$	$1.11^{+0.24}_{-0.22}$	$0.70^{+0.29}_{-0.27}$	$1.09^{+0.11}_{-0.10}$

Table 2: Summary of the combined measurements of the $\sigma \times \text{BR}$ for the five main production and five main decay modes. When uncertainties are separated into two components, the first is the statistical uncertainty and the second is the systematic uncertainty. When only one uncertainty is reported, it is the total uncertainty. From Ref. [6].

The effective Lagrangian of Eq. (9) can be amended by 6 extra Higgs couplings that break the CP symmetry

$$\begin{aligned} \mathcal{L} = & \tilde{\kappa}_g \frac{\alpha_s}{12\pi v} G_{\mu\nu}^a \tilde{G}^{a\mu\nu} H + \tilde{\kappa}_\gamma \frac{\alpha}{2\pi v} A_{\mu\nu} \tilde{A}^{\mu\nu} H + \tilde{\kappa}_{Z\gamma} \frac{\alpha}{\pi v} A_{\mu\nu} \tilde{Z}^{\mu\nu} H \\ & - i \left(\tilde{\kappa}_t \sum_{f=u,c,t} \frac{m_f}{v} \bar{f}_L f_R + \tilde{\kappa}_b \sum_{f=d,s,b} \frac{m_f}{v} \bar{f}_L f_R + \tilde{\kappa}_\tau \sum_{f=e,\mu,\tau} \frac{m_f}{v} \bar{f}_L f_R + h.c. \right) H, \end{aligned} \quad (11)$$

where $\tilde{F}_{\mu\nu} = \epsilon_{\mu\nu\rho\sigma} F^{\rho\sigma}$ is the dual field-strength of $F_{\mu\nu}$. It is certainly tempting to consider new sources of CP violation in the Higgs sector, potentially bringing in one of the necessary ingredients for a successful baryogenesis scenario. It should be noted [37] that these CP violating couplings would induce quark and electron electric dipole moments at one- (for $\tilde{\kappa}_\gamma$ and $\tilde{\kappa}_{Z\gamma}$) or two-loops (for $\tilde{\kappa}_f$). Unless the Yukawa couplings of the Higgs to the electron and light quarks are significantly reduced compared to their SM values, these constraints severely limit the possibility to observe any CP violating signal in the Higgs sector at the LHC.

The coefficient κ_3 can be accessed only through double Higgs production processes, hence it will remain largely unconstrained at the LHC and a future machine like an ILC [38] or a future very

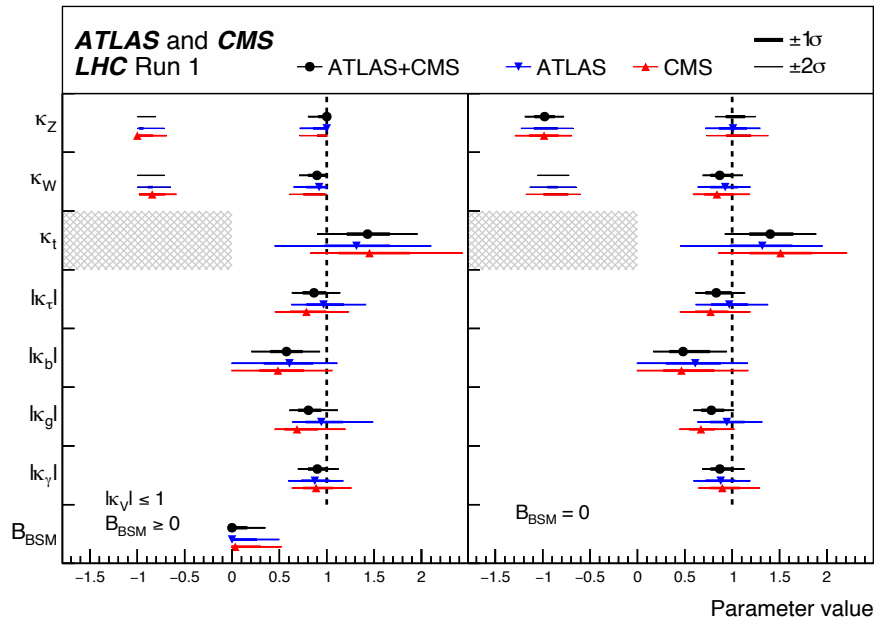


Fig. 5: ATLAS-CMS combined measurements of coupling modifiers. From Ref. [6].

high-energy circular collider might be needed to pin down this coupling [39]. The LHC will also have a limited sensitivity on the coefficient κ_τ since the lepton contribution to the Higgs production cross section remains subdominant and the only way to access the Higgs coupling is via the $H \rightarrow \tau^+\tau^-$ and possibly $H \rightarrow \mu^+\mu^-$ channels. Until the associated production of a Higgs with a pair of top quarks is observed, the Higgs coupling to the top quark is only probed indirectly via the one-loop gluon fusion production or the radiative decay into two photons. However, these two processes are only sensitive to the two combinations $(\kappa_t + \kappa_g)$ and $(\kappa_t + \kappa_\gamma)$ and a deviation in the Higgs coupling to the top quark can in principle always be masked by new contact interactions to photons and gluons (for a discussion, see Ref. [40]). In the next section, we shall discuss how individual information on $\kappa_{\gamma,g,t}$ can be obtained by studying either the hard recoil of the Higgs boson against an extra jet or the off-shell Higgs production in $gg \rightarrow h^* \rightarrow ZZ \rightarrow 4\ell$.

8 Beyond inclusive single Higgs measurements

So far the LHC has mostly produced Higgs bosons on-shell in processes with a characteristic scale around the Higgs mass. This gives a rather good portrait of the Higgs couplings around the weak scale itself. However to fully accomplish its role as a UV regulator of the scattering amplitudes, what matter are the couplings of the Higgs at asymptotically large energy. To probe the Higgs couplings at large energy, one can rely on the associated production with additional boosted particle(s) but the price to pay is to deal with significantly lower production rates.

8.1 Boosted Higgs

The dominant production mode of the Higgs at the LHC is the gluon fusion channel. This is a purely radiative process. The lightness of the Higgs boson plays a malicious role and makes it impossible to disentangle short- and long-distance contribution to the total rate. This limitation is embodied in the Higgs low energy theorem [33,34] that prevents one from resolving the loop contribution itself (the NLO gluon fusion inclusive cross section for a finite and infinite top mass differ only by 1%, see Ref. [41]). New Physics could modify the top Yukawa and also generate a contact interaction to the gluons without

leaving any impact on the total rate, provided that $\kappa_t + \kappa_g = 1$. Concrete examples are top partners in composite Higgs models or mixed stops in the MSSM. Still, extra radiation in the $gg \rightarrow h$ process will allow one to explore the structure of the top loop. When the extra radiation carries away a large amount of energy and boosts the Higgs boson, the process effectively probes the ultraviolet structure of the top loop. Notice that the extra radiation cannot be in the form of a photon, as the amplitude for $gg \rightarrow h + \gamma$ vanishes due to Furry's theorem. One is therefore led to consider the production of h in association with a jet.

Figure 6 gives the sensitivity on the boosted analysis to resolve the κ_t - κ_g degeneracy plaguing the inclusive rate measurement [40]. Similar results have been obtained in Refs. [42,43] and a more realistic analysis of $h \rightarrow 2\ell + \text{jet}$ via $h \rightarrow \tau\tau$ and $h \rightarrow WW^*$ has been performed in Ref. [44].

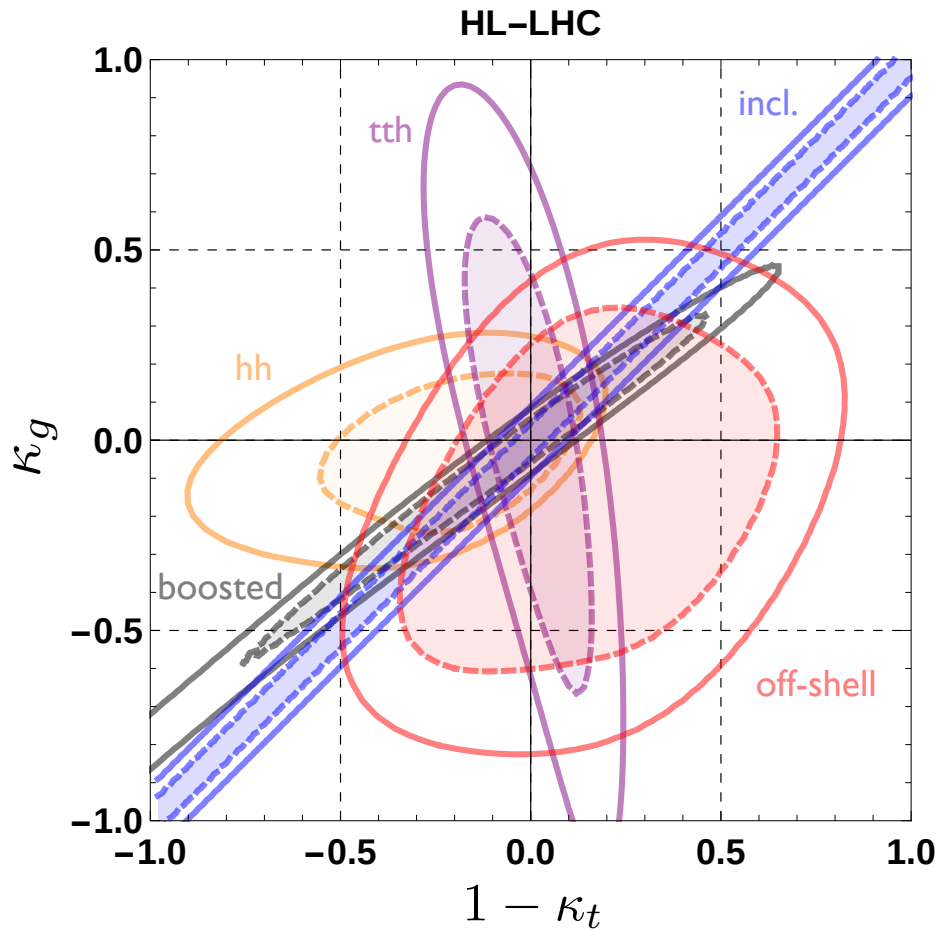


Fig. 6: 95% (solid) and 68% (dashed) exclusion contours in the (κ_t, κ_g) plane obtained from HL-LHC projections (assuming a 14 TeV pp run with 3 ab^{-1} of integrated luminosity): inclusive Higgs measurements (blue), $t\bar{t}h$ (purple), off-shell (red), boosted (gray), and double Higgs production (orange). From Ref. [45].

It should be noted that the $gg \rightarrow h + \text{jet}$ process has been computed only at leading order with the full mass dependence on the fermion running in the loops. The theoretical uncertainty can be estimated by relying on the NNLO K -factors computed in the $m_t \rightarrow \infty$ limit. It is however clear that an exact NLO computation of the SM Higgs p_T spectrum would be very welcome.

8.2 Off-shell Higgs

As for any other quantum particle, the influence of the Higgs boson is not limited to its mass shell. In 2014, the CMS and ATLAS collaborations reported the differential cross-section measurement of $pp \rightarrow$

$Z^{(*)}Z^{(*)} \rightarrow 4\ell, 2\ell 2\nu$ ($\ell = e, \mu$) at high invariant-mass of the ZZ system [46]. This process receives a sizable contribution from a Higgs produced off-shell by gluon fusion [47]. As such, this process potentially carries information relevant for probing the EFT at large momenta and could thus reveal the energy-dependence of the Higgs couplings controlled by higher-dimensional operators with extra derivatives. It has been proposed [48] to use the off-shell Higgs data to bound, in a model-independent way, the Higgs width. However this bound actually holds under the assumption that the Higgs couplings remain unaltered over a large range of energy scales and thus applies only to very specific models. Instead, the off-shell measurement offers a rather unique access to the structure of the Higgs couplings at high energy. Again this channel reveals itself to be particularly efficient to disentangle the long and short distance contribution to the Higgs production by gluon fusion.

Figure 6 also shows the sensitivity on the off-shell analysis to resolve the κ_t - κ_g degeneracy plaguing the inclusive rate measurement [49]. For a recent discussion of off-shell Higgs production within the SM and beyond and an extensive list of references, see Ref. [50].

9 Conclusions

The first run of the LHC operations crowned the Standard Model as the successful description of the fundamental constituents of matter and their interactions to the tiniest details, from the QCD jet production over many orders of magnitude, to the multiple productions of electroweak gauge bosons as well as the production of top quarks. Undeniably, the Higgs boson discovery will remain the acme of the LHC run 1 and it has profound theoretical and phenomenological implications. The LHC run 2 at $\sqrt{s} = 13$ TeV has already beautifully confirmed the pivotal role of the Higgs boson in the Standard Model and it is expected that on its way towards its full high-luminosity run, the LHC will provide invaluable and crucial experimental information on the physics behind the breaking of the electroweak symmetry and it carries the hopes to finally reveal the first cracks in the SM grounds. If naturalness turned out to be a good guide, the LHC should soon find new states and revolutionize the field. If we are not so lucky and such new states are too heavy for the LHC reach, we might still detect indirectly their presence through the deviations they can induce on the Higgs properties. Precise measurements of such properties are therefore crucial and could be extremely useful to guide future direct searches at higher energies, either at the LHC itself or at other future facilities.

The Higgs boson might also be a portal to a hidden sector whose existence is anticipated to account for the total matter and energy budget of the Universe. The Higgs boson could also be one key agent in driving the early exponentially growing phase of our Universe and thus allowing large scale structures to emerge from original quantum fluctuating seeds.

The search for the Higgs boson has occupied the particle physics community for the last 50 years. With the Higgs discovery in 2012, High Energy Physics experiences a profound change in paradigm: What used to be the missing particle in the SM now quickly turns into a tool both to explore the manifestations of the SM and to possible venture into the physics landscape beyond. Whatever the LHC will reveal next, the exploring of new territory is on-going and for sure we, as high-energy practitioners, are living in exciting times!

Acknowledgements

I would like to thank N. Ellis, M. Mulders, E. Carrera for the organization of CLASHEP and the invitation to report on the latest developments on Higgs physics. I also thank all the students, and the other lecturers, for the nice atmosphere during the school. I would like also to thank M. Carena, M. Kado and V. Sharma for their collaboration on the Higgs chapter of the Review of Particle Physics [6], a lot material of which has been used for these lecture notes.

References

- [1] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model*, Phys. Rept. **457** (2008) 1–216, arXiv:hep-ph/0503172 [hep-ph].
- [2] L. Reina, *TASI 2011: lectures on Higgs-Boson Physics*, 2013. arXiv:1208.5504 [hep-ph].
- [3] S. Dittmaier and M. Schumacher, *The Higgs Boson in the Standard Model - From LEP to LHC: Expectations, Searches, and Discovery of a Candidate*, Prog. Part. Nucl. Phys. **70** (2013) 1–54, arXiv:1211.4828 [hep-ph].
- [4] A. Pomarol, *Higgs Physics, in Proceedings, 2014 European School of High-Energy Physics (ESHEP 2014): Garderen, The Netherlands, June 18 - July 01 2014*, pp. 59–77. 2016. arXiv:1412.4410 [hep-ph].
- [5] C. Csaki, C. Grojean, and J. Terning, *Alternatives to an Elementary Higgs*, Rev. Mod. Phys. **88** (2016) no. 4, 045001, arXiv:1512.00468 [hep-ph].
- [6] M. Carena, C. Grojean, M. Kado, and V. Sharma, *Status of Higgs boson physics, in Review of Particle Physics, Chin. Phys., Chin. Phys. C40* (2016) 100001. <http://pdg.lbl.gov/2016/reviews/rpp2016-rev-higgs-boson.pdf>.
- [7] LHC Higgs Cross Section Working Group, de Florian, D. et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, arXiv:1610.07922 [hep-ph].
- [8] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys.Rev.Lett. **13** (1964) 321–323.
- [9] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys.Lett. **12** (1964) 132–133.
- [10] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys.Rev.Lett. **13** (1964) 508–509.
- [11] G. Guralnik, C. Hagen, and T. Kibble, *Global Conservation Laws and Massless Particles*, Phys.Rev.Lett. **13** (1964) 585–587.
- [12] ATLAS Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys.Lett. **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].
- [13] CMS Collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys.Lett. **B716** (2012) 30–61, arXiv:1207.7235 [hep-ex].
- [14] R. Rosenfeld, *Physics Beyond the Standard Model*, in these proceedings (2016) .
- [15] G. Giudice, C. Grojean, A. Pomarol, and R. Rattazzi, *The Strongly-Interacting Light Higgs*, JHEP **0706** (2007) 045, arXiv:hep-ph/0703164 [hep-ph].
- [16] R. Contino, C. Grojean, M. Moretti, F. Piccinini, and R. Rattazzi, *Strong Double Higgs Production at the LHC*, JHEP **1005** (2010) 089, arXiv:1002.1011 [hep-ph].
- [17] CMS Collaboration, Khachatryan, V. et al., *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV*, Eur. Phys. J. **C75** (2015) no. 5, 212, arXiv:1412.8662 [hep-ex].
- [18] A. J. Buras, C. Grojean, S. Pokorski, and R. Ziegler, *FCNC Effects in a Minimal Theory of Fermion Masses*, JHEP **1108** (2011) 028, arXiv:1105.3725 [hep-ph].
- [19] R. Harnik, J. Kopp, and J. Zupan, *Flavor Violating Higgs Decays*, JHEP **1303** (2013) 026, arXiv:1209.1397 [hep-ph].
- [20] CMS Collaboration, Khachatryan, V. et al., *Search for Lepton-Flavour-Violating Decays of the Higgs Boson*, Phys. Lett. **B749** (2015) 337–362, arXiv:1502.07400 [hep-ex].
- [21] CMS Collaboration, *Search for Lepton Flavour Violating Decays of the Higgs Boson in the mu-tau final state at 13 TeV*, . <http://inspirehep.net/record/1468550>.
- [22] ATLAS Collaboration, Aad, G. et al., *Search for lepton-flavour-violating decays of the Higgs and*

- Z bosons with the ATLAS detector*, arXiv:1604.07730 [hep-ex].
- [23] LHC Higgs Cross Section Working Group, Dittmaier, S. et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, arXiv:1101.0593 [hep-ph].
- [24] LHC Higgs Cross Section Working Group, Dittmaier, S. et al., *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, arXiv:1201.3084 [hep-ph].
- [25] LHC Higgs Cross Section Working Group, Heinemeyer, S. et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, arXiv:1307.1347 [hep-ph].
- [26] ATLAS and CMS Collaborations, Aad, G. et al., *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, arXiv:1503.07589 [hep-ex].
- [27] D. Buttazzo, G. Degrassi, P. P. Giardino, G. F. Giudice, F. Sala, et al., *Investigating the near-criticality of the Higgs boson*, JHEP **1312** (2013) 089, arXiv:1307.3536 [hep-ph].
- [28] M. Farina, M. Perelstein, and B. Shakya, *Higgs Couplings and Naturalness in λ -SUSY*, JHEP **1404** (2014) 108, arXiv:1310.0459 [hep-ph].
- [29] G. Panico and A. Wulzer, *The Composite Nambu-Goldstone Higgs*, Lect. Notes Phys. **913** (2016) pp.1–316, arXiv:1506.01961 [hep-ph].
- [30] ATLAS and CMS Collaborations, G. Aad et al., *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV*, JHEP **08** (2016) 045, arXiv:1606.02266 [hep-ex].
- [31] R. S. Gupta, A. Pomarol, and F. Riva, *BSM Primary Effects*, Phys.Rev. **D91** (2015) 035001, arXiv:1405.0181 [hep-ph].
- [32] LHC Higgs Cross Section Working Group, David, A. et al., *LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle*, arXiv:1209.0040 [hep-ph].
- [33] J. R. Ellis, M. K. Gaillard, and D. V. Nanopoulos, *A Phenomenological Profile of the Higgs Boson*, Nucl.Phys. **B106** (1976) 292.
- [34] M. A. Shifman, A. Vainshtein, M. Voloshin, and V. I. Zakharov, *Low-Energy Theorems for Higgs Boson Couplings to Photons*, Sov.J.Nucl.Phys. **30** (1979) 711–716.
- [35] R. Contino, M. Ghezzi, C. Grojean, M. Muhlleitner, and M. Spira, *Effective Lagrangian for a light Higgs-like scalar*, JHEP **1307** (2013) 035, arXiv:1303.3876 [hep-ph].
- [36] M. Ciuchini, E. Franco, S. Mishima, and L. Silvestrini, *Electroweak Precision Observables, New Physics and the Nature of a 126 GeV Higgs Boson*, JHEP **1308** (2013) 106, arXiv:1306.4644 [hep-ph].
- [37] D. McKeen, M. Pospelov, and A. Ritz, *Modified Higgs branching ratios versus CP and lepton flavor violation*, Phys.Rev. **D86** (2012) 113004, arXiv:1208.4597 [hep-ph].
- [38] K. Fujii et al., *Physics Case for the International Linear Collider*, arXiv:1506.05992 [hep-ex].
- [39] R. Contino et al., *Physics at a 100 TeV pp collider: Higgs and EW symmetry breaking studies*, arXiv:1606.09408 [hep-ph].
- [40] C. Grojean, E. Salvioni, M. Schlaffer, and A. Weiler, *Very boosted Higgs in gluon fusion*, JHEP **1405** (2014) 022, arXiv:1312.3317 [hep-ph].
- [41] M. Grazzini and H. Sargsyan, *Heavy-quark mass effects in Higgs boson production at the LHC*, JHEP **1309** (2013) 129, arXiv:1306.4581 [hep-ph].
- [42] A. Banfi, A. Martin, and V. Sanz, *Probing top-partners in Higgs+jets*, JHEP **1408** (2014) 053, arXiv:1308.4771 [hep-ph].
- [43] A. Azatov and A. Paul, *Probing Higgs couplings with high p_T Higgs production*, JHEP **1401**

- (2014) 014, arXiv:1309.5273 [hep-ph].
- [44] M. Schlaffer, M. Spannowsky, M. Takeuchi, A. Weiler, and C. Wymant, *Boosted Higgs Shapes*, Eur.Phys.J. **C74** (2014) no. 10, 3120, arXiv:1405.4295 [hep-ph].
- [45] A. Azatov, C. Grojean, A. Paul, and E. Salvioni, *Resolving gluon fusion loops at current and future hadron colliders*, JHEP **09** (2016) 123, arXiv:1608.00977 [hep-ph].
- [46] CMS Collaboration, Khachatryan, V. et al., *Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs*, Phys. Lett. **B736** (2014) 64–85, arXiv:1405.3455 [hep-ex].
- [47] E. N. Glover and J. van der Bij, *Z boson pair production via gluon fusion*, Nucl.Phys. **B321** (1989) 561.
- [48] F. Caola and K. Melnikov, *Constraining the Higgs boson width with ZZ production at the LHC*, Phys.Rev. **D88** (2013) 054024, arXiv:1307.4935 [hep-ph].
- [49] A. Azatov, C. Grojean, A. Paul, and E. Salvioni, *Taming the off-shell Higgs boson*, JETP **3** (2014) no. 147, 410, arXiv:1406.6338 [hep-ph].
- [50] N. Kauer, *Off-shell Higgs signal and total width determination at the LHC*, PoS **FFP14** (2016) 114, arXiv:1502.02581 [hep-ph].

Physics Beyond the Standard Model

R. Rosenfeld

ICTP South American Institute for Fundamental Research & Instituto de Física Teórica
Universidade Estadual Paulista, São Paulo, Brazil

Abstract

In these three lectures I review the need to go beyond the Standard Glashow-Weinberg-Salam Model and discuss some of the approaches that are explored in this direction.

Keywords

Lectures; beyond the standard model; supersymmetry; particle physics; effective lagrangian; hidden sector.

1 Introduction

In these lectures I took a somewhat different approach to introduce the possibility of new physics beyond the Standard Model (SM). Usually one finds in most lectures discussions about well motivated models proposed to solve the most pressing issue of the SM: the hierarchy problem. Hence one usually presents the three extensions that tackles this issue: supersymmetric models, composite models and models with compact extra dimensions.

At the expense of being less motivated, I decided to not introduce these grand schemes and instead be more model-independent and making only minimal modifications to the SM, restricting myself to the scalar sector. These modifications already gives a rough idea of some the consequences of expanding the SM.

In this manner after a discussion of the hierarchy problem I started with the assumption that no new particles will be directly found at accelerators, possibly due to the large mass scales involved. In this case new physics would manifest itself in the form of new higher dimensional operators induced by integrating out these heavy new degrees of freedom and suppressed by the scale of new physics. This is the spirit of the Effective Theory approach. Afterwards I introduce minimal modifications in the scalar sector, which already produces modified Higgs couplings and a dark matter candidate. In the final lecture the traditional models for physics beyond the SM, such as SUSY, Composite Higgs and Extra Dimensions are briefly touched upon.

I must say right away that in my opinion a full-fledged write-up of these lectures, which still need much improvement, is not essential at this point. Hence these notes are intended as just a brief guide to what was discussed. The slides of my lectures, as well as for the other lectures of the School can be found at:

<http://physicschool.web.cern.ch/PhysicSchool/CLASHEP/CLASHEP2015/programme.html>

There is some overlap with the excellent lectures by Christophe Grojean on Higgs Physics in this School.

In addition, there are several lectures on Physics Beyond the Standard Model, among which I list:

- Lectures by Alex Pomarol at the 2010 CERN School [1];
- Lectures by Eduardo Pontón at the 2012 TASI School [2];
- Lectures by Joe Lykken at the 2009 CERN School [3];
- Lectures by Tony Gherghetta at the 2009 TASI School [4]

Therefore in the following I will present a simple sketch of my lectures pointing to some references where more details can be found.

2 First lecture

In this lecture I started by recalling the astounding success of the SM, exemplified by the muon and electron magnetic moments, the Z boson line shape and the number of neutrinos, precision electroweak measurements at LEP, measurements of cross sections for SM processes at the LHC and, finally, the historical discovery of a Higgs boson in 2012. I also briefly mentioned the recent measurements of Higgs couplings which are in agreement with SM predictions within experimental errors.

Next the usual shortcomings of the SM were mentioned, noticing that it does not explain several issues:

- the 19 free parameters necessary to explain the observed phenomena (flavor problem, absence of strong CP violation, etc);
- origin of electroweak symmetry breaking;
- neutrino masses;
- dark matter;
- origin of the matter-antimatter asymmetry;
- dark energy;
- inflation;
- gravity.

In addition, we discussed some of the more conceptual problems related to the scalar sector of the SM: perturbative unitarity, triviality, vacuum stability and, especially, the hierarchy or naturalness problem. I recommend the review of Giudice [5] for a very lucid explanation of the naturalness problem.

In particular, the hierarchy problem arises from the absence of a symmetry that could protect the Higgs mass from receiving dangerous quadratic quantum corrections. It has played a major role in the development of models beyond the SM. Favorite models such as Supersymmetry, Composite Higgs and Compact Extra Dimensions have naturalness as their main motivation. All solutions to the naturalness/hierarchy problem lead to new physics at scales not much above the electroweak scale. The discovery of a light Higgs boson and the absence of any new particles or deviations of couplings at the LHC have put theories motivated by the naturalness principle under stress. It has been called “the LHC battle for naturalness”¹.

If the mass scale of new physics is beyond the reach of the LHC their main effects can be parameterized in an effective lagrangian that includes higher dimensional terms involving only SM fields that respect the known SM symmetries. This approach is very general and can be called agnostic in the sense that it does not depend on the details of the underlying model. The number of terms is finite but can be large. At dimension-6, there are 59 terms that can be added to the SM lagrangian [6].

Under certain assumptions, the absence of new physics results in constraints on the energy scale associate to it. However, one must keep in mind that a given observable may be sensitive to a combination of dimension-6 operators and on the other hand each dimension-6 operator may affect more than one observable. Typically, barring spurious cancellations among different operators one obtains bounds from LEP electroweak precision measurements from LEP1 alone that the scale of new physics Λ should be larger than $O(10 \text{ TeV})$. LEP2 and LHC bounds are less restrictive, roughly $\Lambda > 300 \text{ GeV}$.

The least experimentally constrained operators are the ones involving the third generation, such as top quark dipole operators. These are also expected to receive contributions if new physics couples dominantly to the third generation, which is the case for several SM extensions.

I also briefly mentioned that an Effective Lagrangian approach can also be used to parameterize the interaction of a dark matter sector to the SM [7] and this has been used in experimental searches [8].

¹See <https://indico.cern.ch/event/290373/>

I concluded the first lecture remarking that Effective Lagrangians are an indirect and agnostic way to study new physics. One might say it represents the lamp post approach – trying to find new phenomena hiding in error bars. It is difficult to derive firm conclusion in this approach: bounds usually depend on combinations of Wilson coefficients and energy scale(s) of new physics. In the end we will only be convinced of new physics by direct evidence!

3 Second lecture

The Higgs boson may be the first fundamental scalar particle found in Nature. It is conceivable that there are more scalar particles out there in a “hidden sector”. They may communicate to us only via the Higgs: the Higgs acts like a portal between the SM and this new sector.

In the second lecture I discuss the simplest extension of the SM: the addition of a singlet real scalar field which we denote by S interacting only with the Higgs doublet via a renormalizable term in the potential. This is not motivated by any grand principle such as naturalness but illustrates some phenomena common to some more complete extensions.

Already in this simple extension one has 2 choices (with different phenomenology) in writing the potential: to allow or not for S to have a vacuum expectation value (vev).

3.1 (S) $I=0$

If S has a nonzero vev it can mix with the Higgs boson, generating two mass eigenstates that we call H_1 and H_2 , and identifying H_1 with the 125 GeV scalar found at the LHC. There are three additional parameters with respect to the SM, related to the S mass, self-coupling and coupling to the Higgs doublet.

In this model there are some phenomenological consequences:

- all Higgs couplings are reduced by a common factor of $\cos \theta$, where θ is a mixing angle. Hence all Higgs partial widths are reduced by $\cos^2 \theta$ with respect to the SM value;
- couplings of the second Higgs to gauge bosons and fermion are the same of a SM Higgs reduced by $\sin \theta$;
- there are new processes (depending on the mass of H_2):
 $H_2 \rightarrow H_1 H_1$ if $m_{H_2} > 250$ GeV and $H_1 \rightarrow H_2 H_2$ if $m_{H_2} < 62.5$ GeV.

There are many bounds in this model coming from:

- perturbativity of couplings;
- vacuum stability (potential bounded from below);
- EW precision measurements: modified couplings, new loop contributions from H_2 ;
- LEP direct searches (low mass H_2);
- LHC direct searches (high mass H_2);
- Higgs couplings at LHC ($H_1 \rightarrow \gamma\gamma, 4f$) (modification of widths due to mixing, possible new contribution to H_1 width for light H_2);
- partial unitarization.

The bounds on the parameters of this simple extension can be found in [9, 10].

The possibility of a resonant di-Higgs production through $pp \rightarrow H_2 \rightarrow H_1 H_1$ for a heavy H_2 is interesting since the SM cross section for double Higgs production is very small. In particular, if H_2 is very heavy the decays of H_1 will be boosted and jet substructure techniques can be used to search for this process in the final state with 4 b quarks [11]. This has been searched for by CMS [12] and ATLAS [13]. There was also a search in the rarer $bb\gamma\gamma$ channel by CMS [14].

3.2 $(S) = 0$

It is possible that there is an unbroken Z_2 symmetry in the scalar sector under which $S \leftrightarrow -S$. This symmetry forbids the field S to develop a vacuum expectation value. It also makes the S boson stable. Hence this model is arguably the simplest extension of the SM with a dark matter candidate [15]. In fact, since S has self-interactions this is a model of self-interacting dark matter. The Z_2 symmetry also forbids a mixing between the new scalar field with the Higgs field. However, the Higgs boson can decay into two dark matter particles, $H \rightarrow SS$ leading to invisible Higgs decays [16, 17].

A term such as $\lambda_{HS} S^2 H^2$ controls:

- $SS \leftrightarrow SM SM$ (annihilation to SM particles, that determine DM relic abundance);
- $SN \rightarrow SN$ (elastic scattering off nucleons, that determine DM direct detection);
- $H \rightarrow SS$ (invisible Higgs decay).

A term such as S^4 controls DM self-interactions.

This model has few free parameters and is already severely constrained. For a recent analysis see [18].

One comment about these models is that it introduces a new physical scale: the mass of the particle S . This in turn generates a hierarchy problem since the S particle induces a new contribution to the Higgs boson mass. Again, naturalness implies that the S particle can not be too heavy, typically:

$$M_S^2 < \frac{16\pi^2}{\lambda_{HS}} M_H^2$$

Many extensions of the SM build on this simple class of models, just adding more scalar fields: complex singlet, 2-Higgs doublets (inert or active, SUSY), Higgs triplets, etc.

Although the vacuum stability issue at high energies can be ameliorated, the simple model discussed here was not built to avoid the naturalness problem, which is arguably the guiding principle to BSM. The next lecture is about models that were motivated by the hierarchy/naturalness problem.

4 Third lecture

In this lecture I begin by discussing the naturalness problem for the electron mass when taking into account the electron self-energy due to its own electrical field. The self-energy contribution is actually divergent and this problem was eventually solved in Quantum Field Theory, where one knows that the electron mass is protected by chiral symmetry - the self-energy of the electron computed in QFT is proportional to the electron mass itself.

This type of mechanism led 't Hooft to conjecture the following "dogma" in 1980 that he called naturalness: "at any energy scale μ a physical parameter or set of physical parameters $\alpha_i(\mu)$ is allowed to be very small only if the replacement $\alpha_i(\mu) = 0$ would increase the symmetry of the system". For example, setting the mass of the electron to zero restores chiral symmetry in QED. This dogma was promoted into a principle in later years.

Setting the Higgs mass to zero in the SM does not increase any symmetry. There is no natural reason for why the Higgs boson should be light. In fact, there are quantum quadratic contributions to the Higgs boson mass that makes it sensitive to very high energy scales. Requiring that in the SM the contributions to the Higgs boson mass from a scale Λ is smaller than the Higgs boson mass itself requires $\Lambda < 600$ GeV. But LHC has ruled out New Physics at this scale. Thus the LHC has shown that SM is not natural.

This is the "naturalness" motivation to go BSM: find a mechanism that can explain why $M_H \ll \Lambda$ for large values of the cut-off representing a physical scale where New Physics should show up.

This requires either a new symmetry to protect the Higgs mass or a mechanism to lower the cut-off. In the first category we may list:

- Supersymmetry (cancellation of quadratic divergences);
- Shift symmetry (Higgs as a pseudo-Nambu-Goldstone boson);
- Conformal symmetry (Higgs as a dilaton);

whereas the latter case includes

- Flat extra-dimensions (Large Extra Dimensions, Universal Extra Dimensions);
- Warped Extra Dimensions (Randall-Sundrum models)

I cannot go into the details about these different alternatives in these Proceedings. There is a vast literature on these naturalness-motivated BSM. It suffices to say that even in the models mentioned above the absence of New Physics at the LHC is calling into question the naturalness principle. The second run of the LHC at 13 TeV that has just started should bring very important information in the coming years.

5 Acknowledgements

I would like to thank Nick Ellis, Martijn Mulders, Kate Ross and especially Edgar Carrera for the organization of CLASHEP, the invitation to lecture and for the hospitality in Ecuador.

References

- [1] A. Pomarol, CERN Yellow Report CERN-2012-001, 115-151 [arXiv:1202.1391 [hep-ph]].
- [2] E. Ponton, arXiv:1207.3827 [hep-ph].
- [3] J. D. Lykken, CERN Yellow Report CERN-2010-002, 101-109 [arXiv:1005.1676 [hep-ph]].
- [4] T. Gherghetta, arXiv:1008.2570 [hep-ph].
- [5] G. F. Giudice, PoS EPS **-HEP2013**, 163 (2013) [arXiv:1307.7879 [hep-ph]].
- [6] B. Grzadkowski, M. Iskrzynski, M. Misiak and J. Rosiek, JHEP **1010**, 085 (2010) [arXiv:1008.4884 [hep-ph]].
- [7] G. Busoni, A. De Simone, E. Morgante and A. Riotto, Phys. Lett. B **728**, 412 (2014) [arXiv:1307.2253 [hep-ph]].
- [8] V. Khachatryan *et al.* [CMS Collaboration], Eur. Phys. J. C **75**, no. 5, 235 (2015) [arXiv:1408.3583 [hep-ex]].
- [9] T. Robens and T. Stefaniak, Eur. Phys. J. C **75**, no. 3, 104 (2015) [arXiv:1501.02234 [hep-ph]].
- [10] A. Falkowski, C. Gross and O. Lebedev, JHEP **1505**, 057 (2015) [arXiv:1502.01361 [hep-ph]].
- [11] M. Gouzevitch, A. Oliveira, J. Rojo, R. Rosenfeld, G. P. Salam and V. Sanz, JHEP **1307**, 148 (2013) [arXiv:1303.6636 [hep-ph]].
- [12] CMS Collaboration [CMS Collaboration], CMS-PAS-HIG-14-013.
- [13] G. Aad *et al.* [ATLAS Collaboration], arXiv:1506.00285 [hep-ex].
- [14] CMS Collaboration [CMS Collaboration], CMS-PAS-HIG-13-032.
- [15] C. P. Burgess, M. Pospelov and T. ter Veldhuis, Nucl. Phys. B **619**, 709 (2001) [hep-ph/0011335].
- [16] M. C. Bento, O. Bertolami, R. Rosenfeld and L. Teodoro, Phys. Rev. D **62**, 041302 (2000) [astro-ph/0003350].
- [17] M. C. Bento, O. Bertolami and R. Rosenfeld, Phys. Lett. B **518**, 276 (2001) [hep-ph/0103340].
- [18] J. M. Cline, K. Kainulainen, P. Scott and C. Weniger, Phys. Rev. D **88**, 055025 (2013) [arXiv:1306.4710 [hep-ph]].

QCD under extreme conditions: an informal discussion

*E.S. Fraga**

J. W. Goethe-University, Frankfurt am Main, Germany

Abstract

We present an informal discussion of some aspects of strong interactions under extreme conditions of temperature and density at an elementary level. This summarizes lectures delivered at the 2013 and 2015 CERN – Latin-American Schools of High-Energy Physics and is aimed at students working in experimental high-energy physics.

Keywords

QCD; heavy-ion collisions; quark-gluon plasma; Polyakov loop; chiral symmetry; lectures.

1 Introduction and motivation: why, where and how

Quantum Chromodynamics (QCD) is an extremely successful theory of strong interactions that has passed numerous tests in particle accelerators over more than 40 years [1]. This corresponds to the behavior of hadrons in the vacuum, including not only the spectrum but also all sorts of dynamical processes. More recently strong interactions, and therefore QCD, has also started being probed in a medium, under conditions that become more and more extreme [2]. Although quite involved theoretically, this is not just an academic problem. In order to make it clear, one should consider three very basic questions, that should always be asked in the beginning: why? where? how?

1.1 Why?

It was realized since the very beginning that strong interactions exhibit two remarkable features that are related but represent properties of complementary sectors of the energy scale. The first one is asymptotic freedom [3], which can be perturbatively demonstrated by an explicit computation of the beta function to a give loop order in QCD [4]. The second, which is consistent with the first but should be seen as totally independent, since it is a property of the nonperturbative vacuum of strong interactions, is color confinement [5]. Even though reality constantly shows that confinement is a property of strong interactions, and therefore should somehow be built in QCD, this proof remains a theoretical open problem so far. Even for the pure Yang-Mills theory, where the bound states correspond to glueballs, the existence of a mass gap is still to be shown after more than half a century of the original paper on nonabelian gauge theories [6]. For this reason, confinement is ranked in the Clay Mathematics Institute list of unsolved Millennium problems [7].

Much more than a cute (and very tough) mathematical problem, this is certainly among the most important theoretical and phenomenological problems in particle physics, since hidden there is the real origin of mass, as we feel in our everyday lives and experience with ordinary (and not so ordinary) matter. Although the Higgs mechanism provides a way to give mass to elementary particles in the Standard Model [8], most of what constitutes the masses of hadrons come from interactions. For instance, more than 90% of the proton mass originates in quark and gluon condensates [9]. So, in spite of the fantastic success of the Standard Model [8], we do not understand a few essential mechanisms.

Extremely high temperatures and densities bring us to an energy scale that facilitates deconfinement, and matter under such extreme conditions can behave in unexpected ways due to collective effects.

*On leave from Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

This is, of course, a way to study the mechanism of confinement (by perturbing or modifying this state of matter). This leads us also to a deeper yet childish motivation, that of understanding what happens if we keep making things hotter and hotter, or keep squeezing things harder and harder [10]. These questions can be reformulated in a more technical fashion as ‘what is the inner structure of matter and the nature of strong interactions under extreme conditions of temperature and density?’. In experiments, one needs to “squeeze”, “heat” and “break”. From the theoretical point of view, one needs a good formulation of in-medium quantum field theory, using QCD or effective theories.

It is clear that the challenge is enormous. Although confinement seems to be a key feature of hadrons, and manifests also in relevant scales such as f_π or Λ_{QCD} , it only *seems* to be present in QCD. So far, controlled lattice simulations show strong evidence of confinement in the pure gauge theory [11]. As hinted previously, however, the theory is nonperturbative at the relevant scales, so that analytic methods are very constrained. And, although lattice simulations have developed to provide solid results in several scenarios, they are not perfect. And, more important, they are not Nature. To make progress in understanding, or at least collecting important facts, one needs it all: experiments and observations, lattice simulations, the full theory in specific (solvable to some extent) limits and effective models. And also combinations, whenever possible, to diminish the drawbacks of each approach.

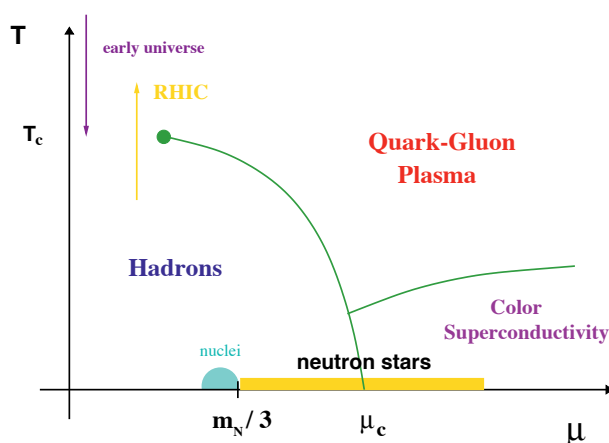


Fig. 1: Cartoon of a phase diagram for strong interactions. Extracted from Ref. [12]

Whichever the framework chosen, collective phenomena will play a major role. Although somewhat put aside in the so-called microscopic “fundamental” particle physics, collective effects can affect dramatically the behavior of elementary particles in a medium under certain conditions. Besides the well-known examples of BCS and BEC phases in condensed matter systems [13], and also in dense quark matter [14], it was recently found that photons can form a Bose-Einstein condensate [15]. In fact, the textbook case of water and its different phases is quite illustrative of the richness that comes from collective phenomena that would hardly be guessed from the case of very few or non-interacting elementary particles.

In terms of the thermodynamics, or many-body problem, the basic idea is to perturb the (confined) vacuum to study confinement by heating (temperature), squeezing or unbalancing species (chemical potentials for baryon number, isospin, strangeness, etc) and using classical external fields (magnetic, electric, etc), so that the system is taken away from the confined phase and back. One can also relate (or not) confinement to other key properties of strong interactions, such as chiral symmetry. And, from the theorist standpoint, draw all possible phase diagrams of QCD and its “cousin theories” (realizations of QCD with parameters, such as the number of colors or flavors, or the values of masses, that are not realized in Nature) to learn basic facts. There are several examples, one well-known being the ‘Columbia plot’, where one studies the nature of the phase transitions and critical lines on the $(m_u = m_d, m_s)$

plane. Nevertheless, if one draws a cartoon of the phase diagram in the temperature vs. quark chemical potential, for instance Fig. 1, and compares it to computations from effective models, lattice simulations and freeze-out points extracted from high-energy heavy ion collision data, one sees that the points still scatter in a large area [16]. So, there is still a long way ahead.

1.2 Where?

According to the Big Bang picture and the current description of the evolution of the early universe [17], we expect that at about $10^{-5}s$ after the Big Bang a soup of quark-gluon plasma (in the presence of electrons, photons, etc) has undergone a phase transition to confined hadrons. This was, of course, the first realization of a QCD transition. This process was thermally driven and happened at very low baryon chemical potential.

It is quite remarkable that the scales of strong interactions allow for the experimental reproduction of analogous conditions in high-energy ultra-relativistic heavy ion collisions in the laboratory [18]. In a picture by T. D. Lee, these collisions are seen as heavy bulls that collide and generate new states of matter [19]. Such experiments are under way at BNL-RHIC [20] and CERN-LHC [21], and will be part of the future heavy ion programs at FAIR-GSI [22] and NICA [23].

For obvious reasons, it is common to refer to such experiments as “Little Bangs”. However, one should be cautious with this point. In spite of the fact that the typical energy scales involved need to be the same, as well as the state of matter created, the so-called quark-gluon plasma [24], the relevant space-time scales differ by several orders of magnitude. Using a simple approximation for the equation of state,

$$3p \approx \epsilon \approx \frac{\pi^2}{30} N(T) T^4, \quad (1)$$

where p is the pressure, ϵ the energy density and $N(T)$ the number of relevant degrees of freedom, we can easily estimate the typical sizes involved. The radius of the universe at the QCD phase transition epoch, as given by the particle horizon in a Robertson-Walker space-time [25], where the scale factor grows as $a(t) \sim t^n$, is given by ($n = 1/2$ and $N(T) \sim 50$ at this time for QCD)

$$L_{\text{univ}}(T) \approx \frac{1}{4\pi} \left(\frac{1}{1-n} \right) \left(\frac{45}{\pi N(T)} \right)^{1/2} \frac{M_{\text{Pl}}}{T^2} = \frac{1.45 \times 10^{18}}{(T/\text{GeV})^2 \sqrt{N(T)}} \text{fm}. \quad (2)$$

Here M_{Pl} is the Planck mass, and it is clear that the system is essentially in the thermodynamic limit.

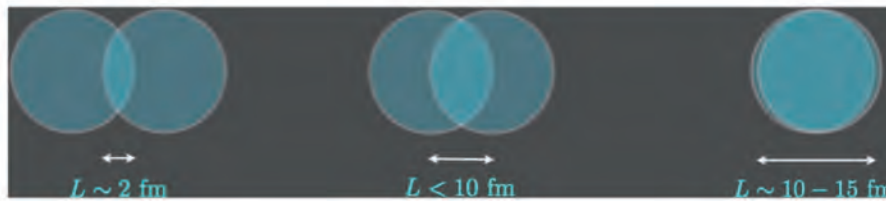


Fig. 2: Cartoon representing non-central heavy ion collisions and how they affect the size of the system.

On the other hand, in heavy ion collisions the typical length scale of the system is $L_{\text{QGP}} \lesssim 10 - 15 \text{ fm}$, so that the system can be very small, especially if one considers non-central collisions [26] (see Fig. 2). One can develop analogous arguments for the time scales given by the expansion rates, finding that the whole process in the early universe happens adiabatically, whereas in heavy ions it is not even clear whether the system can achieve thermal equilibrium, given the explosive nature of the evolution in this case. So, there are certainly large differences (in time and length scales) between Big and Little Bangs...

Keeping this caveat in mind, heavy ion experiments have been investigating new phases of matter at very high energies for more than a decade, producing an awesome amount of interesting data and a richer picture of strong interactions (see Ref. [27] for a review).

In the realization of the Big and Little Bangs one is always in the high temperature and low density (small baryon chemical potential) sector of the phase diagram of strong interactions. However, high densities (at very low temperatures) can also probe new states of hadronic matter, and that is what is expected to be found in the core of compact stars [28]. There, new phases, condensates and even color superconductivity may be present. In particular, the deconfinement and chiral transitions might affect significantly the explosion mechanism in supernovae [28] via modifications in the equation of state.

After a neutron (or hybrid) star is formed, densities in its core can in principle reach several times the nuclear saturation density $n_0 = 0.16 \text{ fm}^{-3} = 3 \times 10^{14} \text{ g/cm}^3$, which corresponds to squeezing ~ 2 solar masses into a sphere of $\sim 10 \text{ km}$ of radius. To describe these objects, one needs General Relativity besides in-medium quantum field theory.

1.3 How?

The reader is hopefully already convinced that, in order to describe the phenomenology of the phase structure and dynamics of strong interactions under extreme conditions, one needs all possibilities at disposal: theory, effective modeling, etc. We do not have one problem ahead, but a myriad of different problems. So, one has to make a choice. Our focus here will be the equation of state, of which we will discuss a few aspects.

At this point, we are lead again to the “why” question. And the answer is because, besides carrying all the thermodynamic equilibrium information we may be interested in, it is also the basic crucial ingredient for dynamics, structure, etc. In fact, the phase diagram topology is determined in every detail by the full knowledge of the pressure $p(T, \mu, B, \dots)$. This will determine all phases present as we dial different knobs, or control parameters, such as temperature or chemical potentials.

The structure of a compact star, for instance, is given by the solution of the Tolman-Oppenheimer-Volkov (TOV) equations [28], which encode Einstein’s General Relativity field equations in hydrostatic equilibrium for a spherical geometry:

$$\frac{dp}{dr} = -\frac{GM(r)\epsilon(r)}{r^2 \left[1 - \frac{2GM(r)}{r}\right]} \left[1 + \frac{p(r)}{\epsilon(r)}\right] \left[1 + \frac{4\pi r^3 p(r)}{\mathcal{M}(r)}\right], \quad (3)$$

$$\frac{d\mathcal{M}}{dr} = 4\pi r^2 \epsilon(r) ; \quad \mathcal{M}(R) = M. \quad (4)$$

Given the equation of state $p = p(\epsilon)$, one can integrate the TOV equations from the origin until the pressure vanishes, $p(R) = 0$. Different equations of state define different types of stars (white dwarfs, neutron stars, strange stars, quark stars, etc) and curves on the mass-radius diagram for the families of stars.

Furthermore, to describe the evolution of the hot plasma created in high-energy heavy ion collisions, one need to make use of hydrodynamics, whose fundamental equations encode the conservation of energy-momentum ($\partial_\mu T^{\mu\nu} = 0$) and of baryon number (or different charges) ($\partial_\mu n_B v^\mu = 0$, with $v^\mu v_\mu = 1$). These represent only five equations for six unknown functions, the additional constraint provided by the equation of state. Hence, it is clear that we really need the equation of state to make any progress.

In principle, we have all the building blocks to compute the equation of state. The Lagrangian of QCD is given, so one would have “simply” to compute the thermodynamic potential, from which one can extract all relevant thermodynamic functions. The fact that the vacuum of QCD is highly nonperturbative, as discussed previously, makes it way more complicated from the outset. As we know, QCD matter becomes simpler at very high temperatures and densities, T and μ playing the role of the momentum

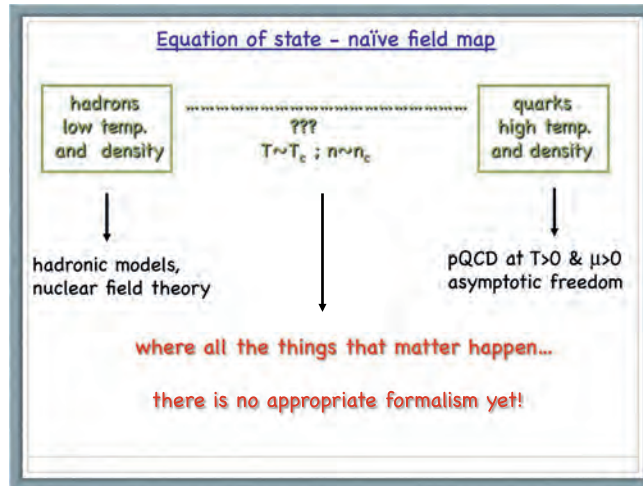


Fig. 3: Cartoon of the naïve field map for the equation of state for strong interactions.

scale in a plasma, but very complicated in the opposite limit. On top of that, T and μ are, unfortunately, not high enough in the interesting cases, so that the physically relevant region is way before asymptotic freedom really kicks in. Perturbative calculations are still an option, but then one has to recall that finite-temperature perturbative QCD is very sick in the infrared, and its naïve formulation breaks down at a scale given by $g^2 T$ [29]. This is known as Linde’s problem: at this scale, for a $(\ell + 1)$ -loop diagram for the pressure, for $\ell > 3$ all loops contribute to the term of order g^6 even for weak coupling [29].

The situation does not look very promising, as illustrated by the cartoon of Fig. 3 which shows that there is no appropriate formalism to tackle with the problem in the physically relevant region for the phase structure, namely the critical regions. However, there are several ways out. Some popular examples being: very intelligent and sophisticated “brute force” (lattice QCD), intensive use of symmetries (effective field theory models), redefining degrees of freedom (quasiparticle models), “moving down” from very high-energy perturbative QCD, “moving up” from hadronic low-energy (nuclear) models. And we can and should also combine these possibilities, as discussed previously.

2 Symmetries of QCD and effective model building

2.1 The simplest approach: the bag model

Before discussing the building of effective models based on the symmetries, or rather approximate symmetries, of QCD, let us consider a very simple description: the MIT bag model [29] applied to describe the thermodynamics of strong interactions.

The model incorporates two basic ingredients, asymptotic freedom and confinement, in the simplest and crudest fashion: bubbles (bags) of perturbative vacuum in a confining medium, including eventual $O(\alpha_s)$ corrections. Asymptotic freedom is implemented by considering free quarks and gluons inside color singlet bags, whereas confinement is realized by imposing that the vector current vanishes on the boundary.

Then, confinement is achieved by assuming a constant energy density for the vacuum (negative pressure), encoded in the so-called bag constant B , a phenomenological parameter extracted from fits to hadron masses. B can also be viewed as the difference in energy density between the QCD and the perturbative vacua. A hadron energy (for a spherical bag) receives contributions from the vacuum and

the kinetic energy, so that its minimum yields

$$E_h^{\min} = \frac{16}{3} \pi R_h^3 B, \quad (5)$$

and the hadron pressure (at equilibrium)

$$p_h = \frac{\partial E_h}{\partial V} = -B + \frac{\text{const}}{4\pi R^4} = 0. \quad (6)$$

Assuming the existence of a deconfining transition, the pressure in the quark-gluon plasma phase within this model is given by

$$p_{\text{QGP}} = \left(\nu_b + \frac{7}{4} \nu_f \right) \frac{\pi^2 T^4}{90} - B, \quad (7)$$

whereas the pressure in the hadronic phase (taking, for simplicity, a pion gas) is given by

$$p_\pi = \nu_\pi \frac{\pi^2 T^4}{90}, \quad (8)$$

neglecting masses. Here, we have the following numbers of degrees of freedom: $\nu_\pi = 3$, $\nu_b = 2(N_c^2 - 1)$ and $\nu_f = 2N_c N_f$ for pions, gluons and quarks, respectively.

For instance, for $N_c = 3$, $N_f = 2$ and $B^{1/4} = 200$ MeV, we obtain the following critical temperature:

$$T_c = \left(\frac{45B}{17\pi^2} \right)^{1/4} \approx 144 \text{ MeV} \quad (9)$$

and a first-order phase transition as is clear from Fig. 4. The value of the critical temperature is actually very good as compared to recent lattice simulations [30], considering that this is a very crude model. On the other hand the nature of the transition, a crossover, is almost by construction missed in this approach.

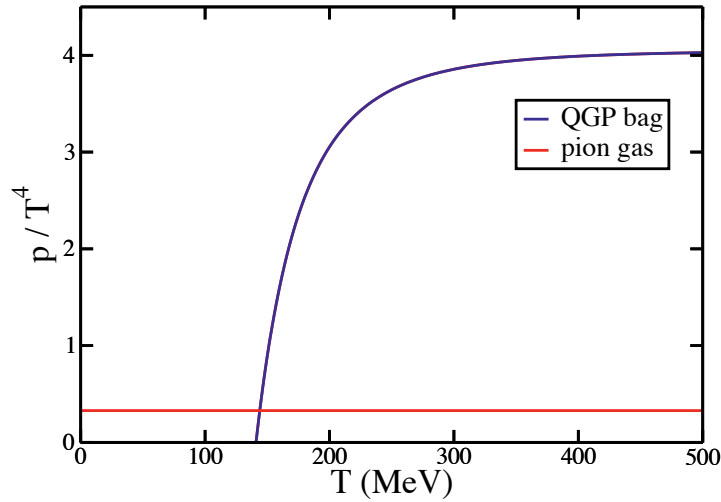


Fig. 4: Pressures in the bag model description.

2.2 Basics of effective model building in QCD

To go beyond in the study of the phases of QCD, one needs to know its symmetries, and how they are broken spontaneously or explicitly. But QCD is very involved. First, it is a non-abelian $SU(N_c)$ gauge theory, with gluons living in the adjoint representation. Then, there are N_f dynamical quarks who live in the fundamental representation. On top of that, these quarks have masses which are all different, which is very annoying from the point of view of symmetries. So, in studying the phases of QCD, we should do it by parts, and consider many “cousin theories” which are very similar to QCD but simpler (more symmetric). In so doing, we can also study the dependence of physics on parameters which are fixed in Nature.

Fig. 5 illustrates the step-by-step process one can follow in assembling the symmetry features present in QCD and learning from simpler theories, as well as cousin theories. Notice that the full theory, whose parameters are given by comparison to the experimental measurements, has essentially no symmetry left. Yet, some symmetries are mildly broken so that a “memory” of them remains. This fact allows us to use “approximate order parameters”, for instance, a concept that is very useful in practice to characterize the chiral and deconfinement transitions.

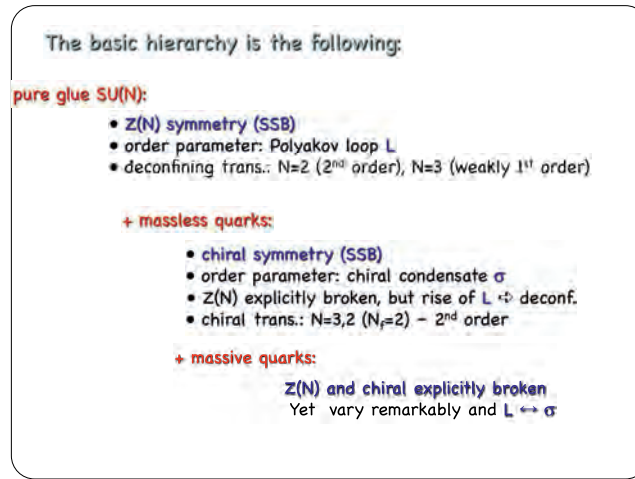


Fig. 5: Basic hierarchy in the step-by-step approach to QCD.

2.3 $SU(N_c)$, $Z(N_c)$ and the Polyakov loop

In the QCD Lagrangian with massless quarks,

$$\mathcal{L} = \frac{1}{2} \text{Tr} F_{\mu\nu} F^{\mu\nu} + \bar{q} i \gamma^\mu D_\mu q, \quad (10)$$

$$D_\mu \equiv (\partial_\mu - ig A_\mu), \quad (11)$$

$$F_{\mu\nu} = \frac{i}{g} [D_\mu(A), D_\nu(A)], \quad (12)$$

we have invariance under local $SU(N_c)$. In particular, we have invariance under elements of the center group $Z(N_c)$ (for a review, see Ref. [31])

$$\Omega_c = e^{i \frac{2n\pi}{N_c} \mathbf{1}}. \quad (13)$$

At finite temperature, one has also to impose the following boundary conditions:

$$A_\mu(\vec{x}, \beta) = + A_\mu(\vec{x}, 0), \quad (14)$$

$$q(\vec{x}, \beta) = -q(\vec{x}, 0). \quad (15)$$

Any gauge transformation that is periodic in τ will do it. However, 't Hooft noticed that the class of possible transformations is more general. They are such that

$$\Omega(\vec{x}, \beta) = \Omega_c, \quad \Omega(\vec{x}, 0) = \mathbf{1}, \quad (16)$$

keeping the gauge fields invariant but not the quarks.

For pure glue this $Z(N_c)$ symmetry is exact and we can define an order parameter - the Polyakov loop:

$$L(\vec{x}) = \frac{1}{N_c} \text{Tr} \mathcal{P} \exp \left[ig \int_0^\beta d\tau \tau^a A_0^a(\vec{x}, \tau) \right], \quad (17)$$

with L transforming as

$$L(\vec{x}) \mapsto \Omega_c L(\vec{x}) \mathbf{1} = e^{i \frac{2n\pi}{N_c}} L(\vec{x}). \quad (18)$$

At very high temperatures, $g \sim 0$, and $\beta \mapsto 0$, so that

$$\langle \ell \rangle = e^{i \frac{2n\pi}{N_c}} \ell_0, \quad \ell_0 \sim 1, \quad (19)$$

and we have a N -fold degenerate vacuum, signaling spontaneous symmetry breaking of global $Z(N_c)$. At $T = 0$, confinement implies that $\ell_0 = 0$. Then, $\ell_0 = 0$ can be used as an order parameter for the deconfining transition:

$$\ell_0 = 0, \quad T < T_c; \quad \ell_0 > 0, \quad T > T_c. \quad (20)$$

Usually the Polyakov loop is related to the free energy of an infinitely heavy test quark via (confinement: no free quark)

$$\langle \ell \rangle = e^{-F_{test}/T}. \quad (21)$$

See, however, the critical discussion in Ref. [31].

The analysis above is valid only for pure glue, i.e. with no dynamical quarks. However, we can still ask whether $Z(3)$ is an approximate symmetry in QCD. On the lattice, in full QCD, one sees a remarkable variation of ℓ around T_c , so that it plays the role of an approximate order parameter [33]. Notice, however, that $Z(3)$ is broken at high, not low T , just the opposite of what is found in the analogous description of spin systems, such as Ising, Potts, etc [13]. The effective potential for the Polyakov loop is illustrated in Fig. 6.

2.4 Adding quarks: chiral symmetry

In the limit of massless quarks, QCD is invariant under global chiral rotations $U(N_f)_L \times U(N_f)_R$ of the quark fields. One can rewrite this symmetry in terms of vector ($V = R + L$) and axial ($A = R - L$) rotations

$$U(N_f)_L \times U(N_f)_R \sim U(N_f)_V \times U(N_f)_A. \quad (22)$$

As $U(N) \sim SU(N) \times U(1)$, one finds

$$U(N_f)_L \times U(N_f)_R \sim SU(N_f)_L \times SU(N_f)_R \times U(1)_V \times U(1)_A, \quad (23)$$

where we see the $U(1)_V$ from quark number conservation and the $U(1)_A$ broken by instantons.

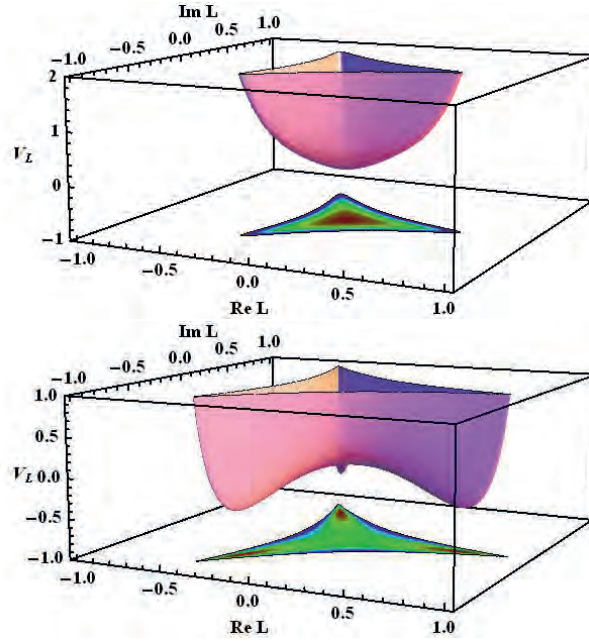


Fig. 6: Effective potential for the Polyakov loop for $T < T_c$ (upper) and $T > T_c$ (lower). Extracted from Ref. [32].

In QCD, the remaining $SU(N_f)_L \times SU(N_f)_R$ is explicitly broken by a nonzero mass term. Take, for simplicity, $N_f = 2$. Then,

$$\mathcal{L} = \frac{1}{4} F_{\mu\nu}^a F^{\mu\nu a} + \bar{\psi}_L \gamma^\mu D_\mu \psi_L + \bar{\psi}_R \gamma^\mu D_\mu \psi_R - m_u (\bar{u}_L u_R + \bar{u}_R u_L) - m_d (\bar{d}_L d_R + \bar{d}_R d_L), \quad (24)$$

so that, for non-vanishing $m_u = m_d$, the only symmetry that remains is the vector isospin $SU(2)_V$. In the light quark sector of QCD, chiral symmetry is just approximate. Then, for massless QCD, one should find parity doublets in the vacuum, which is not confirmed in the hadronic spectrum. Thus, chiral symmetry must be broken in the vacuum by the presence of a quark chiral condensate, so that

$$SU(N_f)_L \times SU(N_f)_R \mapsto SU(N_f)_V, \quad (25)$$

and the broken generators allow for the existence of pions, kaons, etc.

Hence, for massless QCD, we can define an order parameter for the spontaneous breaking of chiral symmetry in the vacuum - the chiral condensate:

$$\langle 0 | \bar{\psi} \psi | 0 \rangle = \langle 0 | \bar{\psi}_L \psi_R | 0 \rangle + \langle 0 | \bar{\psi}_R \psi_L | 0 \rangle, \quad (26)$$

so that this vacuum expectation value couples together the L and R sectors, unless in the case it vanishes. For very high temperatures or densities (low α_s), one expects to restore chiral symmetry, melting the condensate that is a function of T and quark masses and plays the role of an order parameter for the chiral transition in QCD.

Again, the analysis above is valid only for massless quarks. However, we can still ask whether QCD is approximately chiral in the light quark sector. On the lattice (full massive QCD), one sees a remarkable variation of the chiral condensate around T_c , so that the condensate plays the role of an approximate order parameter [33].

In summary, there are two relevant phase transitions in QCD, associated with spontaneous symmetry breaking mechanisms for different symmetries of the action: (i) an approximate $Z(N_c)$ symmetry and deconfinement, which is exact for pure gauge $SU(N_c)$ with an order parameter given by the Polyakov

loop; (ii) an approximate chiral symmetry and chiral transition, which is exact for massless quarks, with an order parameter given by the chiral condensate.

One can try to investigate these phase transitions by building effective models based on such symmetries of the QCD action. Then, the basic rules would be: (i) keeping all relevant symmetries of the action; (ii) trying to include in the effective action all terms allowed by the chosen symmetries; (iii) developing a mimic of QCD at low energy using a simpler field theory; (iv) providing, whenever possible, analytic results at least for estimates and qualitative behavior. Well-known examples are the linear sigma model, the Nambu-Jona-Lasinio model, Polyakov loop models and so on [24]. Although they represent just part of the story, combined with lattice QCD they may provide good insight.

3 A final comment

Instead of conclusions, just a final comment on a point we have already made in the discussion above. To make progress in understanding, or at least in collecting facts about, (de)confinement and chiral symmetry, we need it all: experiments and observations, lattice simulations, theory developments, effective models, and also combinations whenever possible. In that vein, it is absolutely crucial to have theorists and experimentalists working and discussing together.

Acknowledgements

The work of ESF was financially supported by the Helmholtz International Center for FAIR within the framework of the LOEWE program (Landesoffensive zur Entwicklung Wissenschaftlich-Ökonomischer Exzellenz) launched by the State of Hesse.

References

- [1] G. Altarelli, hep-ph/0204179.
- [2] K. Fukushima, J. Phys. G **39** (2012) 013101.
- [3] D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30** (1973) 1343; H. D. Politzer, Phys. Rev. Lett. **30** (1973) 1346.
- [4] T. van Ritbergen, J. A. M. Vermaseren and S. A. Larin, Phys. Lett. B **400** (1997) 379.
- [5] G. 't Hooft, hep-th/0010225.
- [6] C.-N. Yang and R. L. Mills, Phys. Rev. **96** (1954) 191.
- [7] Clay Mathematics Institute, [http : //www.claymath.org/millennium/Yang – Mills_Theory/](http://www.claymath.org/millennium/Yang-Mills_Theory/).
- [8] J. F. Donoghue, E. Golowich and B. R. Holstein, *Dynamics of the standard model*, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **2** (1992) 1.
- [9] S. Pokorski, *Gauge Field Theories* (Cambridge Monographs On Mathematical Physics, 1987).
- [10] F. Wilczek, Phys. Today **53N8** (2000) 22.
- [11] S. Borsanyi, G. Endrodi, Z. Fodor, S. D. Katz and K. K. Szabo, JHEP **1207** (2012) 056.
- [12] E. S. Fraga, Y. Hatta, R. D. Pisarski and J. Schaffner-Bielich, nucl-th/0301062.
- [13] L. E. Reichl, *A Modern Course in Statistical Physics* (Wiley, 2009).
- [14] M. G. Alford, A. Schmitt, K. Rajagopal and T. Schäfer, Rev. Mod. Phys. **80** (2008) 1455.
- [15] Jan Klaers *et al.*, Nature **468** (2010) 545.
- [16] M. A. Stephanov, Prog. Theor. Phys. Suppl. **153** (2004) 139 [Int. J. Mod. Phys. A **20** (2005) 4387].
- [17] E. W. Kolb and M. S. Turner, *The Early Universe*, Front. Phys. **69** (1990) 1; S. Weinberg, *Cosmology* (Oxford University Press, 2008).
- [18] C. Y. Wong, *Introduction to High-Energy Heavy Ion Collisions* (World Scientific, Singapore, 1994). L. P. Csernai, *Introduction to Relativistic Heavy Ion Collisions* (John Wiley and Sons, Chichester, 1994). J. Harris and B. Müller, Ann. Rev. Nucl. Part. Sci. **46** (1996) 71.

- [19] L. McLerran and N. Samios, *T. D. Lee: Relativistic Heavy Ion Collisions and the Riken Brookhaven Center*, BNL-77850-2007-CP.
- [20] *RHIC – Relativistic Heavy Ion Collider*, <http://www.bnl.gov/rhic/>.
- [21] *LHC – Large Hadron Collider*, <http://home.web.cern.ch/about/experiments/>.
- [22] *FAIR – Facility for Antiproton and Ion Research*, <http://www.gsi.de/fair/>.
- [23] *NICA – Nuclotron-based Ion Collider Facility*, <http://nica.jinr.ru/>.
- [24] D. H. Rischke, *Prog. Part. Nucl. Phys.* **52** (2004) 197; K. Yagi, T. Hatsuda and Y. Miake, *Quark-Gluon Plasma: From Big Bang To Little Bang*, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* **23** (2005) 1.
- [25] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, 1972).
- [26] L. F. Palhares, E. S. Fraga and T. Kodama, *J. Phys.* **38** (2011) 085101.
- [27] T. Ullrich, B. Wyslouch and J. W. Harris, *Nucl. Phys. A* **904-905** (2013) pp. 1c.
- [28] N. K. Glendenning, *Compact Stars — Nuclear Physics, Particle Physics, and General Relativity* (Springer, New York, 2000).
- [29] J. I. Kapusta and C. Gale, *Finite-Temperature Field Theory: Principles and Applications* (Cambridge University Press, 2006); M. Le Bellac, *Thermal Field Theory* (Cambridge University Press, 1996).
- [30] O. Philipsen, *Prog. Part. Nucl. Phys.* **70** (2013) 55.
- [31] R. D. Pisarski, hep-ph/0203271.
- [32] A. J. Mizher, M. N. Chernodub and E. S. Fraga, *Phys. Rev. D* **82** (2010) 105016.
- [33] S. Borsanyi, S. Durr, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, D. Nogradi and K. K. Szabo *et al.*, *JHEP* **1208** (2012) 126.

An introduction to cosmology

K. E. Kunze

Departamento de Física Fundamental and IUFFyM, Universidad de Salamanca, Salamanca, Spain

Abstract

Cosmology is becoming an important tool to test particle physics models. We provide an overview of the standard model of cosmology with an emphasis on the observations relevant for testing fundamental physics.

Keywords

Cosmology; lectures; cosmological model; dark matter; dark energy.

1 Introduction

Cosmology is the only part of physics which has the whole universe as its area of research. As such it covers a vast range of scales. Energy scales go from the present day temperature of 10^{-4} eV upto the Planck scale 10^{19} GeV. It aims to describe the evolution of the universe from its very beginning upto today where it has an estimated age of the order of 10^{10} years. Due to its very nature of understanding the universe as a whole cosmology needs input from very different areas of physics. These naturally include astrophysics and theories of gravitation, but also plasma physics, particle physics and experimental physics.

Over the last two decades cosmology has entered a data driven era. A turning point were the first observations with upto then unprecedented precision of the cosmic microwave background (CMB) with the COBE satellite in 1990 [1–3]. Since then there have been a variety of CMB experiments, ground based such as the Atacama Cosmology telescope (ACT) [4] in Chile, the South Pole telescope (SPT) [5] at the South Pole, with balloons such as Boomerang [6], two more satellites, namely the Wilkinson Microwave Anisotropy Probe (WMAP) [7] and Planck [8] and more experiments are planned for the future. The first systematic study of the structure in the local universe was the Center for Astrophysics (CfA) survey of galaxies [9]. One of the most recent ones is the Sloan Digital Sky Survey III [10] with SDSS IV [11] already underway. In 2019 the launch of the ESA mission EUCLID is planned. Its goal is to measure shapes and redshifts of galaxies upto redshifts of $z \sim 2$ thereby allowing to determine the evolution of the recent universe since the time when dark energy became important [12].

In these lectures we will start with the evolution of the universe on very large scales where it is, to a high degree, isotropic and homogeneous. This will be followed by a description of the thermal history of the universe from very early times upto the present epoch including the key events. Observations show that the universe is not perfectly isotropic. At the largest scales this manifests itself in the small temperature anisotropies $\frac{\Delta T}{T} \sim \mathcal{O}(10^{-5})$ in the CMB. These are an imprint of the density perturbations which provided the seeds from which all large scale structure such as galaxies have developed. Therefore the second part of the lectures is dedicated to the inhomogeneous universe, the origin of the temperature anisotropies and polarization of the CMB and large scale structure. The third and last part deals with the two big unknowns in our universe. Different observations such as from the CMB, large scale structure and high redshift supernovae pinpoint the cosmological parameters to around 4% baryonic matter, 25% of cold dark matter and about 71% of dark energy. From the data the physical properties of these components can be constrained. However, upto the present it is neither known what constitutes dark matter nor dark energy. There are many proposals, some of which are rather exotic, but none stands out as a "natural" model.

There are already a number of excellent text books which cover different aspects of these lectures, e.g., [13] [14], [15], [16], [17], [18], [19], [20]. Also the review sections related to cosmology in [21] provide a very good overview.

2 The homogeneous universe

Observations such as the high degree of isotropy of the CMB indicate that globally the universe is well described by a spatially homogeneous and isotropic model. These are the Friedmann-Robertson-Walker solutions of general relativity to which a brief introduction can be found in appendix A. Spatial homogeneity and isotropy mean that physical conditions are the same everywhere and in each direction. It allows to choose a coordinate system such that the four dimensional space time is described by a foliation of spatial hypersurfaces at constant time and the metric is given by

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right]. \quad (1)$$

The parameter k labels the different choices of spatial curvature. It takes the values $k = 0$ for a flat universe, $k = +1$ for a closed universe and $k = -1$ for an open universe. The coordinates r , ϕ and θ determine the spatial *comoving* coordinates on each constant time slice. These coordinates do not change during the evolution of the universe or, in other words, from one constant time slice to the next one. However, as the universe is not static this leads naturally to the notion of *physical* coordinates as well as *physical* scales. As an example, consider two nearby observers (or galaxies or any other astrophysical object) and assume that at some fixed time t_1 they are separated by a distance ℓ_1 . Because of the expansion of the universe all *physical* scales are multiplied by the scale factor $a(t)$ so that $\ell_1 = a(t_1)\ell_{com}$. Hence, at some later time t_2 the *physical* distance between the two objects is given by $\ell_2 = a(t_2)\ell_1/a(t_1)$. Quite often the scale factor today is set to one, $a(t_0) = 1$, in which case at present comoving and physical scales coincide¹. Here we have introduced another common choice, i.e. to denote the present epoch by an index "0". Choosing the coordinate system accordingly the physical radial distance is given by

$$d_p = \int_0^r \frac{adr'}{(1 - kr'^2)^{\frac{1}{2}}} \equiv a(t)f(r). \quad (2)$$

Thus the physical distance between two objects changes locally at a rate $v_p = \dot{a}f(r) = \frac{\dot{a}}{a}d_p(t) = H(t)d_p(t)$, where $H \equiv \frac{\dot{a}}{a}$ is the Hubble parameter. In this section a dot indicates the derivative w.r.t. cosmic time t . Applied to the present epoch $v_p = H_0d_p$ which is also known as Hubble's law. In an expanding universe this is a recession velocity which was first observed in galaxies by Edwin Hubble in 1929. Present observations give a value of the Hubble constant close to $70 \text{ km s}^{-1}\text{Mpc}^{-1}$. There are some variations in the value of H_0 depending on which data are used resulting in a certain tension between different data sets. For example, from observations of supernovae in combination with Cepheid variables $H_0 = (73.8 \pm 2.4) \text{ km s}^{-1}\text{Mpc}^{-1}$ [22] and from the Planck 15 temperature data combined with the Planck 15 gravitational lensing reconstruction the Hubble parameter is found to be $H_0 = (67.8 \pm 0.9) \text{ km s}^{-1}\text{Mpc}^{-1}$ [23].

An important question is how this recession velocity can actually be measured. The answer lies with the observation of cosmological redshift. Consider a distant galaxy at radial coordinate r_1 which emits light at some time t_1 at, say, wavelength λ_e which is observed by an observer at $r = 0$ at the present time t_0 at a wavelength λ_o . Light travels along null geodesics so that $ds^2 = 0$ implying for radial null geodesics

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_0^{r_1} \frac{dr}{(1 - kr^2)^{\frac{1}{2}}} = f(r_1). \quad (3)$$

Light emitted at a time $t_1 + \delta t_1$ reaches the detector at a time $t_0 + \delta t_0$. Since $f(r_1)$ is a constant and it is assumed that the source has no peculiar motion,

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_{t_1 + \delta t_1}^{t_0 + \delta t_0} \frac{dt}{a(t)} \Rightarrow \int_{t_1}^{t_1 + \delta t_1} \frac{dt}{a(t)} = \int_{t_0}^{t_0 + \delta t_0} \frac{dt}{a(t)}. \quad (4)$$

¹Note, however, that in this case the curvature parameter k has to be appropriately rescaled in the case of a non flat universe.

Thus for small $\delta t_i, i = 0, 1, (\lambda c \delta t_i \ll ct_i)$ and approximating $a(t) \sim \text{const.}$ during the time interval of integration it follows that $\delta t_1/a(t_1) = \delta t_0/a(t_0)$ so that

$$1 + z \equiv \frac{\lambda_0}{\lambda_1} = \frac{a(t_0)}{a(t_1)} \quad (5)$$

where z is the redshift. This cosmological redshift is a direct consequence of the expansion of the universe. In case, the scale factor is diminishing a blue shift is observed. In Fig. 1 an example of the observation of cosmological redshift in the spectrum of a galaxy at redshift $z = 0.1437$ together with a reference spectrum of a star at $z = 0$ is shown.

It is difficult to measure distances at very large scales even more so because of the expansion of the universe. In cosmology there are two important distance measures which are the luminosity distance and the angular diameter distance. These rely on the fact that the flux and the angular size of an object could in principle be known. This means that if the observer has independent knowledge of its absolute luminosity or its physical size its distance can be estimated. Objects whose absolute luminosity or physical size are available define the classes of standard candles or standard rulers, respectively. Distances obtained for these objects are the luminosity distance and angular diameter distance, respectively. The luminosity distance D_L is defined by

$$D_L = \left(\frac{L}{4\pi\ell} \right)^{\frac{1}{2}}, \quad (6)$$

where L is the absolute luminosity and ℓ is the visible luminosity which is received by the observer. To determine its evolution with redshift consider a source located at a point P at a coordinate distance r and emitting a signal at some time t , as illustrated in Fig. 2. An observer at a point P_0 observes the signal at a time t_0 . The absolute luminosity is the energy flux (=energy/time) and the visible luminosity ℓ is the energy flux density (=energy/(time \times surface)). The rates of emission and reception of photons are related by $(\delta t_0)^{-1}/(\delta t_e)^{-1} = a(t)/a_0$. Thus the absolute luminosity $L = E_e/\delta t_e = hc/\lambda_e \delta t_e$ and the visible luminosity $\ell = E_0/(4\pi a_0^2 r^2 \delta t_0) = hc/(4\pi a_0^2 r^2 \lambda_0 \delta t_0)$ resulting in

$$D_L = a_0 r (1 + z). \quad (7)$$

Hubble's law implies locally that an object moves away from the observer with a velocity proportional to its physical distance. Expanding the scale factor beyond linear order leads to

$$D_L \simeq H_0^{-1} \left[z + \frac{1}{2}(1 - q_0)z^2 + \dots \right] \quad (8)$$

where q_0 is the deceleration parameter today which in general is defined by $q \equiv -\frac{a\ddot{a}}{\dot{a}^2}$. In order to determine the luminosity distance it is necessary to find objects with a known absolute luminosity which can be used as standard candles. Cepheids provide one example of standard candles which have been used frequently in the past. These are variable stars with a regular change in their apparent magnitude whose period is related to their absolute luminosity. Another class of standard candles are Type Ia supernovae (SN Ia) which have a well known light profile (cf. e.g [25], [26]). They are found in binary systems formed when one of the stars is accreting material from the other star. Reaching a critical mass limit leads to a thermonuclear explosion and subsequently to a sudden increase in the observed light curve. These light curves are well known and can be observed at cosmological distances as they can reach luminosities upto 10^{10} times the luminosity of the sun. The observations of SN Ia have been particularly important in establishing that the expansion of the universe is accelerating in the current epoch.

The other important distance measure is the angular diameter distance D_A which generalizes the concept of the parallax. It is defined by requiring that the angle θ over which the object extends is

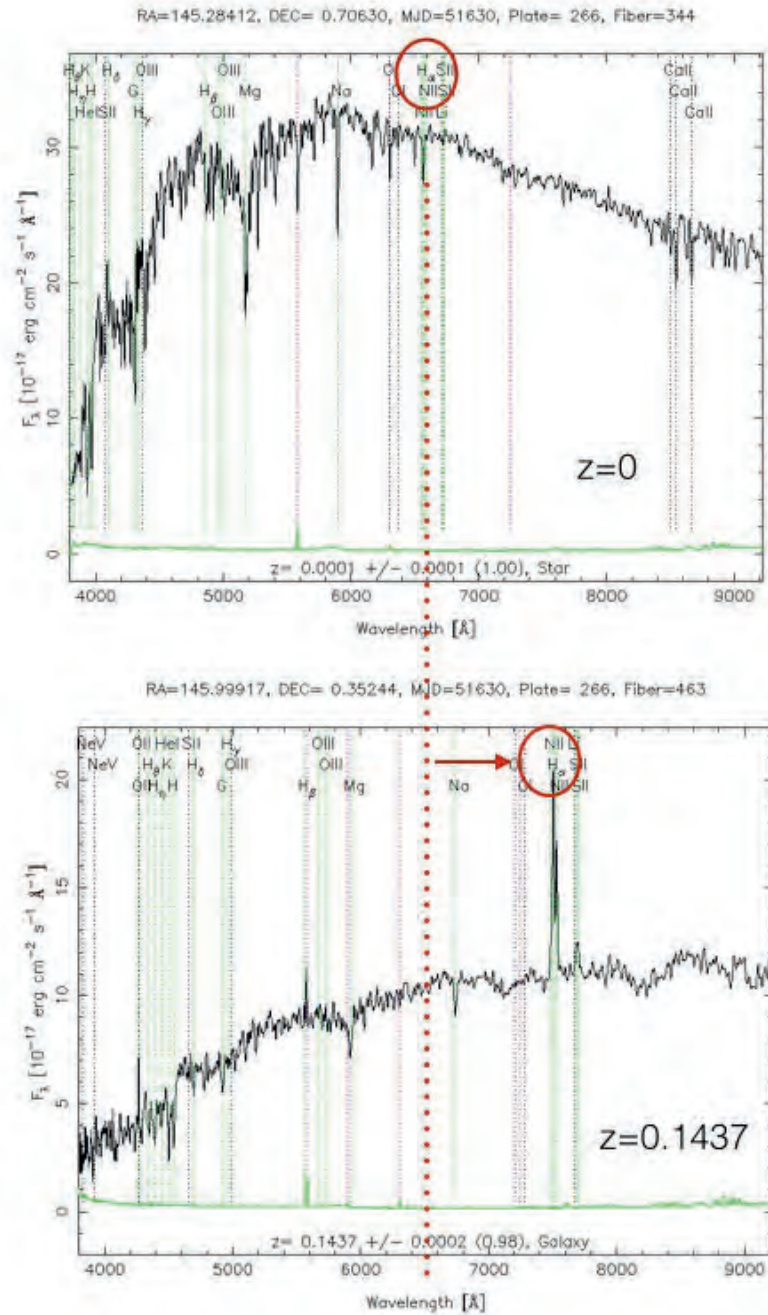


Fig. 1: Observing the cosmological redshift. *Upper panel:* A reference spectrum of a star at redshift $z = 0$. *Lower panel:* Spectrum of galaxy SDSS J094359.79+002108.7 at redshift $z = 0.1437$. Marked by a red circle is in both spectra the H_{α} line of the Balmer series of hydrogen ($\lambda = 6564.7\text{\AA}$ at rest, i.e. at $z = 0$). The red arrow in the lower panel indicates the shift of the H_{α} line (as well as the other spectral lines) towards the long wavelength part of the spectrum due to the expansion of the universe, implying that in this case a cosmological redshift is observed. Whereas in the case of the reference spectrum (*upper panel*) the line is an absorption line in the case of the spectrum of the galaxy (*lower panel*) it is an emission line. The figures have been done using spectra taken from Data Release 7 (DR7) of the Sloan Digital Sky Survey, SDSS-II [24]. The spectra have been obtained from <http://skyserver.sdss.org/dr7/en/tools/>.

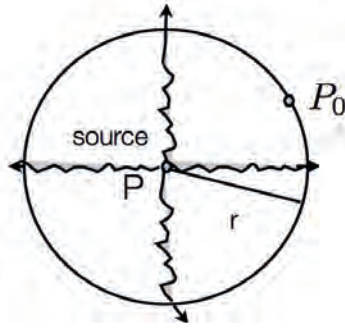


Fig. 2: A point source situated at P is emitting radiation isotropically which is received by an observer at P_0 at a coordinate distance r . Knowing the absolute luminosity of the point source allows the observer to determine its luminosity distance D_L .

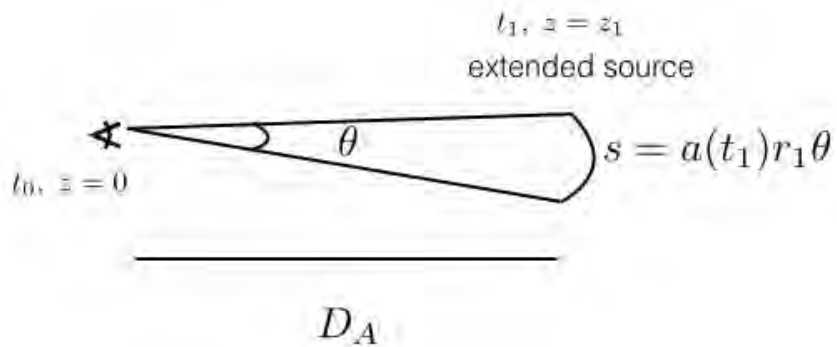


Fig. 3: The angular diameter distance D_A is defined in such a way that $\theta = s/D_A$ where $s = a(t_1)r_1\theta$ is the physical extension of the object.

given, just as in Euclidean space, that is, by the ratio of the transverse size s over the distance D_A to the object (cf Fig. 3). Assuming that the extended source at a redshift z_1 is emitting light at a time t_1 and is located at a comoving radial distance r_1 then for small angles θ , the angular diameter distance is given by $D_A = a(t_1)r_1$. Comparing this with the expression of the luminosity distance D_L (cf. Eq. (7)) it is found that

$$\frac{D_A}{D_L} = (1 + z)^{-2}. \tag{9}$$

The angular diameter distance D_A is only useful if the angular extension of the source is known which in general is a difficult task to measure at cosmological distances. As such it relies on a standard ruler. As will be seen in the next section such a standard ruler does exist in the early universe and is imprinted in the observed angular power spectrum of the temperature fluctuations of the cosmic microwave background.

Due to the high degree of symmetry of the Friedmann-Robertson-Walker (FRW) solutions the description of the evolution of the universe becomes particularly simple. The evolution of the background geometry is encapsulated in one time dependent function which is the scale factor $a(t)$. In the physically relevant cases matter is described by a perfect fluid determined by its energy density $\rho(t)$ and pressure $p(t)$. These are related by an equation of state which has a rather simple form, $p = w\rho$, where w is a constant. In this case the evolution of the background is described by the Friedmann equations

$$H^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\rho \tag{10}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) \tag{11}$$

where $H = \frac{\dot{a}}{a}$ is the Hubble parameter. The evolution of matter is determined by

$$\dot{\rho} + 3H(\rho + p) = 0. \quad (12)$$

In the standard model of cosmology the universe is initially very hot and dense before cooling down as it expands. Thus initially it is dominated by radiation, subsequently by non relativistic matter and at present day by an effective cosmological constant, otherwise also called dark energy which will be discussed in more detail in section 3. These different epochs are described by a perfect fluid with equation of state with $w = \frac{1}{3}$ for radiation, $w = 0$ for non relativistic matter and $w = -1$ for an effective cosmological constant. The energy density scales as $\rho \sim a^{-4}$ for a radiation dominated universe and as $\rho \sim a^{-3}$ for a matter dominated universe. In a flat universe ($k = 0$) it is found that the scale factor evolves as $a \sim t^{\frac{1}{2}}$ in a radiation dominated universe and as $a \sim t^{\frac{2}{3}}$ in a matter dominated universe. An important quantity in cosmology is the dimensionless density parameter Ω which is defined by

$$\Omega \equiv \frac{\rho}{\rho_{crit}} \quad (13)$$

where $\rho_{crit} \equiv \frac{3H^2}{8\pi G}$ is the critical energy density which corresponds to the total energy density of a flat FRW universe. Using this definition the Friedmann equation takes the form

$$\frac{k}{H^2 a^2} = \Omega - 1. \quad (14)$$

Moreover, including the different contributions to the total energy density explicitly, then

$$H^2 = \frac{8\pi G}{3}\rho_R + \frac{8\pi G}{3}\rho_M - \frac{k}{a^2} + \frac{\Lambda}{3} \quad (15)$$

which, with the appropriate definitions, can be written as a constraint equation on the sum of the density parameters,

$$\Omega_r + \Omega_m + \Omega_\Lambda - \Omega_K = 1. \quad (16)$$

Here $\Omega_R = \Omega_\gamma + \Omega_\nu$ includes radiation (photons) $\Omega_\gamma = \rho_\gamma/\rho_{crit}$ and relativistic matter such as light neutrinos $\Omega_\nu = \rho_\nu/\rho_{crit}$. Non relativistic matter $\Omega_m = \Omega_b + \Omega_c$ has contributions from baryons $\Omega_b = \rho_b/\rho_{crit}$ and cold dark matter $\Omega_c = \rho_c/\rho_{crit}$. The cosmological constant term or dark energy is $\Omega_\Lambda = \Lambda/3H^2$ and the curvature term $\Omega_K = k/(a^2 H^2)$. Best fit values of 68% confidence limits on the present day values of the density parameters from observations of the cosmic microwave background give for the standard 6-parameter Λ CDM model from the Planck 15 data [23]: $H_0 = (67.31 \pm 0.96) \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.315 \pm 0.013$, $\Omega_\Lambda = 0.685 \pm 0.013$. Including spatial curvature as a free parameter gives the 95% limit $\Omega_K = -0.052^{+0.049}_{-0.055}$. In Fig. 4 the distribution of the main contributions are shown for Planck 15. In section 3 we will discuss the nature of these contributions of which only the baryonic contribution is understood which only makes up about 5% of the total energy density in the universe (cf Fig. 4).

The standard model of cosmology describes the evolution from its very early stages upto the present. The underlying model is the rather simple flat Friedman-Robertson-Walker model which in a sense is quite remarkable. The Friedman-Robertson-Walker solutions for a standard type of matter such as relativistic or non relativistic particles or radiation generally have an initial curvature singularity which is a consequence of the Penrose-Hawking singularity theorems. In the standard model of cosmology, also known as the big bang model, the universe evolves from a very tiny, very hot initial state to the present day very large and rather cold state as indicated by the observed temperature of the cosmic microwave background of 2.73 K. In the following we will give a brief overview of the milestones in the evolution of the universe (cf., e.g., [27]). Before times marked by the Planck time $t \sim 10^{-43} \text{ s}$ nothing is really known since general relativity as a classical theory is not valid anymore. At later times, for temperatures below

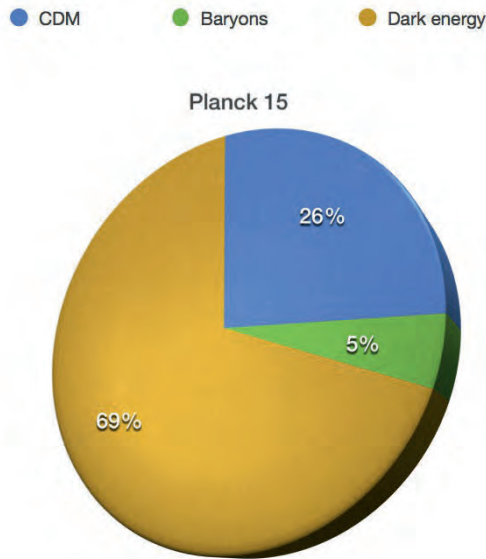


Fig. 4: Distribution of contributions to the total energy density at present from Planck 15 best fit parameter.

$T < 10^{19}$ GeV all particles of the standard model of particle physics constitute the primordial, relativistic plasma. At temperatures 100-300 MeV, at a time of the order of 10^{-5} s the quark-gluon phase transition takes place and free quarks and gluons combine to form baryons and mesons. The next important epoch is primordial nucleosynthesis when the universe has cooled down to temperatures between 10 and 0.1 MeV, corresponding to times between 10^{-2} to 10^2 s. During this time light elements are produced such as hydrogen, helium-4, deuterium, lithium and other light elements. Observations of the primordial abundances of these elements provide the possibility to test the standard model of cosmology very deep into its very early evolution. As will be discussed below in more detail, during this epoch light, standard model neutrinos decouple at around 1 MeV from the rest of the cosmic plasma. At around 1 eV, at a time 10^{11} s, matter-radiation equality takes place when the energy densities in relativistic (“radiation”) and non relativistic matter (“matter”) are equal. At this moment the evolution of the universe changes from radiation dominated to matter dominated. Recombination takes place at around 10^{12} to 10^{13} s when the plasma cooled down enough so that electrons and nuclei form neutral atoms. Shortly afterwards the photons decouple from the cosmic plasma, the universe becomes transparent to radiation which today is observed as the cosmic microwave background (CMB). The temperature of the CMB is not perfectly isotropic but there are small deviations from the background isotropy. Moreover, the radiation is linearly polarized (which one would not expect from Thomson scattering of a random, isotropic ensemble of electrons). These CMB anisotropies can be used to test models of the very early universe as will be discussed in the next section. The temperature fluctuations in the CMB are caused by perturbations in the density field. At later times, under the influence of gravity, some of these fluctuations become gravitationally unstable leading to the first generation of stars which marks the beginning of galaxy and large scale structure formation. This first generation of stars are thought to be important in the reionization of the intergalactic medium (IGM). The ionization state of the IGM can be determined from the observation of light from distant quasars which are active galactic nuclei with extremely large luminosities. The presence of neutral hydrogen in the IGM along the line of sight can be detected by the observation of Ly α absorption at wavelengths shorter than $(1 + z_*)\lambda_{Ly\alpha}$ in the spectrum of the quasar assumed to be at a redshift z_* . Moreover, $\lambda_{Ly\alpha} = 1216 \text{ \AA}$ is the laboratory wavelength corresponding to the Ly α absorption line. Observations show that this absorption line is absent upto redshifts $z \simeq 6$ indicating that the IGM has been completely reionized below this redshift (c.f., more details in section 2).

The description of physical processes in the early universe such as decoupling and recombination requires to understand the thermodynamics of the early universe. Early on, during the radiation dominated era the universe is dense and hot. Thermal equilibrium is established by interactions of particles rapid in comparison with the typical time scale of expansion of the universe. In thermodynamical equilibrium the number density n , the energy density ρ and the pressure p of a gas of particles without strong interactions with g internal degrees of freedom is determined in terms of the distribution function $f(\vec{p})$ in phase space by (cf. e.g., [19])

$$n = \frac{g}{(2\pi)^3} \int f(\vec{p}) d^3p \quad (17)$$

$$\rho = \frac{g}{(2\pi)^3} \int E(\vec{p}) f(\vec{p}) d^3p \quad (18)$$

$$p = \frac{g}{(2\pi)^3} \int \frac{|\vec{p}|}{3E} f(\vec{p}) d^3p, \quad (19)$$

where $E^2 = |\vec{p}|^2 + m^2$ with m and \vec{p} the mass and 3-momentum of each particle, respectively. In kinetic equilibrium the distribution function is given by either the Fermi-Dirac or Bose-Einstein statistics. Hence

$$f(\vec{p}) = [\exp((E - \mu)/T) \pm 1]^{-1} \quad (20)$$

where μ is the chemical potential, +1 refers to the Fermi-Dirac distribution and -1 to Bose-Einstein distribution. If the particles are in chemical equilibrium and, say, species i interacts with species j, k, l such that $i + j \leftrightarrow k + l$ then $\mu_i + \mu_j = \mu_k + \mu_l$. The chemical potential of the photons is set to zero, $\mu_\gamma = 0$. Therefore since a particle (p) and its antiparticle (\bar{p}) annihilate to photons the chemical potential satisfy $\mu_p = -\mu_{\bar{p}}$. In the relativistic limit $T \gg m, T \gg \mu$ the thermodynamical quantities are given by

$$\rho = \frac{7}{8} \frac{\pi^2}{30} g T^4 \quad (21)$$

$$n = \frac{3}{4} \frac{\zeta(3)}{\pi^2} g T^3 \quad (22)$$

for fermions and

$$\rho = \frac{\pi^2}{30} g T^4 \quad (23)$$

$$n = \frac{\zeta(3)}{\pi^2} g T^3 \quad (24)$$

for bosons and $\zeta(3) = 1.20206$ is the value of the Riemann zeta function. In both cases $p = \frac{1}{3}\rho$ is obtained. In the non relativistic limit $m \gg T$, for fermions as well as bosons it is found that

$$n = g \left(\frac{mT}{2\pi} \right)^{\frac{3}{2}} e^{-\frac{m-\mu}{T}} \quad (25)$$

$$\rho = nm \quad (26)$$

$$p = nT, \quad (27)$$

where the last relation implies $p \ll \rho$. Obviously, all N particle species in thermodynamical equilibrium contribute to the total value of the corresponding thermodynamical quantities, so that, e.g., for the energy density,

$$\rho_R = T^4 \sum_{i=1}^N \left(\frac{T_i}{T} \right)^4 \frac{g_i}{2\pi^2} \int_{x_i}^{\infty} \frac{(u^2 - x_i^2)^{\frac{1}{2}} u^2 du}{\exp(u - y_i) \pm 1} \quad (28)$$

where $u \equiv \frac{E}{T_i}$, $x_i \equiv \frac{m_i}{T_i}$ and $y_i \equiv \frac{\mu_i}{T_i}$. Moreover T is the photon temperature. The contributions from non relativistic particle species are subleading in comparison to the contributions from the relativistic ones. Therefore to a good approximation the total energy density is given by

$$\rho_R = \frac{\pi^2}{30} g_* T^4 \quad (29)$$

and similarly the total pressure $p_R = \frac{1}{3} \rho_R = \frac{\pi^2}{90} g_* T^4$ where g_* counts the total number of relativistic degrees of freedom, $m_i \ll T$,

$$g_* = \sum_{i=\text{bosons}} g_i \left(\frac{T_i}{T} \right)^4 + \frac{7}{8} \sum_{i=\text{fermions}} g_i \left(\frac{T_i}{T} \right)^4. \quad (30)$$

This is a function of temperature and hence of time as massive species are relativistic at high enough temperatures, $T \gg m_i$ but become non relativistic once the temperature drops below their rest mass $T \ll m_i$ at which point they stop contributing. At $T \ll 1$ MeV the only relativistic species are 3 species of light neutrinos and the photon. As will be discussed in detail below, at that time photons and neutrinos do not have the same temperature, but rather satisfy $T_\nu = \left(\frac{4}{11} \right)^{\frac{1}{3}} T_\gamma$. Thus $g_*(T \ll \text{MeV}) = 2 \times \frac{7}{8} \times 3 \left(\frac{4}{11} \right)^{\frac{4}{3}} + 2 = 3.36$ assuming Dirac neutrinos which introduces the factor 2 since neutrinos and antineutrinos contribute. For temperatures above 1 MeV also positrons and electrons, each contributing two degree of freedom, are relativistic so that $g_* = 2 + \frac{7}{8}(2 + 2 + 2 \times 3) = \frac{43}{4} = 10.75$. At temperatures above 300 GeV all species of the standard model are relativistic and g_* is of the order of 100. During the radiation dominated epoch $\rho = \rho_R$. Hence using Eqs. (10) and (29) yields to

$$H^2 = 1.66 g_*^{\frac{1}{2}} \frac{T^2}{M_P}, \quad (31)$$

where $M_P = 1.22 \times 10^{19}$ GeV is the Planck mass. Primordial nucleosynthesis or big bang nucleosynthesis (BBN) predicts the abundances of hydrogen and of light elements such as deuterium, helium-3, helium-4, lithium-7 which were synthesized at the end of the first three minutes of the universe. These predictions are in good agreement with observations. BBN provides important constraints on possible deviations from the standard model of cosmology and on new physics Beyond the Standard Model (BSM). One of the key quantities is the ratio of number density of neutrons over the number density of the protons, denoted $\frac{n}{p}$. In thermodynamical equilibrium it is given by

$$\frac{n}{p} = e^{-\frac{Q}{T}} \quad (32)$$

where Q is the difference in the rest masses of neutrons and protons, $Q \equiv m_n - m_p = 1.293$ MeV. Very early on, at very high temperatures, $T \gg 1$ MeV, weak interactions are very efficient so that $\frac{n}{p} = 1$. The neutron-proton interconversion rate is given by $\Gamma_{n \leftrightarrow p} \sim G_F^2 T^5$ where $G_F = 1.166 \times 10^{-5} \text{GeV}^{-2}$ is the Fermi constant. At around a temperature $T_D = 0.7$ MeV this conversion rate drops below the expansion rate determined by the Hubble parameter, $H \sim \sqrt{g_*} T^2 / M_P$. Using Eq. (32) it follows that the neutron to proton ratio at freeze-out at around 0.7 MeV is, $\frac{n}{p} = \frac{1}{6}$. After freeze-out this ratio is only changed by the decay of the free neutrons, $n \rightarrow p + e + \nu_e$. The life time of a free neutron is $\tau = (885.7 \pm 0.8)$ s. After some time t the number density of neutrons is given by $n_n(0) e^{-\frac{t}{\tau}}$. Assuming t to be of the order of 10^2 s the final value of $\frac{n}{p}$ is given by $\frac{n}{p} = \frac{1}{7}$. Using this together with the fact that nearly all neutrons end up in helium-4 yields an estimate of the primordial mass fraction Y_p ,

$$Y_p = \frac{2 \frac{n}{p}}{1 + \frac{n}{p}} \simeq 0.25. \quad (33)$$

The observed value is $Y_p = 0.2465 \pm 0.0097$. The nucleosynthesis chain begins with the formation of deuterium in the process $p(n, \gamma)D$. Because of the low density only 2-body reactions, such as $D(p, \gamma)He^3$, $He^3(D, p)He^4$ are important. The abundances of elements other than helium-4 produced in BBN are in comparison much smaller. Primordial abundances can be observed in metal low regions where the abundances of light elements have not been changed by stellar nucleosynthesis (cf. e.g., [28], [29], [30]).

As described above the Big Bang model fares well with observations. However, there are certain short comings (e.g., [27, 31]). These include the flatness problem, that is the observation that at present day the universe is flat. For example, Planck 15 found that $|\Omega_K| < 0.005$ [23]. A simple calculation shows that this requires a severe fine tuning of the initial conditions at the beginning of standard cosmology. For the sake of argument assume the universe to be matter dominated throughout its evolution. Then with Eq. (14)

$$|\Omega - 1| = \frac{|k|}{a^2 H^2} \sim t^{\frac{2}{3}}. \quad (34)$$

Assuming that the total density parameter today Ω_0 is in the range between 10^{-2} and 10 requires that at the time of BBN at around 1s,

$$|\Omega - 1|_{\text{BBN}} = |\Omega - 1|_0 \left(\frac{t_0}{t_{\text{BBN}}} \right)^{-\frac{2}{3}} \Rightarrow |\Omega - 1|_{\text{BBN}} < 10^{-11}. \quad (35)$$

This indicates that a slight change in the initial conditions leads to a completely different universe. In addition, the horizon problem encapsulates the fact that domains which have been in causal contact have a limited size. The physical horizon distance is given by

$$d_H(t) = a(t) \int_{t_i}^t \frac{dt'}{a(t')} \sim H(t)^{-1} \quad (36)$$

which determines the maximal separation of two points in causal contact. Going back to the time of decoupling shows that at this moment the universe consisted of many causally disconnected domains. This raises the question of how to explain that the amplitude of the fractional temperature fluctuations of the CMB observed in different directions in the sky is of the order of $\frac{\Delta T}{T} \sim 10^{-5}$. These small fluctuations also pose a problem in itself for the Friedmann-Robertson-Walker models which are exactly homogeneous and isotropic. However, for large scale structure formation initial density fluctuations are a necessary ingredient. Therefore the observations of the CMB anisotropies provide a key piece in the puzzle of how galaxies and other structures formed in our universe.

The inflationary paradigm solves these problems of the Big Bang model. Inflation is an era before the beginning of standard cosmology. It is defined as a stage of accelerated expansion of the universe, $\ddot{a} > 0$. De Sitter space-time which is characterized by only a positive cosmological constant and no additional energy density contribution is a typical example of an inflationary solution. In this case $H = \sqrt{\frac{\Lambda}{3}} = \text{const.}$ and effectively, $p = -\rho$. Moreover, in this case the scale factor evolves from some initial time t_i , as

$$a(t) = a(t_i) e^{H(t-t_i)}. \quad (37)$$

The flatness problem is resolved by noticing that in this case the total density parameter is driven towards unity,

$$|\Omega - 1| = \frac{|k|}{a^2 H^2} \sim |k| e^{-2Ht} \rightarrow 0. \quad (38)$$

The solution to the horizon problem is provided by the observation that the exponential behaviour of the scale factor allows to enlarge one causal domain sufficiently within a finite amount of time. Thus the

observable universe originates from one patch in causal contact in the very early universe. The origin of the density fluctuations are assumed to be quantum fluctuations of a scalar field. One of the simplest and most studied models of inflation is driven by the potential energy $V(\phi)$ of such a scalar field ϕ [31]. Assuming that the only matter in the universe is given by this scalar field implies that the energy density and pressure are given by,

$$\rho = \frac{\dot{\phi}^2}{2} + V(\phi) \quad (39)$$

$$p = \frac{\dot{\phi}^2}{2} - V(\phi). \quad (40)$$

Going back to Eq. (11) it can be checked that $\ddot{a} > 0$ since $\rho + 3p < 0$ in this case. The evolution of the scalar field is determined by the Klein-Gordon equation,

$$\ddot{\phi} + 3H\dot{\phi} = -\frac{dV}{d\phi}. \quad (41)$$

The Friedmann equation takes the form

$$H^2 = \frac{8\pi}{3M_P^2} \left(\frac{\dot{\phi}^2}{2} + V(\phi) \right). \quad (42)$$

One important class of inflationary models uses the slow roll approximation which consists in neglecting the $\ddot{\phi}$ term in Eq. (41) as well as the kinetic energy term in the Friedmann equation, so that

$$3H\dot{\phi} = -\frac{dV}{d\phi} \quad (43)$$

$$H^2 = \frac{8\pi}{3M_P^2} V(\phi). \quad (44)$$

This results in the expansion of the universe being driven by the potential energy of the scalar field which is also called the inflaton. There is no clear candidate from particle physics as to what could be the inflaton. It remains one of the challenges to connect inflation with fundamental particle physics. The potential is typically chosen to be an even power of ϕ and the picture is that the scalar field is slowly rolling down its potential. The possibility for slow roll inflation is determined by the shape of the potential which is encoded in the slow roll parameters (cf. e.g. [13]),

$$\epsilon(\phi) = \frac{M_P^2}{16\pi} \left(\frac{V'}{V} \right)^2 \quad (45)$$

$$\eta(\phi) = \frac{M_P^2}{8\pi} \frac{V''}{V}. \quad (46)$$

The conditions for slow roll inflation are $\epsilon \ll 1$ and $|\eta| \ll 1$. The field value at the end of inflation is determined by the condition that the modulus of at least one of the slow parameters reaches unity. The duration and hence the amount of inflation is measured by the number of e-folds,

$$N(t) = \ln \frac{a(t_{fin})}{a(t)} \quad (47)$$

where t_{fin} denotes the end of inflation. In slow roll inflation this can be expressed as a function of ϕ ,

$$N(\phi) = \frac{8\pi}{M_P^2} \int_{\phi}^{\phi_{fin}} d\phi \frac{V}{V'}. \quad (48)$$

The requirement for successful resolution of the problems with standard cosmology puts a lower bound on the total amount of e-folds which is typically in the range between 55 and 65. In the inflationary paradigm the temperature fluctuations and polarization of the CMB as well as the seeds necessary for large scale structure formation have their origin in the quantum fluctuations of the inflaton. During inflation the physical horizon size determined by H^{-1} (cf. Eq. (36)) is approximately a constant. This means that physical wavelengths which were within the horizon at some time are stretched beyond the horizon at some later time (recall that all physical scales are comoving scales multiplied by the scale factor). The horizon crossing takes place at $\lambda = 1/(aH)$ where λ denotes the comoving wavelength. Once a perturbation is outside the horizon its amplitude freezes and it becomes a classical perturbation. The spectrum of fluctuations can be calculated by quantizing the inflaton on the homogeneous background. Quantization of a scalar field on a de Sitter background is a well studied problem. Of course, slow roll inflation is not exactly a de Sitter background but the spectrum can still be calculated approximately. It is interesting to note that in the case of power law inflation for which the scale factor evolves as $a \sim t^p$ and the potential is an exponential potential the spectrum can be found exactly (for a detailed discussion of this case see, e.g. [18]).

The scalar field can be separated into a homogeneous part $\phi(t)$ and a perturbative part $\delta\phi(t, \vec{x})$ such that $\phi(t, \vec{x}) = \phi(t) + \delta\phi(t, \vec{x})$. Considering the simplest case of a massless scalar field then the perturbations satisfy in Fourier space for each comoving wave number \vec{k} the mode equation,

$$\ddot{\delta\phi}_{\vec{k}} + 3H\dot{\delta\phi}_{\vec{k}} + \left(\frac{k}{a}\right)^2 \delta\phi_{\vec{k}} = 0. \quad (49)$$

Quantizing $\delta\phi$ leads to a 2-point function $\langle \delta\phi_{\vec{k}} \delta\phi_{\vec{k}'} \rangle = \left(\frac{H}{2\pi}\right)^2 \delta_{\vec{k}\vec{k}'}$. Therefore on super horizon scales classical fluctuations in the scalar field with an amplitude $|\delta\phi| \simeq \frac{H}{2\pi}$ are induced. This in turn induces curvature perturbations which are imprinted in the CMB. This will be discussed in more detail in the next section. The phases of each wave are random. At each point in space the sum of all waves is described by a Brownian motion in all directions implying Gaussian perturbations.

3 The inhomogeneous universe

The cosmic microwave background provides us with a unique window to the physics of the early universe. To understand its formation it is important to understand the thermal and ionization history of the cosmic plasma. As the universe expands its temperature T corresponding to the photon temperature cools down as $T \propto a^{-1}$. Very early on, deep inside the radiation dominated era, temperatures are high enough so that all particles in the cosmic plasma are strongly coupled by interactions. In the best fit Λ CDM model the initial conditions for the numerical evolution of the perturbations which are imprinted in the CMB as temperature anisotropies and polarization are set after neutrino decoupling at around 1 MeV. At this time matter in the universe consists of a strongly coupled photon-baryon fluid and cold dark matter. To complete the model which best fits the data a cosmological constant Λ has to be added. However, dynamically Λ does not play a role until very close to the present epoch. Dark matter only interacts gravitationally and its presence is required to provide the gravitational potential field to explain the observed curvature fluctuations.

Once the temperature of the primordial plasma is low enough electrons and nuclei, mostly protons, recombine to form neutral atoms, mostly hydrogen. Defining the epoch of recombination by requiring that the ionization fraction is 0.1 determines the corresponding redshift to be $z = 1360$ and the temperature is of the order of 4000K. This is much lower than what would be expected if it just depended on the ionization energy of hydrogen 13.6 eV which corresponds to about 160000 K. This is due to subtleties in the recombination process involving two-photon decay processes (cf., e.g., [32]). The epoch of decoupling is described to be the moment after which (most) photons will not scatter again. Though this is not true for all photons since there is a residual ionization fraction of about 10^{-4} . To be more

precise photon decoupling is defined to be the epoch when the time between scatterings equals the age of the universe. This defines the surface of last scattering and is the origin of what is observed today as the cosmic microwave background (CMB). As already mentioned the CMB is remarkably homogeneous and isotropic. However, there are small temperature fluctuations $\Delta T/T \sim 10^{-5}$ and it is linearly polarized. It is precisely these temperature anisotropies and polarization which constrain the physics of the very early universe and its evolution. As will be discussed later on in more detail at around a redshift $z = 10$ reionization takes place which generates additional CMB anisotropies and polarization.

Temperature fluctuations $\Theta(\hat{\mathbf{n}}) = \delta T/T$ in the direction $\hat{\mathbf{n}}$ on the sky are expanded in spherical harmonics $Y_{\ell m}(\hat{\mathbf{n}})$ such that (cf. e.g. [33])

$$\Theta(\eta, \mathbf{x}, \hat{\mathbf{n}}) = \int \frac{d^3 k}{(2\pi)^3} \sum_{\ell} \sum_{m=-2}^2 \Theta_{\ell}^{(m)} G_{\ell}^m \quad (50)$$

which in general includes contributions from the scalar ($m = 0$), vector ($m = \pm 1$) and tensor ($m = \pm 2$) modes which are uncoupled at linear order. These modes describe the linear perturbations of the metric as well as the energy momentum tensor in Fourier space. They will be discussed below in more detail. Moreover,

$$G_{\ell}^m = (-i)^{\ell} \sqrt{\frac{4\pi}{2\ell+1}} Y_{\ell}^m(\hat{\mathbf{n}}) e^{i\mathbf{k}\cdot\mathbf{x}}. \quad (51)$$

The two-point function is determined by the corresponding angular power spectrum C_{ℓ}^{XY} such that [33]

$$(2\ell+1)C_{\ell}^{XY} = \frac{2}{\pi} \int \frac{dk}{k} \sum_{m=-2}^2 k^3 \langle X_{\ell}^{(m)*}(\eta_0, k) Y_{\ell}^{(m)}(\eta_0, k) \rangle \quad (52)$$

where X and Y denote Θ , E , and B . For completeness, also the polarization modes E and B are included here. The polarization of the CMB will be discussed in more detail below.

The cosmic microwave background has a nearly perfect Planck spectrum at a temperature $T_0 = 2.7255 \pm 0.0006$ K at $1\text{-}\sigma$ [34] which corresponds to the monopole, $\ell = 0$. The dipole corresponding to the multipole $\ell = 1$ has the largest amplitude of the temperature fluctuations at 3.372 ± 0.014 mK (95 CL) [3]. It is due to the motion of the solar system with respect to the CMB. An observer moving with a velocity $\beta = \frac{v}{c}$ relative to an isotropic Planck radiation field of temperature T_0 measures a Doppler shifted temperature pattern,

$$T(\theta) = \frac{T_0(1-\beta^2)^{\frac{1}{2}}}{1-\beta\cos\theta} \simeq T_0 \left(1 + \beta\cos\theta + \frac{\beta^2}{2}\cos(2\theta) + \frac{\beta^2}{2}\cos 2\theta + \mathcal{O}(\beta^3) \right) \quad (53)$$

observing at every point in the sky a black body spectrum with temperature $T(\theta)$. The observed Doppler shift implies that the barycenter of the solar system is moving with a velocity 371 ± 1 km s $^{-1}$ relative to the CMB rest frame in the direction determined by the galactic coordinates $(l, b) = (264.14 \pm 0.15, 48.26 \pm 0.15)$ which is almost orthogonal to the direction of the Galactic center [3]. Since the dipole is a frame-dependent quantity it is possible to define an "absolute rest frame" in which the CMB dipole is zero. Higher order order multipoles, $\ell \geq 2$ carry the information about fluctuations in the matter density and velocity fields as well as the gravitational field from before last scattering of the present day CMB photons upto today. There are quite a large number of observations of the CMB: ground based with radio telescopes, detectors mounted on balloons and satellites. To mention a few examples, the Atacama Cosmology Telescope (ACT) is a 6m radio telescope situated in the Atacama Plateau in Chile. It observes the CMB over a large area of the sky at three frequency channels, 148 GHz, 218 GHz and 277 GHz. Observations of the CMB anisotropies are for multipoles $540 < \ell < 9500$ [4]. Another example is the South Pole Telescope (SPT) located at the south pole and observing similarly the temperature

anisotropies for multipoles $2000 < \ell < 9400$ at three frequency bands, namely, 95 GHz, 150 GHz and 220 GHz [5]. Boomerang was the first balloon experiment to observe the CMB. There were two flights (1998, 2003) launched from McMurdo Station on a circular path around Antarctica lasting about 10 days [6]. The angular power spectrum of the CMB anisotropies was obtained for $75 < \ell < 1025$ in four 150 GHz channels. The first satellite to observe the CMB was COBE (Cosmic Background explorer) which was launched by NASA in 1989. Its measurement of the absolute spectrum of the CMB revealed a nearly perfect black body spectrum. The tiny deviations from the Planck spectrum observed by the COBE/FIRAS instrument constitute the first observational constraints on spectral distortion of the CMB. There are proposals for future space missions, such as PIXIE (Primordial Inflation Explorer) [35], to measure and tighten the constraints on spectral distortions of the CMB together with precise polarization measurements (B-mode). COBE observed the CMB temperature anisotropies at an effective angular resolution of 10 degrees in three frequency channels (31.5 GHz, 53 GHz and 90 GHz) [36]. After COBE the next satellite experiment to observe the CMB temperature anisotropies as well as polarization was WMAP (Wilkinson Microwave Anisotropy Probe) operated by NASA between 2001 and 2010 observing in 5 frequency channels (23,33,41,61 and 94 GHz). Angular power spectra cover a multipole range of $2 \leq \ell < 1000$ [7]. The Planck mission was on board a satellite operated by ESA. It took data between 2009 and 2013. It observed the CMB temperature anisotropies and polarization. The temperature anisotropy angular power spectrum covers the range $2 \leq \ell < 2500$ observed in separate frequency channels covering the range 30-857 GHz [8].

The observed CMB temperature perturbations are the result of perturbations of the perfectly isotropic and spatially homogeneous Friedman-Robertson-Walker background and its matter distribution. The scales on which the CMB is observed are sufficiently large so that perturbations are still in the linear regime. At linear order there are three types of perturbations of an FRW background, namely, scalar, vector and tensor perturbations depending on their behaviour under general coordinate transformations. Starting with a flat FRW background determined by the line element,

$$ds^2 = a^2(\eta)(d\eta^2 - \delta_{ij}dx^i dx^j) \quad (54)$$

the most general linear perturbation of the metric is given by

$$ds^2 = a^2(\eta) [(1 - 2A) d\eta^2 + 2B_i d\eta dx^i - [(1 + 2D) \delta_{ij} + 2E_{ij}] dx^i dx^j], \quad (55)$$

where the metric perturbation variables A , B_i , D and E_{ij} are all functions of space and time and the Einstein summation convention is used which corresponds to summing over repeated indices. Moreover, latin indices take values between 1 and 3. These functions are expanded in scalar, vector and tensor harmonics. In the case of scalar and vector perturbations there is an inherent gauge freedom associated with the metric perturbation variables. This is related to the fact that when perturbing the metric $g_{\mu\nu} \rightarrow g_{\mu\nu} + \delta g_{\mu\nu}$ this can be done using different ways of defining a coordinate system or in other words, slicings, corresponding to the choice of equal time hypersurfaces. Therefore, two perturbations of the same FRW background, with metric tensor, say $\mathcal{G}_{\mu\nu}$ and $\mathcal{H}_{\mu\nu}$, are related by a transformation of the space-time coordinates, $x^\mu \rightarrow \tilde{x}^\mu$ and the corresponding transformation of the metric tensors, $\mathcal{G}_{\mu\nu}(x^\alpha) \rightarrow \mathcal{H}_{\mu\nu}(\tilde{x}^\alpha)$. Since these are perturbations of the same background space-time the change in the perturbed metric tensor has to be calculated at the same coordinate value, that is $\Delta\mathcal{T}_{\mu\nu}(x^\alpha) \equiv \mathcal{H}_{\mu\nu}(x^\alpha) - \mathcal{G}_{\mu\nu}(x^\alpha)$. Calculating this to first order in the perturbations determines the gauge transformations of the metric perturbation variables (cf. Eq. (55)) as well as the corresponding linear perturbations of the energy-momentum tensor. Naturally, the physics must not depend on a gauge choice so all relevant quantities have to be gauge-invariant. Whereas for scalar modes there exists a number of gauge choices, for vector modes there exist only two gauge choices and the tensor modes are described by only gauge-invariant quantities. In the following we focus on the scalar mode and choose the so called conformal Newtonian gauge defined below, in Eq. (61), as well as the gauge-invariant formulation. The perturbation equations are then derived by calculating the first order perturbation of Einstein's equations. For details see, e.g. [37] or [13], [14], [18], [38].

Long before their decoupling photons are tightly coupled to the rest of the cosmic plasma by Thomson scattering. This implies that they are in thermal equilibrium and have a Planck distribution. However, as the universe cools down, Thomson scattering becomes less efficient and photons fall out of equilibrium perturbing the Planck distribution. In order to calculate the CMB temperature anisotropies and polarization of the CMB one has to follow the evolution of the photon phase-space distribution in the perturbed FRW background $f(\eta, \mathbf{x}, \mathbf{n}, q)$, given by

$$f(\eta, \mathbf{x}, \mathbf{n}, q) = f(q) + \delta f(\eta, \mathbf{x}, \mathbf{n}, q), \quad (56)$$

where $q = a(\eta)p(\eta, \mathbf{x})$ is the comoving photon energy and $f(q)$ the black body distribution,

$$f(q) = \frac{1}{e^{\frac{q}{T_0}} - 1}. \quad (57)$$

Moreover \mathbf{n} points along the direction of propagation of the photons. The distribution function is determined by the Boltzmann equation including a collision term $C[f]$ (cf., e.g., [13])

$$\frac{\partial f}{\partial \eta} + \frac{\partial f}{\partial x^i} \frac{dx^i}{d\eta} + \frac{\partial f}{\partial q} \frac{dq}{d\eta} + \frac{\partial f}{\partial n^i} \frac{dn^i}{d\eta} = \frac{df}{d\eta} \equiv C[f]. \quad (58)$$

Defining the brightness function $\Theta \equiv \frac{\delta T}{T}$ yielding

$$f(\eta, \mathbf{x}, \mathbf{n}, q) = \frac{1}{e^{\frac{q}{T_0(1+\Theta)}} - 1} \quad (59)$$

results at first order in

$$\delta f(\eta, \mathbf{x}, \mathbf{n}, q) = -q \frac{df(q)}{dq} \Theta(\eta, \mathbf{x}, \mathbf{n}). \quad (60)$$

The photon trajectory in the perturbed FRW background can be expressed at first order in terms of the functions characterizing the linear perturbations of the metric (cf., Eq. (55)). Since at linear order the perturbation equations for the scalar, vector and tensor modes decouple so do the corresponding Boltzmann equations. The scalar mode is the most important one to establish the (minimal) Λ CDM model. Thus, as way of example, we will consider the Boltzmann equation derived for the scalar mode perturbation of a flat FRW background. The scalar mode perturbation of the metric Eq. (55) in the conformal Newtonian gauge is given by

$$ds^2 = a^2(\eta) \left[-(1 + 2\Psi) d\eta^2 + (1 - 2\Phi) \delta_{ij} dx^i dx^j \right] \quad (61)$$

where Φ and Ψ are called gravitational potentials. This is related to the fact that during the matter dominated era $\Phi = \Psi$ and Newtonian gravity applies on scales well within the horizon with Φ being the Newtonian gravitational potential. For the scalar mode perturbation, the Boltzmann equation for the brightness perturbation at linear order in Fourier space, $\Theta(\eta, \mathbf{k}, \mathbf{n})$, is given by

$$\frac{\partial \Theta}{\partial \eta} + ik\mu\Theta - \frac{\partial \Phi}{\partial \eta} + ik\mu\Psi = C[\Theta] \quad (62)$$

where $\mu = \hat{\mathbf{k}} \cdot \mathbf{n}$. Similarly, a Boltzmann equation for the vector as well as the tensor mode can be derived (for details, see e.g., [37] or [13], [14], [18]). The next step is to expand the brightness perturbation $\Theta(\eta, \mathbf{x}, \mathbf{n})$ in terms of spherical harmonics (cf. equation (50)). This leads to the Boltzmann hierarchy which determines the evolution of the multipole components $\Theta_\ell^{(m)}(\eta, \mathbf{k})$. For the scalar mode the first three multipoles determine the energy density perturbation, $\Theta_0^{(0)} = \frac{\delta\gamma}{4}$, the velocity, $\Theta_1^{(0)} = \frac{V_\gamma}{3}$, and the anisotropic stress of the photon fluid, $\Theta_2^{(0)} = \frac{\pi_\gamma}{12}$. The latter one encodes one of the contributions to

the deviation from a perfect fluid. There exist a variety of numerical programs to solve the Boltzmann hierarchy and calculate the CMB anisotropies. These are all open source. The first one was the COSMOS program [39] followed by CMBFAST [40], CAMB [41], CMBEASY [42] and CLASS [43–46] to mention just a few. A good reference to find links to these and other programs as well as to published data of many CMB observations is the NASA Legacy Archive for Microwave Background Data Analysis LAMBDA at <http://lambda.gsfc.nasa.gov/>.

The minimal model used to fit the observations of the CMB is the 6-parameter Λ CDM model in which case initial conditions for the numerical solutions are set after standard model (SM) neutrino decoupling, thus at $T < 1$ MeV. At this moment, well within the radiation dominated era, in the universe the matter is constituted by the already decoupled SM neutrinos, cold dark matter (CDM) and the tightly coupled baryon-electron-photon fluid. In the latter matter component the Coulomb interaction between electrons and baryons, that is nuclei, mostly protons (75%) and helium-4 nuclei (<25%) cause locally equal number densities of electrons and baryons. Photons are tightly coupled to the electron-baryon fluid via the (non relativistic) Thomson scattering off the (already non-relativistic) electrons and nuclei. The minimal model is based on the scalar mode with adiabatic initial conditions which will be discussed below. Extensions of the minimal Λ CDM model include a tensor mode which would be generated if the relevant cosmological perturbations are generated during inflation. The tensor mode also has a particular signature in the polarization of the CMB as it generates the B-mode of polarization (see below) which is why the observation of a primordial B-mode is considered to be a strong indication that inflation took place in the early universe. Other extensions include massive neutrinos (cf., e.g., [47]) and primordial magnetic fields, e.g., [48–50]. Primordial magnetic fields source all three types of linear perturbation modes. In particular, whereas vector modes decay in the standard Λ CDM model they do get sourced by a primordial magnetic field present before decoupling.

Initial conditions for the cosmological perturbation equations are set well outside the horizon deep inside the radiation dominated era. For scalar modes adiabatic initial conditions and isocurvature initial conditions constitute two different classes of initial conditions. Taking as an example the radiation-baryon fluid then its specific entropy of radiation normalized to the baryon number n_b is given by $S_{\gamma b} = \frac{S_\gamma}{n_b}$. Hence the first order fractional perturbation is given by (cf. e.g. [51])

$$\frac{\delta S_{\gamma b}}{S_{\gamma b}} = \frac{\delta S_\gamma}{S_\gamma} - \frac{\delta n_b}{n_b} = 3 \frac{\delta T_\gamma}{T_\gamma} - \frac{\delta n_b}{n_b} = \frac{\delta \rho_\gamma}{\rho_\gamma + p_\gamma} - \frac{\delta \rho_b}{\rho_b} = \frac{3}{4} \delta_\gamma - \delta_b. \quad (63)$$

In this case imposing adiabaticity implies

$$\frac{\delta S_{\gamma b}}{S_{\gamma b}} = 0 \Rightarrow \frac{3}{4} \delta_\gamma = \delta_b. \quad (64)$$

Eq. (63) can be generalized to any other fluid component with equation of state, $p_j = w_j \rho_j$, where w_j is a constant, so that in general,

$$\frac{\delta S_{\gamma j}}{S_{\gamma j}} = \frac{\delta_\gamma}{w_\gamma + 1} - \frac{\delta_j}{1 + w_j}. \quad (65)$$

Thus for the Λ CDM model adiabatic initial conditions are determined by

$$\frac{1}{4} \delta_\gamma = \frac{1}{4} \delta_\nu = \frac{1}{3} \delta_b = \frac{1}{3} \delta_{cdm}. \quad (66)$$

The amplitude of adiabatic perturbations are characterized by a gauge-invariant quantity, namely, the curvature perturbation on a uniform density hypersurface ζ . In the gauge invariant formalism it is given by (cf., e.g., [14])

$$-\zeta = -\Phi + \mathcal{H}k^{-1}V \quad (67)$$

which for the standard Λ CDM yields to

$$\zeta = \frac{\Delta_\gamma}{4}, \quad (68)$$

where Δ_γ is the gauge-invariant photon energy density contrast. On large scales, this is related to the curvature perturbation on a comoving slicing by $\mathcal{R} \simeq -\zeta$ (cf., e.g., [18]).

The minimal, 6-parameter Λ CDM model includes only the scalar, adiabatic mode to source the density fluctuations necessary to explain the observed CMB temperature anisotropies and polarization. One of the quests of cosmology is to determine the origin of the corresponding primordial curvature fluctuations. A promising candidate is single field inflation as it naturally generates adiabatic initial conditions. At the end of inflation the inflaton decays and the SM particles are created. Since their overall ratios are fixed, this leads to $\delta(n_A/n_B) = 0$ implying $\delta S_{A,B}/S_{A,B} = 0$ [52, 53]. The comoving curvature perturbation from single field inflation is given by

$$\mathcal{R} = -H \frac{\delta\rho}{\dot{\rho}} \Rightarrow \mathcal{R} = -H \frac{\delta\phi}{\dot{\phi}}. \quad (69)$$

Using the two point function of the inflaton fluctuations (cf. after Eq. (49)) the two-point function $\langle \mathcal{R}_{\mathbf{k}} \mathcal{R}_{\mathbf{k}'} \rangle = \frac{2\pi^2}{k^3} \mathcal{P}_{\mathcal{R}} \delta_{\mathbf{k},\mathbf{k}'}$ is then determined by the power spectrum

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{1}{4\pi} \left(\frac{H^2}{\dot{\phi}} \right)^2 \Big|_k \quad (70)$$

which in slow roll inflation can be expressed as [27]

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{1}{24\pi^2 M_P^4} \frac{V}{\epsilon} \Big|_{k=aH} \quad (71)$$

where ϵ is one of the slow roll parameters defined in Eq. (46) calculated at horizon crossing. An effective spectral index can be defined as

$$n - 1 \equiv \frac{d \ln \mathcal{P}_{\mathcal{R}}(k)}{d \ln k} \quad (72)$$

which can be expressed in terms of the slow roll parameters as

$$n(k) - 1 = -6\epsilon + 2\eta. \quad (73)$$

In addition often a running of the spectral index is considered,

$$\frac{dn}{d \ln k} = -16\epsilon\eta + 24\epsilon^2 + 2\xi \quad (74)$$

where a third slow roll parameter ξ is defined by $\xi \equiv \frac{M_P^4}{64\pi^2} \frac{V'V'''}{V^2}$. For example, Planck 15 constrains this scale dependence of the spectral index as [23]

$$\frac{dn}{d \ln k} = -0.0084 \pm 0.0082, \quad \text{Planck } TT + \text{low}P \quad (75)$$

$$\frac{dn}{d \ln k} = -0.0057 \pm 0.0071, \quad \text{Planck } TT, TE, EE + \text{low}P \quad (76)$$

where TT denotes the Planck data determining the temperature autocorrelation function, TE the Planck temperature and E-mode polarization cross correlation data, EE the Planck E-mode auto correlation data and $\text{low}P$ the low ℓ E-mode polarization data.

Isocurvature initial conditions have the defining property of a vanishing total curvature perturbation, $\mathcal{R} = 0$. They do, however, imply entropy perturbations between the photons and other types of matter in the universe as discussed before. In general, these can be generated during multi-field inflation since single-field inflation only excites adiabatic modes. Hence isocurvature initial conditions are determined by

$$S_j \equiv \frac{\delta S_{\gamma j}}{S_{\gamma j}} \neq 0 \quad (77)$$

where the index j indicates the type of matter. Each of these defines an independent isocurvature mode. In the minimal extension of the Λ CDM model isocurvature modes are induced by fluctuations of the baryon number w.r.t to the photon number density, namely the baryon isocurvature mode ($\mathcal{R} = 0$, $S_{cdm} = 0 = S_\nu$, $S_b \neq 0$), as well as the CDM isocurvature mode ($\mathcal{R} = 0$, $S_b = 0 = S_{cdm}$, $S_{cdm} \neq 0$). In principle there are also two additional isocurvature modes for the neutrinos. One is the neutrino density isocurvature mode in which case the initial conditions are determined by $\mathcal{R} = 0$, $S_b = 0 = S_{cdm}$, $S_\nu \neq 0$. It can be excited similarly to the baryon or CDM isocurvature modes. The other is the neutrino velocity mode taking into account the evolution of the neutrino anisotropic stress. However, so far there is no physical mechanism to excite this mode (cf., e.g., [54], [15]).

The resulting angular power spectrum of only isocurvature modes is not compatible with observations. Though, it is possible to have a relatively small contribution from isocurvature modes with the total power spectrum clearly dominated by the adiabatic mode. In general, the different modes can have non-vanishing correlations. Thus there are two parameters to characterize the isocurvature mode contribution. Firstly, there is the primordial isocurvature fraction which in general depends on scale in the parametrization used in [54],

$$\beta_{iso}(k) = \frac{\mathcal{P}_{\mathcal{I}\mathcal{I}}(k)}{\mathcal{P}_{\mathcal{R}}(k) + \mathcal{P}_{\mathcal{I}\mathcal{I}}(k)} \quad (78)$$

where \mathcal{I} denotes the isocurvature mode. Secondly, there is the correlation between the adiabatic mode (\mathcal{R}) and the isocurvature mode (\mathcal{I}) which is encoded in the scale independent correlation fraction,

$$\cos \Delta_{ab} = \frac{\mathcal{P}_{ab}}{(\mathcal{P}_{aa}\mathcal{P}_{bb})^{\frac{1}{2}}} \quad (79)$$

where $a, b = \mathcal{R}, \mathcal{I}$.

For example, the Planck collaboration finds from the analysis of the 2015 temperature data plus polarization at low ℓ for the adiabatic mode plus a CDM isocurvature mode that the 95% CL upper bounds on the scale dependent fractional primordial contribution of isocurvature modes are [54]

$$100\beta_{iso,k=0.002\text{Mpc}^{-1}} = 4.1, \quad 100\beta_{iso,k=0.050\text{Mpc}^{-1}} = 35.4, \quad 100\beta_{iso,k=0.100\text{Mpc}^{-1}} = 56.9 \quad (80)$$

and the scale independent primordial correlation fraction, $\cos \Delta_{\mathcal{R}\mathcal{I}}$ is in the intervall [-30:20].

The Boltzmann hierarchy is solved by a line-of-sight integral [40, 55] which for the brightness perturbation of the scalar mode is given by

$$\frac{\Theta_\ell}{2\ell + 1} = \int_{\eta_{in}}^{\eta_0} d\eta S(k, \eta) j_\ell [k(\eta_0 - \eta)], \quad (81)$$

where $j_\ell(x)$ denote the spherical Bessel functions and $S(k, \eta)$ is the source function which can be written in terms of different physical contributions S_i as (e.g., [14, 56])

$$S = \sum_i S_i. \quad (82)$$

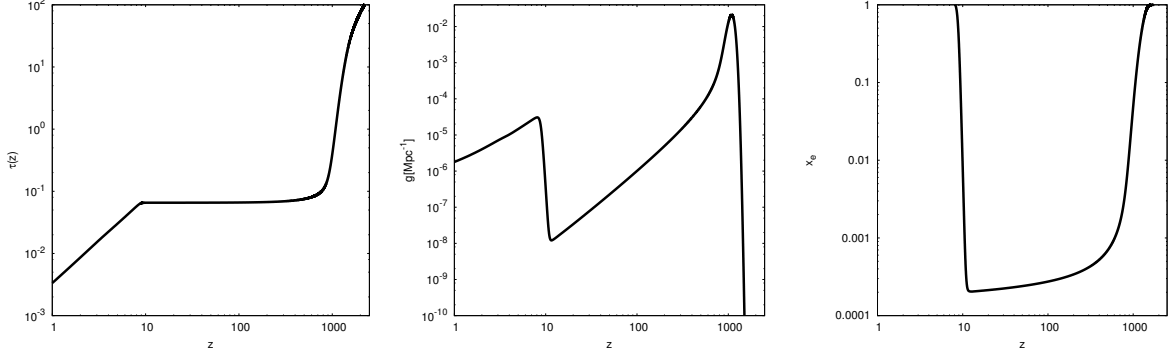


Fig. 5: The optical depth (*left*), the visibility function (*middle*) and the ionization fraction (*right*) for the bestfit parameters from Planck 15 (TT,TE,EE+lowP+lensing+ext(BAO+JLA+ H_0)).

The sum includes the Sachs-Wolfe term which is important at low ℓ , corresponding to large angular scales,

$$S_{SW} = \frac{g}{4} \Delta_\gamma, \quad (83)$$

the integrated Sachs-Wolfe term,

$$S_{ISW} = g(\Phi + \Psi) - e^{-\tau}(\dot{\Phi} + \dot{\Psi}), \quad (84)$$

the Doppler term involving the gauge-invariant baryon velocity V_b

$$S_{dop} = k^{-1} (g\dot{V}_b + \dot{g}V_b) \quad (85)$$

and a polarization term

$$S_{pol} = gP^{(0)} \quad (86)$$

involving the quadrupole of the brightness perturbation as well as of the E-mode polarization. Moreover, $g \equiv \dot{\tau}e^{-\tau}$ is the visibility function and $\tau(t) = \sigma_T \int_t^{t_0} n_e(t) dt$ is the optical depth. The term $e^{-\tau(t)}$ gives the probability that a CMB photon observed today, that is, at t_0 , has not scattered since time t . The optical depth and the visibility function are shown in figure 5. All figures in this section have been done using the bestfit parameters of the Λ CDM base model derived from Planck 15 temperature, polarization data, lensing likelihood and additional data from large scale structure and supernovae (BAO, JLA and H_0) at 68 %CL given by $\Omega_b h^2 = 0.02230 \pm 0.00014$, $\Omega_c h^2 = 0.1188 \pm 0.0010$, $z_{re} = 8.8_{-1.1}^{+1.2}$, $A_s = (2.142 \pm 0.049) \times 10^{-9}$, $n_s = 0.9667 \pm 0.0040$ [23]. As can be seen it is strongly peaked around decoupling at around $z = 1089.90$. Upto a redshift around $z = 8.8$ the visibility function decays monotonously and then rises again due to the complete reionization of the universe at late times. This will be discussed in more detail below. The different contributions to the total angular power spectrum of the brightness perturbation induced by the adiabatic, scalar mode are shown in figure 6. Also shown are the binned Planck 2015 data. As can be appreciated from figure 6 different contributions are important on different scales. On large scales the Sachs-Wolfe plateau can be identified. The subsequent overall peak structure are an imprint of the acoustic oscillations of the baryon-photon fluid in the tight coupling regime, long before last scattering of the photons. This can be seen from the evolution equations of the density contrast and velocities of baryons and photons in the tight-coupling limit for the scalar mode, which are given in the gauge-invariant formalism (cf. [37]),

$$\dot{\Delta}_\gamma = -\frac{4}{3} k V_\gamma, \quad (87)$$

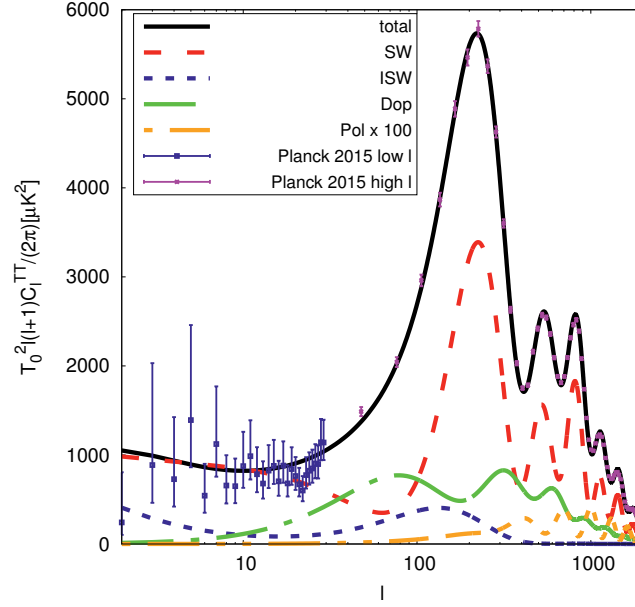


Fig. 6: The angular power spectrum of the temperature autocorrelation function calculated for the bestfit parameters from Planck 15 (TT,TE,EE+lowP+lensing+ext(BAO+JLA+ H_0)). Also shown are the binned Planck 2015 data [57]. The different contributions due to the Sachs-Wolfe effect ("SW"), the integrated Sachs-Wolfe effect ("ISW"), the Doppler term ("Dop") as well as the (amplified) polarization term ("Pol x 100") are also included.

$$\dot{V}_\gamma = k(\Psi - \Phi) + \frac{k}{4}\Delta_\gamma - \frac{k}{6}\pi_\gamma + \tau_c^{-1}(V_b - V_\gamma) \quad (88)$$

and the baryons satisfy,

$$\dot{\Delta}_b = -kV_b - 3c_s^2\mathcal{H}\Delta_b, \quad (89)$$

$$\dot{V}_b = (3c_s^2 - 1)\mathcal{H}V_b + k(\Psi - 3c_s^2\Phi) + kc_s^2\Delta_b + R\tau_c^{-1}(V_\gamma - V_b), \quad (90)$$

where $\tau_c^{-1} = an_e\sigma_T$ is the mean free path of photons between scatterings with σ_T the Thomson cross section. Moreover, $R \equiv \frac{4}{3}\frac{\rho_\gamma}{\rho_b}$ and the $c_s^2 = \frac{\partial \bar{p}}{\partial \bar{\rho}}$ is the adiabatic sound speed calculated in terms of the background quantities. Here the dot denotes the time derivative w.r.t. to conformal time η and $\mathcal{H} = \frac{\dot{a}}{a}$. These equations can be combined to yield a second order differential equation for Δ_γ which describes a forced harmonic oscillator [58],

$$\ddot{\Delta}_\gamma + \frac{\dot{R}_b}{1 + R_b}\dot{\Delta}_\gamma + c_{sb\gamma}^2 k^2 \Delta_\gamma \simeq \frac{4k^2}{3} \frac{2 + R_b}{1 + R_b} \Phi, \quad (91)$$

where $R_b \equiv \frac{1}{R}$ and $c_{sb\gamma}^2 \equiv \frac{1}{3} \frac{1}{R_b + 1}$ is the sound speed of the baryon-photon fluid. For adiabatic initial conditions this is solved by

$$\begin{aligned} \Delta_\gamma(\eta) = & \frac{1}{(1 + R_b)^{\frac{1}{4}}} \left[\Delta_\gamma(0) \cos(kr_s(\eta)) + \frac{\sqrt{3}}{k} \left[\dot{\Delta}_\gamma(0) + \frac{1}{4}\dot{R}_b(0)\Delta_\gamma(0) \right] \sin(kr_s(\eta)) \right. \\ & \left. + \frac{\sqrt{3}}{k} \int_0^\eta d\tau' (1 + R_b(\eta'))^{\frac{3}{4}} \sin[kr_s(\eta) - kr_s(\eta')] F(\eta') \right] \end{aligned} \quad (92)$$

where $F(\eta) \equiv \frac{4k^2}{3} \frac{2 + R_b}{1 + R_b} \Phi$ and the sound horizon is defined by

$$r_s(\eta) \equiv \int_0^\eta c_{sb\gamma} d\eta'. \quad (93)$$

After the first acoustic peak around $\ell = 220$ a clear damping of the amplitude of the subsequent peaks is observed. This is due to the photon diffusion damping. As the universe cools down the number of free electrons drastically diminishes as they combine with the nuclei. This implies that the mean free path of photons increases strongly and tight coupling between the baryon and photon fluids breaks down. However, as this is not an instantaneous process photons diffuse and later on start free streaming within the baryon fluid thereby moving baryons out of the potential wells of the CDM gravitational field. This results in an attenuation of the baryon density perturbation which in turn results in a damping of the corresponding brightness perturbations which manifests itself in the CMB temperature as well as polarization angular power spectra. The resulting damping is clearly seen in the numerical solutions of the Boltzmann hierarchy. However, by using a simple model of a random walk it can also be estimated and understood physically. It is rather interesting that the effect of the damping is well approximated by simply multiplying the amplitudes by a factor $\exp(-k^2/k_D^2(\eta))$ as the comparison between analytical approximations and numerical solutions shows [58]. Here, k_D^{-1} , is the Silk scale or photon diffusion scale. It is roughly the comoving distance a photon can travel since some initial time. It can be estimated by modelling the movement of the photon in the local baryon rest frame as a random walk (e.g., [13]). In this case the mean time between collisions (i.e. scattering off) with electrons is $t_c \sim (\sigma_T n_e)^{-1}$. The average number of steps during a time interval t is thus $N = t/t_c$. Therefore, during a time interval t the photon diffuses a distance of the order $d \sim \sqrt{N}t_c \sim (tt_c)^{1/2}$. This yields the physical photon diffusion scale

$$ak_D^{-1} \simeq \left(\frac{t}{\sigma_T n_e} \right)^{1/2}. \quad (94)$$

Therefore, the comoving photon diffusion scale evolves with the scale factor during radiation domination as $k_D^{-1} \sim a^{3/2}$ and during matter domination as $k_D^{-1} \sim a^{5/4}$. Each scale k^{-1} starts out bigger than the Silk scale at horizon entry. Once $k_D(\eta) = k$ is reached photon diffusion damping sets in.

The amplitude of the angular power spectra depends rather strongly on the cosmological parameters. For example, whereas increasing the baryon energy density parameter strongly increases the first acoustic peak, increasing the total matter energy density lowers the first peak.

Another important observation of the CMB is its linear polarization. This polarization is due to Thomson scattering of photons off free electrons at around last scattering. The polarization of a plane electromagnetic wave is commonly determined by the behaviour of its electric field. The magnetic field is obtained from Maxwell's equations. The most general form of a homogeneous monochromatic plane wave travelling in the direction $\mathbf{k} = k\hat{\mathbf{n}}$ is given by (cf. e.g. [59]),

$$\mathbf{E} = (\epsilon_1 E_1 + \epsilon_2 E_2) e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)} \quad (95)$$

where ϵ_i are the polarization vectors which together with $\hat{\mathbf{n}}$ can be chosen to form a right handed orthonormal basis. The amplitudes E_i can be conveniently written separating out the phase δ_i so that $E_i = s_i e^{i\delta_i}$, $i = 1, 2$. The wave is linearly polarized if there is no phase difference, i.e. $\delta_1 = \delta_2$. In this case the electric field has a constant direction which makes an angle $\theta = \tan^{-1} \left(\frac{E_2}{E_1} \right)$ with the direction ϵ_1 . Otherwise, in general, the wave is elliptically polarized with the particular case of circular polarization for which the amplitudes satisfy $|E_1| = |E_2|$ and the phase difference is $|\delta_1 - \delta_2| = \frac{\pi}{2}$. There is another choice of basis vectors instead of those spanning the plane transverse to the direction of propagation, which is the helicity basis with the basis vectors $\epsilon_{\pm} = \frac{1}{\sqrt{2}} (\epsilon_1 \pm i\epsilon_2)$. The description of radiation requires in general four parameters, namely, its intensity, degree of polarization, plane of polarization and the ellipticity of radiation. These distinct parameters are encoded in a compact way in the four Stokes parameters (cf. e.g., [60]). The Stokes parameters in terms of the linear polarization basis vectors $\epsilon_{1,2}$ are given by (e.g., [59])

$$I = |\epsilon_1 \cdot \mathbf{E}|^2 + |\epsilon_2 \cdot \mathbf{E}|^2 \Rightarrow I = s_1^2 + s_2^2 \quad (96)$$

$$Q = |\boldsymbol{\epsilon}_1 \cdot \mathbf{E}|^2 - |\boldsymbol{\epsilon}_2 \cdot \mathbf{E}|^2 \Rightarrow Q = s_1^2 - s_2^2 \quad (97)$$

$$U = 2\text{Re}[(\boldsymbol{\epsilon}_1 \cdot \mathbf{E})^*(\boldsymbol{\epsilon}_2 \cdot \mathbf{E})] \Rightarrow U = 2s_1s_2 \cos(\delta_2 - \delta_1) \quad (98)$$

$$V = 2\text{Im}[(\boldsymbol{\epsilon}_1 \cdot \mathbf{E})^*(\boldsymbol{\epsilon}_2 \cdot \mathbf{E})] \Rightarrow V = 2s_1s_2 \sin(\delta_2 - \delta_1) \quad (99)$$

These satisfy,

$$I^2 = Q^2 + U^2 + V^2. \quad (100)$$

Moreover, going to the helicity basis it can be shown that $V = s_+^2 - s_-^2$ giving the difference between the intensities of positive and negative helicity states. From equations (96) to (99) it follows that for linearly polarized radiation, $Q \neq 0$, $U \neq 0$ but $V = 0$. For circularly polarized light, $Q=0=U$ and $V \neq 0$. So far we have considered the case of one monochromatic plane wave. However, in general in astrophysics and cosmology one would rather expect a diffuse, not monochromatic radiation which is not 100% polarized but rather partially polarized. This type of radiation can be described by the incoherent sum of unpolarized and fully polarized contributions leading to $Q^2 + U^2 + V^2 < I^2$. This leads to the definition of the degree of polarization p ,

$$p = \frac{\sqrt{Q^2 + U^2 + V^2}}{I} \quad (101)$$

This reduces to the definition of the degree of linear polarization for $V = 0$ [61].

Rotating the linear polarization vectors $\boldsymbol{\epsilon}_{1,2}$ by an angle ψ around the normal $\hat{\mathbf{n}}$, namely, (cf. e.g. [18])

$$\begin{aligned} \boldsymbol{\epsilon}'_1 &= \cos \psi \boldsymbol{\epsilon}_1 + \sin \psi \boldsymbol{\epsilon}_2, \\ \boldsymbol{\epsilon}'_2 &= -\sin \psi \boldsymbol{\epsilon}_1 + \cos \psi \boldsymbol{\epsilon}_2 \end{aligned} \quad (102)$$

leaves I and V invariant while the combination $Q \pm iU$ behaves as a spin-2 variable, i.e.

$$(Q \pm iU)' = e^{\mp 2i\psi} (Q \pm iU). \quad (103)$$

This makes this combination to be the most useful one to describe the polarization of the CMB. Whereas the temperature fluctuations or brightness perturbations are expanded in spin-0 spherical harmonics $Y_{\ell m}$ (cf. equation (50)) the CMB polarization variables are expanded in terms of spin-2 spherical harmonics ${}_{\pm 2}Y_{\ell m}$ such that [33]

$$(Q \pm iU)(\eta, \mathbf{x}, \hat{\mathbf{n}}) = \int \frac{d^3k}{(2\pi)^3} \sum_{\ell} \sum_{m=-2}^2 (E_{\ell}^{(m)} \pm iB_{\ell}^{(m)}) {}_{\pm 2}G_{\ell}^m, \quad (104)$$

where ${}_{\pm 2}G_{\ell}^m = (-1)^{\ell} \sqrt{\frac{4\pi}{2\ell+1}} {}_{\pm 2}Y_{\ell}^m(\hat{\mathbf{n}}) e^{i\mathbf{k}\cdot\mathbf{x}}$. As before, scalar modes correspond to $m = 0$, vector modes to $m = \pm 1$ and tensor modes to $m = \pm 2$. $E_{\ell}^{(m)}$ and $B_{\ell}^{(m)}$ are the E-mode and B-mode of polarization. By using a representation of the spin-2 spherical harmonics in terms of derivative operators it can be shown that these modes can be written in terms of a gradient and a curl of a field, respectively. This gave rise to the notation in terms of E-mode (“*electric*”) and B-mode (“*magnetic*”).

Before photon decoupling the tight coupling of the baryon-photon fluid is achieved by Thomson scattering. Its cross section is given by

$$\frac{d\sigma}{d\Omega}(\mathbf{n}, \boldsymbol{\epsilon}; \mathbf{n}^{(0)}, \boldsymbol{\epsilon}^{(0)}) = \frac{3\sigma_T}{8\pi} |\boldsymbol{\epsilon}^* \cdot \boldsymbol{\epsilon}^{(0)}|^2 \quad (105)$$

where $\boldsymbol{\epsilon}^{(0)}$ and $\boldsymbol{\epsilon}$ are the polarization vectors of the incoming and outgoing photons, respectively. The Thomson cross section is given by $\sigma_T \equiv \frac{8\pi}{3} \left(\frac{e^2}{m_e c^2}\right)^2$. The incoming and outgoing propagation directions

of the photons are $\mathbf{n}^{(0)}$ and \mathbf{n} , respectively. Thomson scattering does not source circular polarization. Therefore in the Λ CDM model $V = 0$ and generally it is not considered in CMB physics though there are some exceptions. The two polarization vectors of the incoming photons can be chosen in spherical coordinates as (cf. e.g. [59])

$$\boldsymbol{\epsilon}_1^{(0)} = \cos\theta (\mathbf{e}_x \cos\phi + \mathbf{e}_y \sin\phi) - \mathbf{e}_z \sin\theta \quad (106)$$

$$\boldsymbol{\epsilon}_2^{(0)} = -\mathbf{e}_x \sin\phi + \mathbf{e}_y \cos\phi. \quad (107)$$

The polarization vectors of the outgoing (scattered) radiation will be chosen along the cartesian coordinate \mathbf{x} - and \mathbf{y} -axes, respectively, so that

$$\boldsymbol{\epsilon}_1 = \mathbf{e}_x \quad (108)$$

$$\boldsymbol{\epsilon}_2 = \mathbf{e}_y \quad (109)$$

Moreover $\mathbf{n}^{(0)}$ and \mathbf{n} make an angle θ . The differential cross section by definition is the ratio of the power radiated in the direction \mathbf{n} with polarization $\boldsymbol{\epsilon}$ per unit solid angle and the unit incident flux which is the power per unit area in the direction \mathbf{n}_0 and polarization $\boldsymbol{\epsilon}_0$. Thus for scattering of electromagnetic waves off one electron at a distance r from the electron [59],

$$\frac{d\sigma}{d\Omega} (\mathbf{n}, \boldsymbol{\epsilon}; \mathbf{n}^{(0)}, \boldsymbol{\epsilon}^{(0)}) = \frac{r^2 \frac{1}{2} |\boldsymbol{\epsilon}^* \cdot \mathbf{E}|^2}{\frac{1}{2} |\boldsymbol{\epsilon}^{(0)*} \cdot \mathbf{E}^{(0)}|^2} \quad (110)$$

where $\mathbf{E}^{(0)}$ and \mathbf{E} are the electric fields of the incoming and scattered radiation far away from the scatterer, respectively. Therefore using equations (105) and (110) and defining partial intensities $I_m \equiv |\boldsymbol{\epsilon}_m^* \cdot \mathbf{E}|^2$ and equivalently for the incident radiation, then

$$I_m = r^{-2} \sum_{n=1,2} \left(\frac{d\sigma}{d\Omega} \right) (\hat{\mathbf{n}}, \boldsymbol{\epsilon}_m; \hat{\mathbf{n}}^{(0)}, \boldsymbol{\epsilon}_n^{(0)}) I_n^{(0)}. \quad (111)$$

Assuming that the incident radiation is unpolarized implies that the intensity only depends on the propagation direction $\mathbf{n}^{(0)}$ and not on the polarization vectors thus $I_n^{(0)} = I_n^{(0)}(\hat{\mathbf{n}}^{(0)})$ (e.g. [62]). Moreover $I_1^{(0)} = \frac{I^{(0)}}{2} = I_2^{(0)}$. Using the expressions (106)-(109) for the polarization vectors and integrating over all incoming directions $\mathbf{n}^{(0)}$ the Stokes parameter Q (cf. equation (97)) reads

$$Q = -\frac{3\sigma_T}{16\pi} \int d\Omega_{\mathbf{n}^{(0)}} r^{-2} I^{(0)}(\mathbf{n}^{(0)}) \cos(2\phi) \sin^2\theta. \quad (112)$$

The brightness perturbation Θ is determined by the fractional perturbation of the intensity I . Using the expansion in terms of spin-0 spherical harmonics $Y_\ell^m(\mathbf{n})$ (cf. equation (50)) it follows that the polarization parameter Q is determined by the quadrupole moment of the brightness perturbation since $\sin^2\theta \cos 2\phi \propto [Y_2^2(\hat{\mathbf{n}}^{(0)}) + Y_2^{-2}(\hat{\mathbf{n}}^{(0)})]$. Moreover, since $U(\hat{\mathbf{n}}^{(0)})$ can be obtained from $Q(\hat{\mathbf{n}}^{(0)})$ by a coordinate transformation $\phi \rightarrow \phi - \frac{\pi}{4}$ (cf. equation (103)) the corresponding expression for U is simply given by equation (112) with $\cos(2\phi)$ replaced by $\sin(2\phi)$. Therefore, U is also sourced by the quadrupole of the brightness perturbation since $\sin^2\theta \propto [Y_2^2(\hat{\mathbf{n}}^{(0)}) - Y_2^{-2}(\hat{\mathbf{n}}^{(0)})]$.

The evolution of the brightness perturbations indicate that only close to photon decoupling the amplitude of the multipoles can grow significantly. Thus the primordial polarization is generated at last scattering. Since it is sourced by the brightness perturbation it also shows the characteristic oscillating structure imprinted by the oscillations of the baryon photon fluid (cf figure 7). Six acoustic peaks in the E-mode autocorrelation functions have been observed by ACTPol [63] and SPTPol [64]. The feature at low ℓ in the angular power spectrum of the auto- and temperature cross correlation function of the E- mode (cf. figure 7) is due to reionization at low redshifts. In the reionized universe the number of

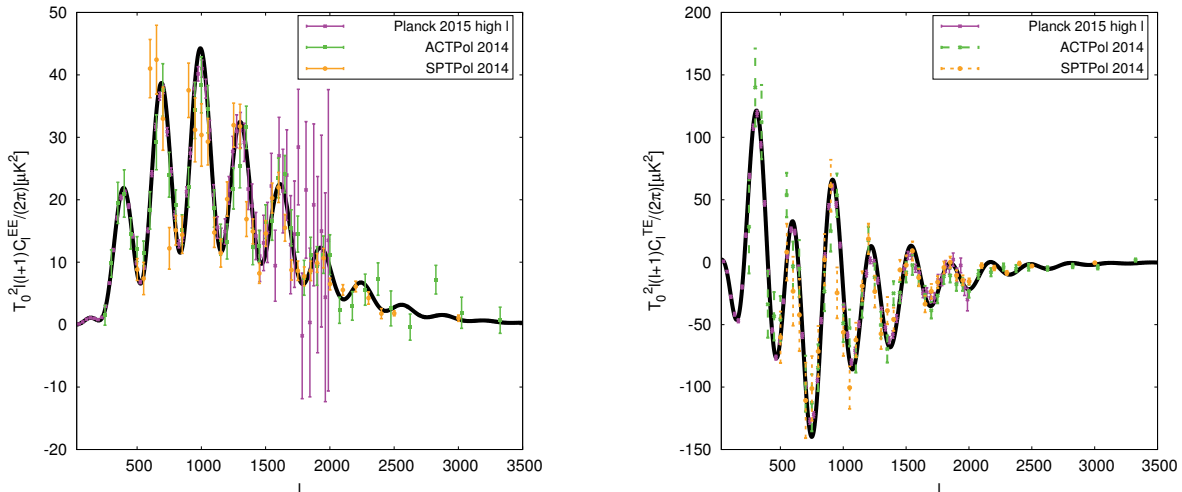


Fig. 7: The angular power spectrum of the E-mode autocorrelation function (*left*) and the cross correlation function of the temperature and E-mode (*right*) for the bestfit parameters from Planck 15 (TT,TE,EE+lowP+lensing+ext(BAO+JLA+ H_0)) together with the Planck 2015 [57], ACTPol [63] and SPTPol [64] data.

free electrons is very large again. Some of the CMB radiation gets re-scattered thereby increasing the polarization signal.

The first observations that the universe today is completely ionized came from observations of quasar spectra. The intergalactic medium (IGM) which permeates the space between galaxies can be probed by its emission and absorption spectra of light from distant background sources. Quasars are ideal background sources since due to their immense power they can be observed upto very high redshifts. Hydrogen is the most abundant element in the universe. If hydrogen is neutral in regions along the line of sight then the Lyman α absorption at the corresponding rest frame wavelengths should lead to the "Gunn-Peterson trough" in the quasar spectrum at wavelengths shorter than the terrestrial reference wavelength at $\lambda_{Ly\alpha} = 121.6$ nm. For objects at redshifts beyond the reionization redshift a complete Gunn-Peterson trough should be observed. Discrete absorption features are observed due to residual regions of neutral hydrogen which give rise to the so-called Lyman α forest seen in quasar spectra. With surveys such as the Sloan Digital Sky Survey (SDSS) a number of quasars at redshifts $z \gtrsim 6$ have been observed. Their spectra show indications for a complete Gunn-Peterson trough and thus imply a lower limit on the reionization redshift at around $z \simeq 6$. Observations of the CMB are well fitted by assuming instantaneous reionization at a redshift z_{reio} . A typical evolution of the ionization fraction is shown in figure 5. Planck 15 data are bestfit for a reionization redshift $z_{reio} = 8.8^{+1.2}_{-1.1}$ at 68%CL (using the temperature and low ℓ polarization data and lensing reconstruction including external data from BAO and supernovae observations) [23]. The origin of this reionization of the IGM lies with the formation of first objects. The first generation of stars, the population III stars, are assumed to be very hot and massive. Their radiation ionizes the intergalactic gas in their vicinity and thus play an important role in the reionization of the universe.

The absolute spectrum of the CMB offers another possibility to gain insight into the physical conditions of the universe during its evolution. The COBE/FIRAS instrument measured the absolute spectrum of the CMB [36]. The data analysis showed that the CMB is very close to a perfect black body spectrum. The deviations from a Planck spectrum are very small indeed. These spectral distortions are classified in two types, namely, μ -type and y -type. They are generated at different times during the evolution of the universe as the spectral properties depend on the interactions of photons with the cosmic plasma. Deviations from the Planck spectrum result when energy is injected into the photon distribution

or the photon occupation number density is changed. This could be caused, e.g., by particle-antiparticle annihilation reactions $X\bar{X} \rightarrow \gamma\gamma$ or particle decay $X \rightarrow \dots\gamma$. Other interesting processes include the damping of density perturbations (cf. the Silk damping discussed above) [65] or damping of primordial magnetic fields [66, 67]. If this happens at very early times, at very high energies, corresponding to redshifts $z > 2 \times 10^6$ then bremsstrahlung and double Compton scattering are very efficient [68]. Since these interactions do not conserve photon numbers and redistribute photon frequencies a Planck spectrum can always be re-established. At later times and thus lower temperatures these interactions are no longer efficient. If the photon energy and/or occupation numbers are suddenly changed for redshifts between 2×10^6 and 4×10^4 photons interact with electrons by elastic Compton scattering. This changes the photon frequency distribution but not the occupation numbers resulting in a deviation from the Planck spectrum in the form of a Bose-Einstein spectrum, $f = [\exp(\mu + h\nu/kT_e) - 1]^{-1}$. This gives rise to a μ -type spectral distortion. For redshifts less than 5×10^4 elastic Compton scattering becomes inefficient. Any energy injection at this later times result in a y -type spectral distortion encoded by expanding the spectrum around the Planck spectrum $B_\nu(T)$ as

$$B_\nu(T) + y \frac{\partial S_y}{\partial y}, \quad y = \int \frac{k(T_e - T_\gamma)}{m_e c^2} d\tau \quad (113)$$

where T_e is the electron (matter) temperature and T_γ the photon temperature. τ is the optical depth (as defined above). This separation into epochs when either μ -type or y -type spectral distortions are generated has been re-fined by including an intermediate era generating an i -type spectral distortion [69]. The observational limits of COBE/FIRAS are $|\mu| < 9 \times 10^{-5}$ and $|y| < 1.5 \times 10^{-5}$ at 95% CL. There are discussions about future missions, such as the Primordial Inflation Explorer PIXIE [35], a potential NASA mission, to improve the COBE/FIRAS measurements of the absolute spectrum of the CMB.

The matter power spectrum provides the initial conditions for large scale structure formation. Linear density perturbations at some point reach a critical amplitude and enter the non-linear regime resulting in gravitational collapse and giving rise on larger scales to the nontrivial network of voids, sheets and filaments as observed in galaxy surveys as well as numerical simulations. The total matter power perturbation Δ_m has a contribution from the perturbation of the cold dark matter component Δ_c as well as in the baryons Δ_b ,

$$\Delta_m \equiv R_c \Delta_c + R_b \Delta_b, \quad (114)$$

where during the matter dominated era $R_i \equiv \frac{\rho_i}{\rho_{matter}}$, $i = b, c$. The corresponding matter spectrum is given by

$$P_m(k) = \frac{k^3}{2\pi^2} |\Delta_m|^2. \quad (115)$$

The total linear matter spectrum for the adiabatic mode is shown in figure 8. The evolution of the matter perturbation is determined in general by the continuity equation, (cf., e.g., [70])

$$\frac{\partial \Delta}{\partial t} + \frac{1}{a} \nabla \cdot (1 + \Delta) \mathbf{u} = 0 \quad (116)$$

and the Euler equation

$$\frac{\partial \Delta}{\partial t} + H \mathbf{u} + \frac{1}{a} [(\mathbf{u} \cdot \nabla) \mathbf{u} + \nabla \phi] = 0 \quad (117)$$

where $\mathbf{u} = \mathbf{v} - H\mathbf{x}$ is the peculiar velocity. Moreover, the gravitational potential ϕ is determined by the Poisson equation

$$\nabla^2 \phi = 4\pi G \rho a^2 \Delta. \quad (118)$$

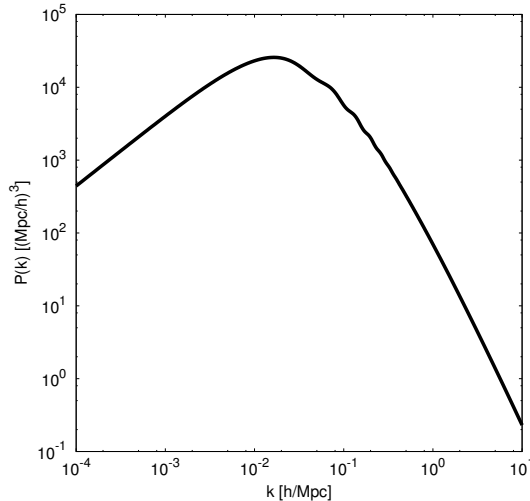


Fig. 8: The total linear matter power spectrum for the bestfit parameters from Planck 15 (TT,TE,EE+lowP+lensing+ext(BAO+JLA+ H_0)).

In the linear regime the equations of linear cosmological perturbation theory are recovered on sub horizon scales. Beyond the linear regime there are approaches to treat the evolution in a semianalytical way such as the Zeldovich approximation or the adhesion model. However, in order to obtain a complete picture of the evolution of the gravitational instabilities and subsequently the distribution of mass in the universe it is necessary to resort to numerical simulations. One of the largest simulations done is the Millenium Run of the Virgo Consortium [71] which uses the publicly available GADGET-2 code [72]. These gravitational N-body simulations generally confirm the picture of how the complex web of cosmic voids and filaments forms from the initial, linear density perturbations.

4 Dark matter and dark energy

Determining the velocity dispersions of galaxies in clusters Fritz Zwicky found the first evidence already in 1933 that there is more matter in the universe than is visible. Typically in spiral galaxies and galaxy clusters the member objects such as stars, gas clouds or galaxies follow a circular movement around the center of the corresponding structure. Plotting this radial dependence of the velocities yield the rotation curves. These show a flattening at large distances from the center which is an indication of the presence of non luminous matter. The rotational velocities can be estimated from the Doppler shift of spectral line(s) in the corresponding spectra. In particular, in galaxies the presence of neutral hydrogen gas clouds (HI clouds) allows to use the Doppler shifts of their 21cm line emission to determine the rotation velocities. The 21cm line or 1420 MHz radiation emission is due to the transition between the two hyperfine levels of the 1s ground state of hydrogen. This transition is highly forbidden with a transition probability of $2.6 \times 10^{-15} \text{ s}^{-1}$. However, there is a significant signal due to the large amount of hydrogen in the universe. Observations of the 21 cm line emission have been an important tool in galactic astronomy. It is becoming now the future of cosmological observations to explore the universe at high redshifts before reionization during the so called dark ages. These refer to the epoch after last scattering and before the formation of the first luminous objects such as the first generation stars (cf. e.g. [73]).

The dynamics can be described by a rather simple model which allows to get an estimate of the total mass of the structure (cf., e.g., the description in [16, 17]). The rotation curves determine the rotational velocity $v_{rot}(r)$ as a function of radial distance r from the center. Thus applying Newtonian dynamics yields to,

$$\frac{GM_r}{r^2} = \frac{v_{rot}^2(r)}{r} \rightarrow M_r = \frac{v_{rot}^2(r)r}{G}, \quad (119)$$

where M_r is the mass within a radius r . The luminous part of matter in spiral galaxies traces out a central region which has spherical geometry and a disk where the spiral arms are located. For a star in the central region one expects a linear evolution with radial distance of the rotational velocity. This uses the observation that typically in the central region the matter density is constant, thus $M_r \propto r^3$ and

$$\frac{Gr^3}{r^2} = \frac{v_{rot}^2}{r} \Rightarrow v_{rot} \sim r. \quad (120)$$

Since the mass of the central region is dominant over the contribution from the disk to a good approximation it can be assumed that on scales much beyond the central region the mass within a radius r is constant. Thus for a star or gas cloud outside the central region the rotational velocity behaves as $v_{rot} \sim r^{-\frac{1}{2}}$. Therefore the velocity should increase close to the center and decrease at large distances from the galactic center. However, the observed rotation curves show a distinct flattening reaching a plateau at large values of r . This behaviour can be explained by the presence of a galactic halo. Since it is not observed by electromagnetic radiation but only by its gravitational interaction it has to be made of dark matter. Observations indicate that the halo mass constitutes 80-90% of the total mass of a galaxy. This means that light traces only a small part of matter in galaxies and clusters of galaxies. Extrapolating to even larger scales, only a small of the total energy density is visible. The necessity of dark matter was already encountered in section 3 discussing the observations of the CMB. The bestfit models clearly require cold dark matter, e.g. the Planck 15 temperature and polarization data constrain $\Omega_{CDM}h^2 = 0.02225 \pm 0.00016$ at 68% CL [23]. Only including cold dark matter allows to generate the corresponding density perturbations in the photon radiation field to fit the observed CMB temperature and polarization anisotropies.

The big question now, of course, is what actually is dark matter. Primordial nucleosynthesis or BBN (big bang nucleosynthesis) predicts that the baryonic matter density parameter is $\Omega_b \simeq 0.05$. However the luminous part of matter in galaxies and galaxy clusters is estimated to have a density parameter of the order $\Omega_{lum} \simeq 0.01$. Therefore it is clear that part of dark matter is baryonic. Baryonic dark matter could be made of Jupiter-like objects or brown dwarfs. The latter are stars with masses less than 10% of the solar mass implying that their core temperatures are not high enough to start the nuclear reaction chain which fuels the radiation emitted by stars. Other possible contributions to dark matter could come from black holes. All these objects are called collectively MACHOs which stands for massive astrophysical halo objects. There is observational evidence for MACHOs since they act as gravitational lenses leading to the phenomenon of micro lensing. When a MACHO crosses the line-of-sight between us and a distant star it causes the star to appear brighter. This is due to the gravitational lens effect of the MACHO, namely, redirecting light into the direction of the observer and hence leading to a maximum in brightness. Observing this characteristic local maximum in the time depend light curves of distant stars could be evidence of a microlensing event. The MACHO project has identified several such microlensing events observing stars in the Magellanic Clouds and the Galactic Bulge [74].

The dominant part of dark matter in the universe is non baryonic. The standard theory is that these are particles created in the very early universe which are stable enough to survive at least until the present day. Candidates are massive neutrinos, axions and in general weakly interacting massive particles (WIMPs). These could be for example supersymmetric partners of the standard model of particle physics.

It is interesting to discuss the case of neutrinos in more detail. There are three families of light neutrinos in the standard model of particle physics. Oscillations between different neutrino flavors indicate limits on mass differences, e.g., measuring the flux of $\bar{\nu}_e$ from distant nuclear reactors the bestfit to the KamLAND data is $\Delta m^2 = 6.9 \times 10^{-5} \text{eV}^2$ for a mixing angle θ determined by $\sin^2 2\theta = 0.91$ [75]. A different bound on the masses of neutrinos can be obtained from cosmology as follows. Within standard big bang cosmology light neutrinos decouple at around 1 MeV as was described above. At this time neutrinos are relativistic implying that their temperature after decoupling from the rest of the cosmic plasma evolves as $T \propto 1/a$. At the time of neutrino decoupling, photons, neutrinos and the rest of matter which was in thermal equilibrium have the same temperature. Upto temperatures $T_{m_e} \simeq m_e \sim 0.5$

MeV the temperatures of photons and neutrinos are the same. However, below these temperatures electron-positron pair annihilate thereby reducing the number of relativistic degrees of freedom g_* . Before electron-positron annihilation the number of relativistic degrees of freedom is determined by photons, electrons and positrons, namely (cf. [17, 19])

$$g_*(\gamma, e^+, e^-) = 2 + \frac{7}{8} \times 4 = \frac{11}{2}. \quad (121)$$

At temperatures much below T_{m_e} only the photons contribute to g_* , i.e. $g_*(\gamma) = 2$. The entropy in a comoving volume is conserved implying that

$$S = g_*(Ta)^3 = \text{const.} \quad (122)$$

This yields to

$$\frac{(Ta)_{T < T_{m_e}}^3}{(Ta)_{T > T_{m_e}}^3} = \frac{g_{*, T > T_{m_e}}}{g_{*, T < T_{m_e}}} = \frac{11}{4}. \quad (123)$$

As the temperature of the neutrinos evolves as $1/a$ it follows that $(aT_\nu)_{T > T_{m_e}} = (aT_\nu)_{T < T_{m_e}}$. Since at temperatures $T > T_{m_e}$, neutrinos and photons have the same temperature, $T_\nu = T_\gamma$, equation (123) implies that for $T < T_{m_e}$

$$\frac{T_\gamma}{T_\nu} = \left(\frac{11}{4}\right)^{\frac{1}{3}} \simeq 1.4. \quad (124)$$

Therefore at present there is a background of (light) neutrinos at a temperature

$$T_{\nu,0} = \left(\frac{11}{4}\right)^{-\frac{1}{3}} T_{\gamma,0} \Rightarrow T_{\nu,0} \simeq 1.95 \text{ K}. \quad (125)$$

However, so far there has been no direct detection due to the obvious experimental difficulties. The number density of neutrinos and antineutrinos today is $n_\nu = \frac{3}{11} n_\gamma$ yielding $n_\nu = 113 \text{ cm}^{-3}$. This leads to a bound on the total mass assuming that the neutrino contribution does not overclose the universe, namely satisfying the condition, $\Omega = \frac{\rho}{\rho_c} \leq 1$. This implies the upper bound

$$\sum_{e,\mu,\tau} m_\nu c^2 \leq 47 \text{ eV}. \quad (126)$$

Light neutrinos as dark matter constitute hot dark matter since they are relativistic at decoupling ($T \sim 1 \text{ MeV}$). However, as they stream freely under the influence of gravity they tend to damp out density perturbations. Therefore the fraction of hot dark matter has to be subdominant w.r.t. cold dark matter.

The number of neutrino species N_{eff} can be constrained with BBN and CMB data. For SM neutrinos the relativistic energy density is given by

$$\rho_r = \rho_\gamma \left[1 + \frac{7}{8} \left(\frac{4}{11}\right)^{\frac{4}{3}} \right]. \quad (127)$$

A different value for N_{eff} implies a change in the expansion rate and hence the temperature at freeze-out of the conversion reaction between protons and neutrons, $n \leftrightarrow p$. This affects the ratio $\frac{n}{p} = e^{-\frac{Q}{T_D}}$ and the primordial helium fraction (cf equation (33)). Recent measurements of the primordial deuterium as well the helium fraction yield [21]

$$N_{eff}^{BBN} = 2.88 \pm 0.16. \quad (128)$$

Moreover, changes in the expansion rate at the time of photon decoupling t_{dec} have implications for the sound horizon r_s which is determined by

$$r_s = \int_0^{t_{dec}} c_s \frac{dt}{a} = \int_0^{a_{dec}} \frac{c_s}{a^2 H} da \quad (129)$$

as well as the photon diffusion damping scale k_D

$$k_D^{-2}(z) = \int_z^\infty \frac{dz}{6H(z)(1+R)\dot{\tau}} \left(\frac{16}{15} + \frac{R^2}{1+R} \right) \quad (130)$$

with the differential optical depth $\dot{\tau} = n_e \sigma_T \frac{a}{a_0}$ and the evolution of the baryon over photon fraction $R = \frac{3}{4} \frac{\Omega_{b,0}}{\Omega_{\gamma,0}} (1+z)^{-1}$. Thus increasing the effective number of relativistic degrees of freedom decreases the damping scale leading to an increase of small scale anisotropies. Planck 2015 temperature and polarization data only [23] constrain the contribution from light neutrinos and any other dark radiation such that at 68% CL $N_{eff} = 2.99 \pm 0.20$. When combined with large scale structure data this constraint becomes $N_{eff} = 3.04 \pm 0.18$.

The experimental evidence of neutrino oscillations implies that neutrinos are massive. The standard Λ CDM model assumes a mass hierarchy which is dominated by the heaviest neutrino mass eigenstate implying effectively one massive and two massless neutrinos with $\sum m_\nu = 0.06$ eV. However, a degenerate mass hierarchy is not excluded. The Planck 2015 temperature and polarization data only [23] constrain the sum of the neutrino masses to be $\sum m_\nu < 0.49$ eV and including large scale structure data leads to the upper bound $\sum m_\nu < 0.17$ eV at 95% CL. The absolute neutrino masses can be detected for example using the induced β decay resulting in limits on $m(\nu_e)$. Tritium β decay experiments such as KATRIN (e.g., [76]) and PTOLEMY [77] are expected to provide the strongest bounds.

The CMB clearly indicates the presence of a substantial amount of cold dark matter. The term weakly interacting massive particles (WIMPs) has been introduced to describe hypothetical particles which could constitute the cold dark matter in the universe. These are particles which are assumed to be non relativistic at the time of their decoupling from the rest of the cosmic plasma. WIMP candidates are, e.g., supersymmetric particles. The flat FRW model imposes the constraint on the density parameter of any WIMP candidate χ that $\Omega_\chi < 1$ (e.g. [17]). Since at present WIMPs are non relativistic, the density parameter is determined by

$$\Omega_\chi = \frac{N(T_0)m}{\rho_c} \quad (131)$$

assuming a WIMP candidate χ with mass m . Moreover, $N(T_0) = N(T_D) \left(\frac{T_0}{T_D} \right)^3$ is the number density today and T_0 and T_D are the temperature today and at "freeze-out" of $\chi\bar{\chi}$ pair production, respectively. Assuming that freeze-out takes place during the radiation-dominated era the Hubble parameter is given by Eq. (31) and the expression for the number density of non relativistic species, Eq. (25). Moreover at freeze-out the temperature T_D can be estimated by the condition $N(T_D)\langle\sigma v\rangle = H(T_D)$ where σ is the $\chi\bar{\chi}$ annihilation cross section and v their relative velocity. Therefore the condition $\Omega_\chi < 1$ can be used to constrain model parameters.

There are direct as well as indirect experimental searches for WIMP detection within our galaxy. The former one uses hypothetical WIMP-nucleon scattering within a detector. These signals might have angular and time dependence due to the motion of the earth. There is a whole range of direct experimental searches going on presently, such as XENON1T [78] or LUX [79]. Indirect searches use decay products and/or excess radiation of potential WIMP decay or annihilation interactions. There have been indications of excess γ ray radiation from the galactic center from the FERMI satellite (e.g., [80]) as well as positron excess from data of the the AMS [81] as well as the HEAT [82] and PAMELA [83] experiments.

As was already discussed in section 2 observations show that at present the largest contribution to the energy density in the universe is due to dark energy. A term which has been coined to deal with its unknown origins. Apart from the observations of supernovae and the CMB anisotropies there is also evidence from large scale structure. Here we are going to focus on the evidence from baryon acoustic oscillations (BAO) which have been observed in galaxy correlation functions. In section 3 the acoustic oscillations in the tightly coupled baryon-photon fluid have been discussed. These give rise to the observed acoustic peak structure in the angular power spectrum of the temperature and polarization anisotropies of the CMB radiation. Moreover, these also manifest themselves as wiggles in the matter power spectrum (cf., Fig. 8). This has been observed, e.g., in the Sloan Digital Sky Survey (SDSS) [84, 85]. In real space, in the two-point galaxy correlation function these acoustic oscillations in the photon-fluid lead to the already well-detected BAO peak determined by the size of the sound horizon at decoupling (cf. Eq. (129)), e.g. in the SDSS [86] the BAO peak is detected at $100h^{-1}$ Mpc. This is the largest scale upto which a sound wave can travel before the sound speed in the baryon fluid decreases dramatically due to the decoupling of the photons. There are several factors which complicate the detection of the BAO peak in galaxy redshift surveys due to their low redshift such as the nonlinearity of the density matter field, bias and redshift distortions due to small scale velocities.

The preferred angular separation scale of galaxies marked by the BAO peak can be used as a standard ruler since it is related to the known scale of the sound horizon at decoupling, r_s . This provides a measurement across the line of sight determining $r_s/D_A(z)$ and hence the angular diameter distance D_A as a function of z . There is also the possibility to measure the redshift separation along the line of sight given by $r_s H(z)$. The uncertainty to convert redshifts into distances along and across the line of sight has led to the definition of the parameter [86]

$$D_V(z) = \left[(1+z)^2 D_A(z) \frac{cz}{H(z)} \right]^{\frac{1}{3}} \quad (132)$$

whose graph as a function of redshift results in the equivalent of BAO Hubble diagram indicating the present day acceleration of the universe (e.g., from the WiggleZ survey, [87, 88]).

Upto now it is an open problem of what actually causes the observed acceleration of the universe. There are many possible explanations ranging from a cosmological constant, scalar fields, a gravity theory different from Einstein general relativity, brane world models, etc. However, there is no compelling candidate. In terms of interpreting the data the simplest approach is to assume a cosmological constant adding just one parameter, Ω_Λ which has an equation of state $p = -\rho$. Considering a more general equation of state $p = w\rho$ where w is a constant the constraint from observations of the CMB such as, e.g., Planck 15 plus astrophysical data gives $w = -1.006 \pm 0.045$ [23].

If one were to try to connect the observed value of Ω_Λ with the expectation value of a vacuum energy density a simple estimate shows that there is a huge discrepancy between prediction and observation. Describing vacuum fluctuations in terms of a set of harmonic oscillators at zero point energy yields the total energy per unit volume of all oscillators (e.g. [17])

$$\epsilon = \frac{\hbar}{4\pi^2} \int \omega_k k^2 dk. \quad (133)$$

Introducing an upper cut-off k_m corresponding to a maximal energy E_m and considering the relativistic limit leads to $\epsilon = \frac{E_m^4}{16\pi^2(\hbar c)^3}$. A natural energy scale in the problem is the Planck scale M_P for which $\epsilon \sim 10^{121}$ GeV leading to

$$\frac{\rho_\Lambda}{\epsilon} \simeq 10^{-121} \quad (134)$$

for $\rho_\Lambda = \Omega_\Lambda \rho_c$ and $\Omega_\Lambda \sim 0.7$ and $\rho_c \simeq 5 \text{ GeV m}^{-3}$. This encapsulates the cosmological constant problem (cf., e.g. [89]).

5 Conclusions

In these lectures we have provided an overview of the current cosmological model together with observational evidence. Roughly 95 % of the energy density of the universe is in the "dark" sector for which there is observational evidence but no clear theoretical understanding. There is no shortage of proposals of models ranging from modifying gravity "by hand", string theory inspired models to various extensions of the standard model of particle physics. However, so far there is no obvious ("natural") model which stands out. There will be more data and at higher precision from astrophysical and cosmological observations in the near future. These can also provide useful additional constraints on particle physics models.

6 Acknowledgements

I would like to thank the organizers for inviting me to present these lectures at the CERN Latin-American School of High Energy Physics CLASHEP 2015, which took place in Ibarra, Ecuador. It was a very interesting school and the organization was marvellous. I would like to thank NORDITA in Stockholm, Sweden, for hospitality where part of these lecture notes were written. Financial support by Spanish Science Ministry grants FPA2015-64041-C2-2-P, FIS2012-30926 and CSD2007-00042 is gratefully acknowledged.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

We acknowledge the use of the Legacy Archive for Microwave Background Data Analysis (LAMBDA), part of the High Energy Astrophysics Science Archive Center (HEASARC). HEASARC/LAMBDA is a service of the Astrophysics Science Division at the NASA Goddard Space Flight Center.

Appendices

A A brief introduction to the Friedmann-Robertson-Walker models

In the standard model of cosmology gravity is described by general relativity. As in special relativity space and time are united to describe a space-time. The invariant line element of special relativity is given by $ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu$ where

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A.1})$$

is the Minkowski metric. In general relativity this is determined by $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ for a general metric $g_{\mu\nu}$. Space-time is curved. The dynamics is determined by Einstein's equations,

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu} \quad (\text{A.2})$$

where the Ricci tensor is defined by $R_{\mu\nu} = \frac{\partial\Gamma_{\mu\nu}^\sigma}{\partial x^\sigma} - \frac{\partial\Gamma_{\mu\sigma}^\nu}{\partial x^\nu} + \Gamma_{\rho\sigma}^\nu\Gamma_{\mu\nu}^\rho - \Gamma_{\rho\nu}^\sigma\Gamma_{\mu\sigma}^\rho$, the Christoffel symbols are defined by $\Gamma_{\nu\rho}^\mu = \frac{g^{\mu\sigma}}{2} \left(\frac{\partial g_{\sigma\rho}}{\partial x^\nu} + \frac{\partial g_{\nu\sigma}}{\partial x^\rho} - \frac{\partial g_{\nu\rho}}{\partial x^\sigma} \right)$ and $R = g^{\mu\nu}R_{\mu\nu}$ is the Ricci scalar. Repeated indices in pairs with one contravariant and one covariant are summed over. Greek indices take values between 0 and 3. $T_{\mu\nu}$ is the energy-momentum tensor which is conserved, satisfying $\nabla_\nu T^{\mu\nu} = 0$. Speaking in an illustrative way, Einstein's equations encapsulate that geometry determines matter distribution and evolution and vice versa. Space-time is curved by the presence of matter. That is why, e.g., the trajectory of light from distant sources is deviated by the sun. This deflection angle is one of the classical tests of general relativity.

Einstein's equations are very complex. There are no general solutions known. It is always necessary to assume some degree of symmetry in order to find solutions. The Friedmann-Robertson-Walker solutions are isotropic and homogeneous and are described by the metric

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2\theta d\phi^2) \right]. \quad (\text{A.3})$$

The parameter k has only three values: $k = 0$ for a flat universe, $k = -1$ for an open one and $k = 1$ for a closed universe. In many cases matter can be described by a perfect fluid with 4-velocity u^μ whose energy momentum tensor is given by

$$T_{\mu\nu} = [\rho(t) + P(t)] u_\mu u_\nu + p(t)g_{\mu\nu} \quad (\text{A.4})$$

where $\rho(t)$ and $p(t)$ are the energy density and pressure, respectively, which are only functions of time in a Friedmann-Robertson-Walker background.

References

- [1] J. C. Mather, E. Cheng, R. Shafer, C. Bennett, N. Boggess, *et al.*, "A Preliminary measurement of the Cosmic Microwave Background spectrum by the Cosmic Background Explorer (COBE) satellite," *Astrophys.J.*, vol. 354, pp. L37–L40, 1990.
- [2] J. C. Mather, E. Cheng, D. Cottingham, R. Eplee, D. Fixsen, *et al.*, "Measurement of the Cosmic Microwave Background spectrum by the COBE FIRAS instrument," *Astrophys.J.*, vol. 420, pp. 439–444, 1994.
- [3] D. Fixsen, E. Cheng, J. Gales, J. C. Mather, R. Shafer, *et al.*, "The Cosmic Microwave Background spectrum from the full COBE FIRAS data set," *Astrophys.J.*, vol. 473, p. 576, 1996, astro-ph/9605054.
- [4] S. Das, T. Louis, M. R. Nolta, G. E. Addison, E. S. Battistelli, *et al.*, "The Atacama Cosmology Telescope: Temperature and Gravitational Lensing Power Spectrum Measurements from Three Seasons of Data," 2013, 1301.1037.
- [5] C. Reichardt, L. Shaw, O. Zahn, K. Aird, B. Benson, *et al.*, "A measurement of secondary cosmic microwave background anisotropies with two years of South Pole Telescope observations," *Astrophys.J.*, vol. 755, p. 70, 2012, 1111.0932.
- [6] W. C. Jones *et al.*, "A Measurement of the angular power spectrum of the CMB temperature anisotropy from the 2003 flight of BOOMERANG," *Astrophys. J.*, vol. 647, pp. 823–832, 2006, astro-ph/0507494.
- [7] G. Hinshaw *et al.*, "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," *Astrophys.J.Suppl.*, vol. 208, p. 19, 2013, 1212.5226.

- [8] P. Ade *et al.*, “Planck 2013 results. XVI. Cosmological parameters,” 2013, 1303.5076.
- [9] V. de Lapparent, M. J. Geller, and J. P. Huchra, “A Slice of the universe,” *Astrophys. J.*, vol. 302, pp. L1–L5, 1986.
- [10] S. Alam *et al.*, “The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III,” *Astrophys. J. Suppl.*, vol. 219, no. 1, p. 12, 2015, 1501.00963.
- [11] K. S. Dawson *et al.*, “The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data,” *Astron. J.*, vol. 151, p. 44, 2016, 1508.04473.
- [12] L. Amendola *et al.*, “Cosmology and fundamental physics with the Euclid satellite,” *Living Rev. Rel.*, vol. 16, p. 6, 2013, 1206.1225.
- [13] D. H. Lyth and A. R. Liddle, *The primordial density perturbation: Cosmology, inflation and the origin of structure*. 2009.
- [14] R. Durrer, “The cosmic microwave background,” *Cambridge, UK: Cambridge Univ. Pr. (2008)*, 2008.
- [15] J. Lesgourgues, G. Mangano, G. Miele, and S. Pastor, “Neutrino cosmology,” *Cambridge, UK: Cambridge Univ. Pr. (2013)*, 2013.
- [16] P. Coles and F. Lucchin, “Cosmology: The Origin and evolution of cosmic structure,” *Chichester, UK: Wiley (1995) 449 p*, 1995.
- [17] D. H. Perkins, “Particle astrophysics,” *Oxford, UK: Univ. Pr. (2003) 256 p*, 2003.
- [18] P. Peter and J.-P. Uzan, “Primordial cosmology,” *Oxford, UK: Univ. Pr. (2009)*, 2009.
- [19] E. W. Kolb and M. S. Turner, “The Early Universe,” *Front. Phys.*, vol. 69, pp. 1–547, 1990.
- [20] H. Mo, F. van den Bosch, and S. White, “Galaxy Formation and Evolution,” *Cambridge, UK: Cambridge Univ. Pr. (2010)*, 2010.
- [21] K. A. Olive *et al.*, “Review of Particle Physics,” *Chin. Phys.*, vol. C38, p. 090001, 2014 and 2015 update.
- [22] A. G. Riess, L. Macri, S. Casertano, H. Lampeitl, H. C. Ferguson, A. V. Filippenko, S. W. Jha, W. Li, and R. Chornock, “A 3% Solution: Determination of the Hubble Constant with the Hubble Space Telescope and Wide Field Camera 3,” *Astrophys. J.*, vol. 730, p. 119, 2011, 1103.2976. [Erratum: *Astrophys. J.* 732, 129 (2011)].
- [23] P. A. R. Ade *et al.*, “Planck 2015 results. XIII. Cosmological parameters,” 2015, 1502.01589.
- [24] K. N. Abazajian *et al.*, “The Seventh Data Release of the Sloan Digital Sky Survey,” *Astrophys. J. Suppl.*, vol. 182, pp. 543–558, 2009, 0812.0649.
- [25] S. Perlmutter, “Supernovae, dark energy, and the accelerating universe: The Status of the cosmological parameters,” *Int. J. Mod. Phys.*, vol. A15S1, pp. 715–739, 2000. [715(1999)].
- [26] B. P. Schmidt *et al.*, “The High Z supernova search: Measuring cosmic deceleration and global curvature of the universe using type Ia supernovae,” *Astrophys. J.*, vol. 507, pp. 46–63, 1998, astro-ph/9805200.
- [27] A. R. Liddle and D. H. Lyth, “Cosmological inflation and large scale structure,” *Cambridge, UK: Univ. Pr. (2000) 400 p*, 2000.
- [28] S. Burles, K. M. Nollett, J. W. Truran, and M. S. Turner, “Sharpening the predictions of big bang nucleosynthesis,” *Phys. Rev. Lett.*, vol. 82, pp. 4176–4179, 1999, astro-ph/9901157.
- [29] D. Tytler, J. M. O’Meara, N. Suzuki, and D. Lubin, “Deuterium and the baryonic density of the universe,”
- [30] F. Iocco, G. Mangano, G. Miele, O. Pisanti, and P. D. Serpico, “Primordial Nucleosynthesis: from precision cosmology to fundamental physics,” *Phys. Rept.*, vol. 472, pp. 1–76, 2009, 0809.0631.
- [31] A. D. Linde, “Particle physics and inflationary cosmology,” *Contemp. Concepts Phys.*, vol. 5, pp. 1–362, 1990, hep-th/0503203.

- [32] A. Challinor and H. Peiris, “Lecture notes on the physics of cosmic microwave background anisotropies,” *AIP Conf. Proc.*, vol. 1132, pp. 86–140, 2009, 0903.5158.
- [33] W. Hu and M. J. White, “CMB anisotropies: Total angular momentum method,” *Phys. Rev.*, vol. D56, pp. 596–615, 1997, astro-ph/9702170.
- [34] D. J. Fixsen, “The Temperature of the Cosmic Microwave Background,” *Astrophys. J.*, vol. 707, pp. 916–920, 2009, 0911.1955.
- [35] A. Kogut *et al.*, “The Primordial Inflation Explorer (PIXIE): A Nulling Polarimeter for Cosmic Microwave Background Observations,” *JCAP*, vol. 1107, p. 025, 2011, 1105.2044.
- [36] D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright, “The Cosmic Microwave Background spectrum from the full COBE FIRAS data set,” *Astrophys. J.*, vol. 473, p. 576, 1996, astro-ph/9605054.
- [37] H. Kodama and M. Sasaki, “Cosmological Perturbation Theory,” *Prog. Theor. Phys. Suppl.*, vol. 78, pp. 1–166, 1984.
- [38] S. Weinberg, “Cosmology,” *Oxford, UK: Oxford Univ. Pr. (2008) 593 p*, 2008.
- [39] E. Bertschinger, “COSMICS: cosmological initial conditions and microwave anisotropy codes,” 1995, astro-ph/9506070.
- [40] U. Seljak and M. Zaldarriaga, “A Line of sight integration approach to cosmic microwave background anisotropies,” *Astrophys. J.*, vol. 469, pp. 437–444, 1996, astro-ph/9603033.
- [41] A. Lewis, A. Challinor, and A. Lasenby, “Efficient computation of CMB anisotropies in closed FRW models,” *Astrophys. J.*, vol. 538, pp. 473–476, 2000, astro-ph/9911177.
- [42] M. Doran, “CMBEASY: an object oriented code for the cosmic microwave background,” *JCAP*, vol. 0510, p. 011, 2005, astro-ph/0302138.
- [43] J. Lesgourgues, “The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview,” 2011, 1104.2932.
- [44] D. Blas, J. Lesgourgues, and T. Tram, “The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes,” *JCAP*, vol. 1107, p. 034, 2011, 1104.2933.
- [45] J. Lesgourgues, “The Cosmic Linear Anisotropy Solving System (CLASS) III: Comparison with CAMB for LambdaCDM,” 2011, 1104.2934.
- [46] J. Lesgourgues and T. Tram, “The Cosmic Linear Anisotropy Solving System (CLASS) IV: efficient implementation of non-cold relics,” *JCAP*, vol. 1109, p. 032, 2011, 1104.2935.
- [47] J. Lesgourgues and S. Pastor, “Massive neutrinos and cosmology,” *Phys. Rept.*, vol. 429, pp. 307–379, 2006, astro-ph/0603494.
- [48] A. Mack, T. Kahniashvili, and A. Kosowsky, “Microwave background signatures of a primordial stochastic magnetic field,” *Phys. Rev.*, vol. D65, p. 123004, 2002, astro-ph/0105504.
- [49] J. R. Shaw and A. Lewis, “Massive Neutrinos and Magnetic Fields in the Early Universe,” *Phys. Rev.*, vol. D81, p. 043517, 2010, 0911.2714.
- [50] K. E. Kunze, “Effects of helical magnetic fields on the cosmic microwave background,” *Phys. Rev.*, vol. D85, p. 083004, 2012, 1112.4797.
- [51] P. D. Naselsky, D. I. Novikov, and I. D. Novikov, “The Physics of the Cosmic Microwave Background,” Aug. 2011.
- [52] D. Langlois, “Lectures on inflation and cosmological perturbations,” *Lect. Notes Phys.*, vol. 800, pp. 1–57, 2010, 1001.5259.
- [53] S. Weinberg, “Can non-adiabatic perturbations arise after single-field inflation?,” *Phys. Rev.*, vol. D70, p. 043541, 2004, astro-ph/0401313.
- [54] P. A. R. Ade *et al.*, “Planck 2015 results. XX. Constraints on inflation,” 2015, 1502.02114.
- [55] W. Hu and M. J. White, “CMB anisotropies: Total angular momentum method,” *Phys. Rev.*, vol. D56, pp. 596–615, 1997, astro-ph/9702170.

- [56] J. Lesgourgues and T. Tram, “Fast and accurate CMB computations in non-flat FLRW universes,” *JCAP*, vol. 1409, no. 09, p. 032, 2014, 1312.2697.
- [57] N. Aghanim *et al.*, “Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters,” *Submitted to: Astron. Astrophys.*, 2015, 1507.02704.
- [58] W. Hu and N. Sugiyama, “Anisotropies in the cosmic microwave background: An Analytic approach,” *Astrophys. J.*, vol. 444, pp. 489–506, 1995, astro-ph/9407093.
- [59] J. D. Jackson, “Classical Electrodynamics,” 1998.
- [60] S. Chandrasekhar, “Radiative transfer,” 1960.
- [61] J. Tinbergen, “Astronomical Polarimetry,” Aug. 2005.
- [62] S. Dodelson, “Modern Cosmology,” *Amsterdam, Netherlands: Academic Pr. (2003) 440 p*, 2003.
- [63] S. Naess *et al.*, “The Atacama Cosmology Telescope: CMB Polarization at $200 < \ell < 9000$,” *JCAP*, vol. 1410, no. 10, p. 007, 2014, 1405.5524.
- [64] A. T. Crites *et al.*, “Measurements of E-Mode Polarization and Temperature-E-Mode Correlation in the Cosmic Microwave Background from 100 Square Degrees of SPTpol Data,” *Astrophys. J.*, vol. 805, no. 1, p. 36, 2015, 1411.1042.
- [65] W. Hu, D. Scott, and J. Silk, “Power spectrum constraints from spectral distortions in the cosmic microwave background,” *Astrophys. J.*, vol. 430, pp. L5–L8, 1994, astro-ph/9402045.
- [66] K. Jedamzik, V. Katalinic, and A. V. Olinto, “A Limit on primordial small scale magnetic fields from CMB distortions,” *Phys. Rev. Lett.*, vol. 85, pp. 700–703, 2000, astro-ph/9911100.
- [67] K. E. Kunze and E. Komatsu, “Constraining primordial magnetic fields with distortions of the black-body spectrum of the cosmic microwave background: pre- and post-decoupling contributions,” *JCAP*, vol. 1401, p. 009, 2014, 1309.7994.
- [68] W. Hu and J. Silk, “Thermalization and spectral distortions of the cosmic background radiation,” *Phys. Rev.*, vol. D48, pp. 485–502, 1993.
- [69] R. Khatri and R. A. Sunyaev, “Beyond y and μ : the shape of the CMB spectral distortions in the intermediate epoch, $1.5 \times 10^4 < z < 2 \times 10^5$,” *JCAP*, vol. 1209, p. 016, 2012, 1207.6654.
- [70] A. Cooray and R. K. Sheth, “Halo models of large scale structure,” *Phys. Rept.*, vol. 372, pp. 1–129, 2002, astro-ph/0206508.
- [71] V. Springel *et al.*, “Simulating the joint evolution of quasars, galaxies and their large-scale distribution,” *Nature*, vol. 435, pp. 629–636, 2005, astro-ph/0504097.
- [72] V. Springel, “The Cosmological simulation code GADGET-2,” *Mon. Not. Roy. Astron. Soc.*, vol. 364, pp. 1105–1134, 2005, astro-ph/0505010.
- [73] J. R. Pritchard and A. Loeb, “21-cm cosmology,” *Rept. Prog. Phys.*, vol. 75, p. 086901, 2012, 1109.6012.
- [74] C. L. Thomas *et al.*, “Galactic Bulge microlensing events from the MACHO Collaboration,” *Astrophys. J.*, vol. 631, pp. 906–934, 2005, astro-ph/0410341.
- [75] K. Eguchi *et al.*, “First results from KamLAND: Evidence for reactor anti-neutrino disappearance,” *Phys. Rev. Lett.*, vol. 90, p. 021802, 2003, hep-ex/0212021.
- [76] G. Drexlin, V. Hannen, S. Mertens, and C. Weinheimer, “Current direct neutrino mass experiments,” *Adv. High Energy Phys.*, vol. 2013, p. 293986, 2013, 1307.0101.
- [77] S. Betts *et al.*, “Development of a Relic Neutrino Detection Experiment at PTOLEMY: Princeton Tritium Observatory for Light, Early-Universe, Massive-Neutrino Yield,” in *Community Summer Study 2013: Snowmass on the Mississippi (CSS2013) Minneapolis, MN, USA, July 29-August 6, 2013*, 2013, 1307.4738.
- [78] E. Aprile *et al.*, “Physics reach of the XENON1T dark matter experiment,” *Submitted to: JCAP*, 2015, 1512.07501.
- [79] D. S. Akerib *et al.*, “First spin-dependent WIMP-nucleon cross section limits from the LUX exper-

- iment,” 2016, 1602.03489.
- [80] F. Calore, I. Cholis, C. McCabe, and C. Weniger, “A Tale of Tails: Dark Matter Interpretations of the Fermi GeV Excess in Light of Background Model Systematics,” *Phys. Rev.*, vol. D91, no. 6, p. 063003, 2015, 1411.4647.
 - [81] L. Accardo *et al.*, “High Statistics Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–500 GeV with the Alpha Magnetic Spectrometer on the International Space Station,” *Phys. Rev. Lett.*, vol. 113, p. 121101, 2014.
 - [82] J. J. Beatty *et al.*, “New measurement of the cosmic-ray positron fraction from 5 to 15-GeV,” *Phys. Rev. Lett.*, vol. 93, p. 241102, 2004, astro-ph/0412230.
 - [83] O. Adriani *et al.*, “An anomalous positron abundance in cosmic rays with energies 1.5–100 GeV,” *Nature*, vol. 458, pp. 607–609, 2009, 0810.4995.
 - [84] M. Tegmark *et al.*, “Cosmological Constraints from the SDSS Luminous Red Galaxies,” *Phys. Rev.*, vol. D74, p. 123507, 2006, astro-ph/0608632.
 - [85] W. J. Percival, S. Cole, D. J. Eisenstein, R. C. Nichol, J. A. Peacock, A. C. Pope, and A. S. Szalay, “Measuring the Baryon Acoustic Oscillation scale using the SDSS and 2dFGRS,” *Mon. Not. Roy. Astron. Soc.*, vol. 381, pp. 1053–1066, 2007, 0705.3323.
 - [86] D. J. Eisenstein *et al.*, “Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies,” *Astrophys. J.*, vol. 633, pp. 560–574, 2005, astro-ph/0501171.
 - [87] C. Blake *et al.*, “The WiggleZ Dark Energy Survey: testing the cosmological model with baryon acoustic oscillations at $z=0.6$,” *Mon. Not. Roy. Astron. Soc.*, vol. 415, pp. 2892–2909, 2011, 1105.2862.
 - [88] C. Blake *et al.*, “The WiggleZ Dark Energy Survey: mapping the distance-redshift relation with baryon acoustic oscillations,” *Mon. Not. Roy. Astron. Soc.*, vol. 418, pp. 1707–1724, 2011, 1108.2635.
 - [89] S. Weinberg, “The Cosmological Constant Problem,” *Rev. Mod. Phys.*, vol. 61, pp. 1–23, 1989.

Experimental Facilities at the High Energy Frontier

P. Jenni

Albert-Ludwigs-University Freiburg, Germany, and CERN, Geneva, Switzerland

Abstract

The main theme of the lectures covered the experimental work at hadron colliders, with a clear focus on the Large Hadron Collider (LHC) and on the roadmap that led finally to the discovery of the Higgs boson. The lectures were not a systematic course on machine and detector technologies, but rather tried to give a physics-motivated overview of many experimental aspects that were all relevant for making the discovery. The actual lectures covered a much broader scope than what is documented here in this write-up. The successful concepts for the experiments at the LHC have benefitted from the experience gained with previous generations of detectors at lower-energy machines. The lectures included also an outlook to the future experimental programme at the LHC, with its machine and experiments upgrades, as well as a short discussion of possible facilities at the high energy frontier beyond LHC.

Keywords

Large hadron collider; detector; trigger; particle physics; Higgs particle; physics beyond the Standard Model.

1 Introduction

Experimental facilities at the High Energy Frontier (HEF) is a very broad topic, impossible to cover in even four lectures. Already during the lectures, but even much more so in this write-up, the main theme will be the long journey to the Higgs boson discovery and beyond at the Large Hadron Collider (LHC).

The HEF is only one of the three main pillars of experimental particle physics. It complements experiments that study the physics at the intensity frontier and at the cosmic frontier. The three areas have considerable overlap that provide for exciting synergies. Without exaggeration one can claim that particle physics has never been such an exciting research field with great promises for new fundamental discoveries as we live through these years. The HEF in itself is a broad and lively field with complementary approaches using either proton-proton (pp), electron-positron (e⁺e⁻) or eventually electron-proton (ep) collisions. For these lectures, however, the LHC will be in the focus, given the topical discovery of the Higgs boson and LHC's unique status of being the only running high energy collider for many years to come.

The discovery of a scalar boson, which shows within the present statistical precision achieved all the expected properties for the famous Higgs boson, announced by the ATLAS and CMS Collaborations on 4th July 2012 [1,2], was the culminating experimental triumph for the Standard Model [3-11]. The Standard Model (SM) of particle physics is one of the most remarkable achievements of physics over the past 50 years. Its descriptive and predictive power has been experimentally demonstrated with unprecedented accuracy in many generations of experiments ranging from low to high energies. The SM comprises the fundamental building blocks of all visible matter, with the three fermion families of quarks and leptons, and their interactions via three out of the four fundamental interactions mediated by bosons, namely the massless photon for the

electromagnetism, the heavy W and Z bosons for the weak force, these two interactions unified in the electroweak theory, and the massless gluons for the strong interaction.

In order to solve the mystery of generation of mass, a spontaneous symmetry-breaking mechanism was proposed introducing a complex scalar field that permeates the entire Universe. This mechanism, known as Brout-Englert-Higgs (BEH) mechanism [6-9], gives the W and Z their large masses and leaves the photon massless. Interaction with the scalar field generates masses to the quarks and leptons in proportion to the strength of their couplings to it. This field leads to an additional massive scalar boson as its quantum, called the Higgs boson. After the discovery of the W and Z bosons in the early 1980s, the hunt for the Higgs boson, considered to be the keystone of the SM, became a central theme in particle physics, and also a primary motivation for the LHC. Finding the Higgs boson would establish the existence of the postulated BEH field, and thereby marking a crucial step in the understanding of Nature.

There is a vast body of literature available for ‘telling the story’ of the Higgs discovery at the LHC. Somehow unjustified and selectively, the following accounts are heavily based on a few didactic articles co-authored by the lecturer with longtime colleagues since the early years of the LHC adventure [12,13] and hadron collider experiments even before [14]. For scientifically rigorous accounts and full references the reader is referred to the original publications [1,2] and the full references therein, as well as to the updated public reference lists of publications by ATLAS [15] and CMS [16].

2 Hadron Collider Experiments Preceding the LHC

The history of hadron colliders started almost 45 years ago. A dedicated account of the evolution of hadron collider experimentation is given in [14], containing also detailed references, where the very impressive growth and sophistication, both for the detector concepts as well as for the analysis methods, is discussed.

The first hadron collider was the about 1 km circumference CERN Intersecting Storage Rings (ISR), commissioned in 1971 with proton-proton (pp) collision energies E between 23 and 63 GeV. Experiments were located at eight beam crossing points of the two separate magnet rings. The CERN SPS Collider (SppbarS), colliding p and antiprotons ($p\bar{p}$) with $E = 546$ and 630 GeV, followed in 1981. The $p\bar{p}$ were produced using intense proton beams hitting a target, and collected in a dedicated accumulator ring, and their phase space was reduced to dense bunches by an electromagnetic feedback system derived from sensing energy fluctuations in the beam itself (stochastic cooling). The 1.3 Tesla conventional bending magnets were housed in the 6.9 km SPS tunnel. The two general-purpose detectors, UA1 and UA2, where the W and Z were discovered, shared the collider with a few smaller experiments. The Fermilab $p\bar{p}$ Tevatron Collider with $E = 2$ TeV used 4.2 Tesla superconducting magnets in a 6.3 km ring. It began operation in 1987, serving two major detectors, CDF (Collider Detector Facility) and $D\bar{0}$ (named after its location on the ring), which provided the top quark discovery, and also a few small specialized experiments. Finally, the pp Superconducting Super Collider (SSC) in Texas would have had $E = 40$ TeV, but was cancelled by the U.S. Congress in 1993 well before its construction completion.

Hadron collider detectors have evolved in size and complexity as physics questions have changed, and as the technology advanced. Each generation of detectors built on the previous experience. The growth in size stemmed in part from the increase in typical particle energies, but also in response to the need for higher precision and hence more sophisticated detectors.

The early ISR experiments, as customary from previous fixed target experiments, had rather specific physics goals that were addressed with detectors covering limited solid angles using the detector technologies from the fixed target era. Toward the end of the ISR period, new discoveries

stimulated a new detector paradigm. The requirement to study new massive particles, such as the W and Z that transmit the weak force, and the search for the top quark, in which several decay particles emerge over a large angular region and at large momenta transverse to the beam axis (p_T), stimulated experiments to become more inclusive in their coverage. Thus, starting with the UA1 and UA2 experiments, the general purpose detectors followed a more hermetic design with cylindrical shells of subdetectors surrounding the beam vacuum pipe and covering nearly the full 4π solid angle. The innermost ‘tracker’ layer records the ionization tracks of charged particles moving in a magnetic field to allow the measurement of their momenta, followed by ‘calorimeters’ to measure the energies of hadrons, electrons and photons, and an outer layer that measures and identifies the muons which penetrate the calorimeter. Implementation in specific experiments varied, depending on the physics emphasis and the available technology. Figure 1 illustrates with some examples this detector evolution.

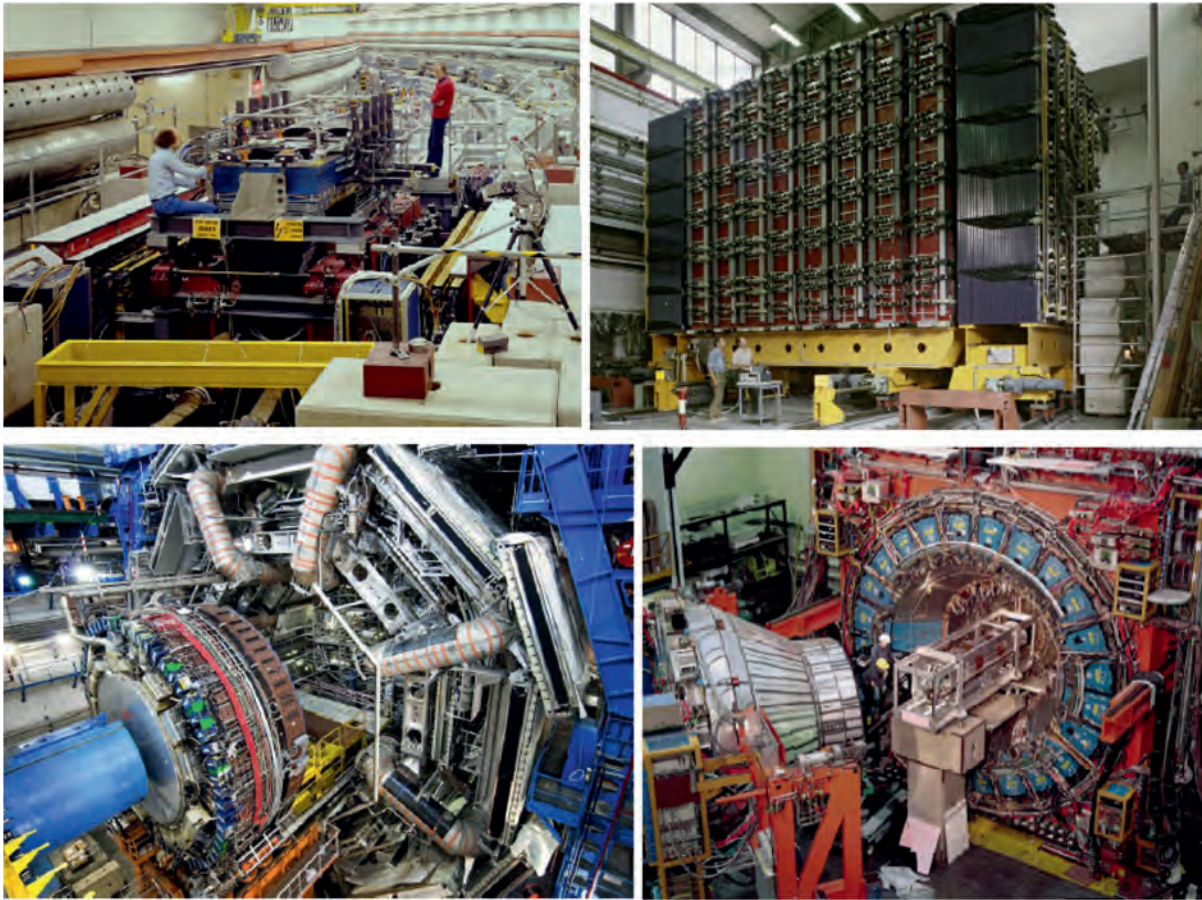


Fig. 1: Clockwise from top left the R702 experiment at the ISR, the UA1 experiment at the SpbarS, the CDF experiment at the Tevatron, and the ATLAS experiment at the LHC, whose approximate diameters are 6, 10, 12 and 24 m respectively.

The trackers evolved from multiwire proportional chambers and drift chambers that measured the location of charged particle ionizations with precisions of hundreds of μm , to higher resolution scintillating fibers and ultimately silicon microstrip and pixel detectors with few μm resolutions. To allow track momentum measurement, solenoids with magnetic field along the beam axis were favoured as this optimized the transverse momentum resolution needed by the high p_T physics programme. The fields increased from about 1 Tesla in earlier detectors to 4 Tesla in the LHC CMS solenoid. The resolution improvement from silicon detectors brought the new capability to sense the short distance between the production and decay of hadrons containing a b -quark, a signature of many kinds of new physics.

The calorimeters were designed to fully contain the energy deposits of electrons, photons and hadrons through a cascade of collisions, each producing several lower energy particles until the shower multiplication dies. The shower containment distance is a function of the radiation length for electrons and photons and of the nuclear interaction length for hadrons. In sampling calorimeters, as used in all hadron collider experiments before the LHC, layers of absorber are interleaved with active detectors that measure the ionization signal which is proportional to the incident particle energy. The two length scales dictate different design choices for the electromagnetic (em) and hadronic (had) calorimeter sections. The energy resolution of the calorimeters is controlled by statistical fluctuations in the shower process and in the number of particles traversing the active layer, as well as by calibration variations and inhomogeneity due to, for example, cables and supports. The 4π calorimeter coverage allowed measurement of almost all produced particles so as to determine the total p_T of visible particles. Since the initial colliding hadrons have zero p_T , large ‘missing’ transverse energy (E_T^{miss}) can be ascribed to non-interacting particles such as neutrinos.

Muons, the only observable charged particles that can penetrate the calorimeter and magnet, are identified in large-area position detectors in the outer detector shell. The muon detectors were typically embedded in the iron return yokes of the magnet and relied on a precise momentum measurement in the tracker, but in the LHC ATLAS detector with its open air-core toroids outside the calorimeter a second precise momentum measurement can be made.

With the increased luminosity, the total collision rate grew from 3×10^5 Hz at the SpbarS, to 3×10^7 Hz at the Tevatron and now to 10^9 Hz at the LHC. Although the ability to record data for offline analysis has grown with modern computing technology, the number of bytes of data describing an event has also grown and it remained until recently impractical to record collisions at more than a few 100 Hz. Complex ‘trigger’ systems have evolved to select the few potentially interesting events from the flood of all collisions. Reliable triggering is key for the success of hadron collider experiments, as a rejected event is lost forever. The SpbarS triggers used fast indications of single particles like electrons and muons to signal an interesting event, although a smaller stream was selected by more elaborate analyses in microprocessors. At the Tevatron, three-level triggers were introduced in which the first deadtime-less decisions were made in single chip processors based on tracker, calorimeter or muon information, followed by microprocessor decisions combining several subdetectors, and finally a farm of processors running a simplified version of offline processing. The multiple level triggering gave data logging at 50 Hz with a dead time loss of about 5%. The LHC experiments retained the multi-level triggering with much more power at each stage to achieve fully efficient logging rates of a few 100 Hz. The growing luminosity also gives an increase in the number of events besides the one of interest occurring in the same bunch crossing. At the LHC, such ‘pileup’ gives dozens of extra events superimposed on the one of interest, and is a major challenge for physics analyses.

3 The LHC Project

The LHC project must be seen as a global scientific adventure, combining the accelerator complex, the experiment collaborations with their detectors, a worldwide computing grid, and a motivating theory community, that started more than 30 years ago. Obviously the main issue is the fundamental physics it addresses, but the project itself is truly ‘a marvel of technology’, a detailed account of which can be found in [17], and specific technological highlights are featured in [18], from which some of the technical discussion is reproduced here.

3.1 Historical Setting and Time Line

A comprehensive history of the years leading to the LHC can be found in [19]. The great success in making the experimentally “clean” W and Z boson discoveries, despite the huge hadronic backgrounds, at the CERN SpbarS Collider in the early 1980s was crucial for the community to dare

to even dream of a future powerful high-energy hadron collider in order to make a decisive search for the Higgs boson. The idea that the tunnel for the, at that time, still future Large Electron-Positron (LEP) machine should be able to house, at some distant time, the LHC, was already in the air in the late 1970s [20]. Thankfully, those leading CERN at the time had the vision to plan for a tunnel with dimensions that could accommodate it. Enthusiasm for an LHC surfaced strongly in 1984 at a CERN-ECFA workshop in Lausanne entitled “LHC in the LEP Tunnel”, which brought together working groups that comprised machine experts, theorists and experimentalists.

With the promise of great physics at the LHC, several motivating workshops and conferences followed, where the formidable experimental challenges started to appear manageable, provided that enough R&D work on detectors would be carried out. Highlights of these “LHC experiment preliminaries” were the 1987 Workshop in La Thuile of the so-called “Rubbia Long-Range Planning Committee” and the large Aachen ECFA LHC Workshop in 1990. Finally, in March 1992, the famous conference entitled “Towards the LHC Experimental Programme”, took place in Evian-les-Bains, where several proto-collaborations presented their designs in “Expressions of Interest”. This was also the time when CERN created an international peer review committee for the LHC experiments (LHCC). Moreover, from the early 1990s, CERN’s LHC Detector R&D Committee (DRDC), which reviewed and steered R&D collaborations, greatly stimulated innovative developments in detector technology.

It cannot be stressed enough how important the many years of R&D were that preceded the final detector construction for the LHC experiments. Technologies had to be taken far beyond their state-of-the-art of the late 1980s in terms of performance criteria in the anticipated harsh LHC environment, like granularity and speed of readout, radiation resistance, reliability, but also considering buildable sizes of the detector components and number of units, and very importantly at an affordable cost. For many detector subsystems there were initially a few parallel developments pursued as options, because it was not guaranteed from the onset that a given proposed technology would finally fulfil all the necessary requirements. Increasingly more realistic prototypes were developed, in a learning process for both the detector communities and the industries involved.

Some of the major technology decisions were taken by the Collaborations before the submission of the Technical Proposals to the LHCC end of 1994, which were finally approved early in 1996. For other choices the R&D needed more time, and they could only be made in the subsequent years from 1996 to the early 2000s, thereby defining the timing for the final Technical Design Reports of the various detector components.

The long duration of the LHC project until the Higgs discovery is illustrated in Table 1, with a few selected milestones concerning the LHC and the general-purpose experiments.

3.2 The Collider

In the LHC the two counter-rotating beams travel in separate channels in the arcs around the ring and consist of many particle bunches separated by a small distance. They are made to collide at the centre of the detectors with a small crossing angle. In the case of the LHC, the nominal number of bunches is 2808, the distance between bunches is 7.5 m and the crossing angle 285 μ rad. The layout of the collider is shown in Fig.2. There are eight arcs (bending radius of 2804 m), where the beams are bent, and eight straight-sections used for utilities or collision insertions. Four insertions are equipped with experimental detectors, where the two counter-rotating beams can be brought to collision.

Table 1: The LHC Timeline

1984	Workshop on a Large Hadron Collider in the LEP tunnel, Lausanne, Switzerland.
1987	Workshop on the Physics at Future Accelerators, La Thuile, Italy. The Rubbia “Long-Range Planning Committee” recommends the Large Hadron Collider as the right choice for CERN’s future.
1990	LHC Workshop, Aachen, Germany (discussion of physics, technologies and detector design concepts).
1992	General Meeting on LHC Physics and Detectors, Evian-les-Bains, France (with four general-purpose experiment designs presented).
1993	Three Letters of Intent evaluated by the CERN peer review committee LHCC. ATLAS and CMS selected to proceed to a detailed technical proposal.
1994	The LHC accelerator approved for construction, initially in two stages.
1996	ATLAS and CMS Technical Proposals approved.
1997	Formal approval for ATLAS and CMS to move to construction (materials cost ceiling of 475 MCHF).
1997	Construction commences (after approval of detailed Technical Design Reports of detector subsystems).
2000	Assembly of experiments commences, LEP accelerator is closed down to make way for the LHC.
2008	LHC experiments ready for pp collisions. LHC starts operation. An incident stops LHC operation.
2009	LHC restarts operation, pp collisions recorded by LHC detectors.
2010	LHC collides protons at high energy (centre of mass energy of 7 TeV).
2012	LHC operates at 8 TeV: discovery of a Higgs boson.

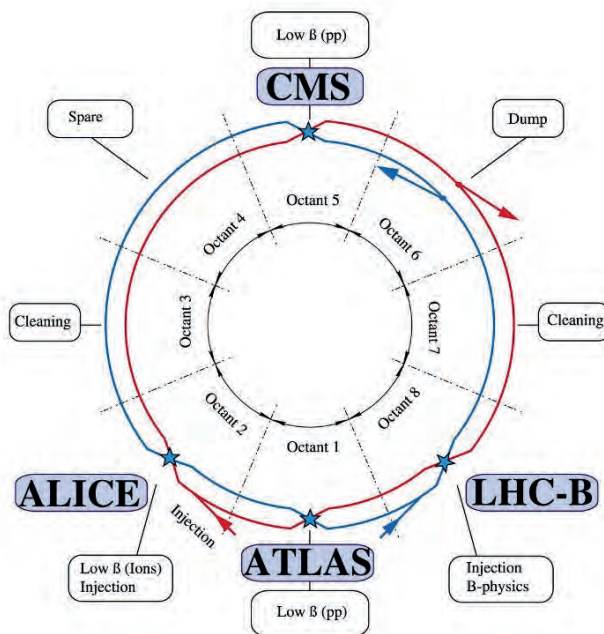


Fig. 2: Layout of the Large Hadron Collider

Just to recall, the two most important parameters of a proton-proton collider are the energy of the collisions and the luminosity, a parameter proportional to the number of collision events per second. The energy is related to the discovery potential of new particles of higher mass and the luminosity to the production of a relevant number of the desired events in a reasonable running time. The luminosity is a measure of the quality of the beams and their collisions, which is the result of careful design and mastering of several phenomena.

The *beam energy* is proportional to the product of B (the magnetic field of the main dipole magnets) and ρ (the bending radius of the arcs). Obviously, since ρ was fixed by the dimensions of the LEP tunnel, it was important to aim at the highest possible field B . Prior to the LHC, three large accelerators are/were based on superconducting magnets: the Tevatron (Fermilab, Chicago), HERA (DESY, Hamburg) and RHIC (BNL, Brookhaven). All of these make/made use of classical NbTi cables cooled by helium at a temperature slightly above 4.2 K (He-I). In each case the fields are below or around 5 Tesla. The choice for LHC was to use NbTi superconductors cooled at a lower temperature, namely 1.9 K provided by superfluid helium (He-II). The gain due to the lower temperature is about 1.5 Tesla or an additional 1.3 TeV in energy. It is interesting to note that the solution adopted for the LHC reconciled two requirements, namely the quest for the highest possible magnetic field and a design such that the magnets could be constructed in existing industry. This design with 8.3 Tesla magnetic field in the dipoles resulted in a nominal beam energy of 7 TeV. Out of the almost ten thousands magnets forming the collider, the 1232 main dipole magnets (length 16 m, mass 27 tons) represented a major design and constructional effort.

The solution of NbTi conductors at 1.9 K offers another important advantage at the expense of a more complex cryogenic system. It is due to the peculiar transport properties of pressurized superfluid helium, such as high heat capacity, low viscosity, and good effective thermal conductivity: it permits to keep the very large total helium mass static by cooling the long string of magnets with only a very small flow of liquid. The total cold mass of about 35000 tonnes operating at 1.9 K is the coldest spot in the Universe.

The considerable beam power (362 MJ) and the large electromagnetic energy stored in the magnets (11 GJ for the complete system) require a very sophisticated protection system to prevent

damage in case of beam guiding problems or resistive transition (quench). In fact each of the two beams have an energy sufficient to melt 500 kg of copper and the electromagnetic energy of the magnets, if not properly discharged, can provoke large damages.

Another important element is the beam vacuum. The requirements for the beam vacuum are imposed by the beam lifetime and the background to the experiments. The interactions between the protons and rest gas are driven by two processes, i.e. single proton-nuclear collisions in which a proton is lost, and multiple small-angle Coulomb collisions, which provoke an increase of the size of the beam and a decrease of the luminosity. To insure a beam lifetime of a few hours, the residual pressure in the vacuum chamber should not exceed $10^{-10} - 10^{-11}$ mbar. The vacuum chamber in the LHC consists basically two types: i) chambers made from stainless steel at 1.9 K in the arcs; ii) chambers made from copper at ambient temperature in the straight sections covering about 6 km of the circumference.

The *luminosity* L is related to the properties of the colliding beams at the collision point and it is measured in $\text{cm}^{-2} \text{s}^{-1}$. It is proportional to the square of number of particle per bunch (N), the number of bunches around the ring (n_b) and inversely proportional to the transverse dimensions of the beams at the collision point. A small correction factor F takes into account the small angle of the beams at crossing. The product of L with the cross-section σ of the process to be investigated gives the average number of events produced per second. L varies with time, since the stored beams degrade during the collision run. The integrated luminosity is the integral $L(t)dt$ over a certain time period. The luminosity can be considered as a figure of merit of the global quality of the machine. During the 2012 run, with beam energy of 4 TeV, the highest initial luminosity was $7.7 \cdot 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$. The integrated luminosity collected by ATLAS and CMS in 2012 was 21.5 fb^{-1} (delivered 23 fb^{-1}).

3.3 Overview and Motivation for the LHC Detectors

The enthusiasm for the great physics potential of an LHC arose in the community at the Lausanne ECFA-CERN workshop in 1984 already mentioned before. Finding the Higgs boson, the direct experimental manifestation of the Brout-Englert-Higgs mechanism, was clearly a priority for the future of particle physics, but was also expected to be very challenging. Its mass is not predicted by the Standard Model and could have been as high as 1 TeV. This required a search over a broad range of mass, hence ideally suited at a high-energy pp collider where the energy spectrum of the constituents of the protons (quarks and gluons) allow all possible Higgs masses to be looked for at the same time. Because of the predicted low detectable cross-sections a very high-luminosity collider was mandatory.

But the Higgs search was by far not the only reason to stimulate great interest for the LHC. Already in the 1980s there were clear motivations from speculative theories predicting physics Beyond the Standard Model (BSM), the most popular one being Supersymmetry (SUSY) with its characteristic missing transverse energy signatures due to the escaping lightest neutral SUSY particle (LSP). Other hypothetical models predicted new heavy resonances, or leptoquarks (particles containing both quarks and leptons), or even substructure to quarks, and many other exotic ideas were around. It was also realized early on that the LHC would produce huge numbers of heavy flavour particles, opening a new frontier in precision flavour physics. Furthermore, unprecedented exploratory steps could be made in studying the quark-gluon plasma at high density and temperatures when colliding heavy ions, like fully ionized lead nuclei. Motivated by these physics prospects, CERN opted ultimately for an experimental programme with two very large general-purpose detectors (ATLAS and CMS), two large apparatus optimized respectively for flavour physics (LHCb) and for heavy ion collisions (ALICE), complemented later by three much smaller specialized experiments (TOTEM, total cross-section and forward physics; LHCf, measuring forward neutral particle production; MoEDAL, monopole search).

The detection of the Higgs boson played a particularly important role in the design of the *general-purpose experiments ATLAS and CMS*. In the region of low mass ($114 < m_H < 150$ GeV), the two channels considered mostly suited for unambiguous discovery were the decay to two photons and the decay to two Z bosons, each decaying in turn into e^+e^- or $\mu^+\mu^-$, where one or both of the Z bosons could be virtual. As the natural width of the low-mass Higgs boson is < 10 MeV, the width of any observed peak would be entirely dominated by the instrumental mass-resolution. This meant that in designing the general-purpose detectors, considerable care was placed on the value of the magnetic field strength, on the precision tracking systems and on high-resolution em calorimeters. The high-mass region, as well as the signatures from supersymmetry, drove the need for good resolution for jets and missing transverse energy (E_T^{miss}), implying also almost full 4π hadronic calorimetry coverage.

The choice of the field configuration determined the overall design for these experiments. It was also well understood that to stand the best chance of making discoveries at the new energy scale of the LHC - and in the harsh conditions generated by about a billion pairs of protons interacting every second, several tens every bunch crossing - would require the invention of new technologies while at the same time pushing existing ones to their limits. In fact, a prevalent saying was “We think we know how to build a high-energy, high-luminosity hadron collider – but we don’t have the technology to build a detector for it”. In reality of course both turned out to be difficult and demanded technological breakthroughs. Early on it was realized that detectors will have to face eventually even luminosities beyond the initial LHC design to reach some of the ultimate physics goals. That the general-purpose experiments have worked so marvellously well since the start-up of the LHC is a testament to the difficult technology-choices made by the conceivers and the critical decisions made during the construction of these experiments. It is noteworthy that indeed the very same elements mentioned above were crucial in the discovery of a Higgs boson.

Very different challenges were faced for the two *special-purpose experiments LHCb and ALICE*, which are reflected in their very different specific designs. The only aspects they have in common is their operation at lower luminosity, typically at $2 \cdot 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ or lower, avoiding basically multiple events per bunch crossing, and their specific optimizations for particle identifications, as will be discussed later.

It is far beyond the scope of this lecture note to give a comprehensive description of these four sophisticated instruments that have been developed with very considerable R&D efforts, culminating in many large-scale prototype measurement campaigns in particle beams at CERN and other accelerator laboratories, over the 1990s and early 2000s. The construction of the various components took place over about 10 years, starting in the second half of the 1990s, in universities, national laboratories, and industries. Typically, after local initial testing, the components were sent to CERN, where after reception tests they were assembled and installed in the experimental caverns, followed by commissioning tests. During all this time, from the first Letters of Intent in 1992 to the operation phase, CERN’s LHCC played an important role closely guiding and monitoring the experiments. Figure 3 shows photographs of the four detectors during their late installation phase, before the detectors were completely closed for operation.

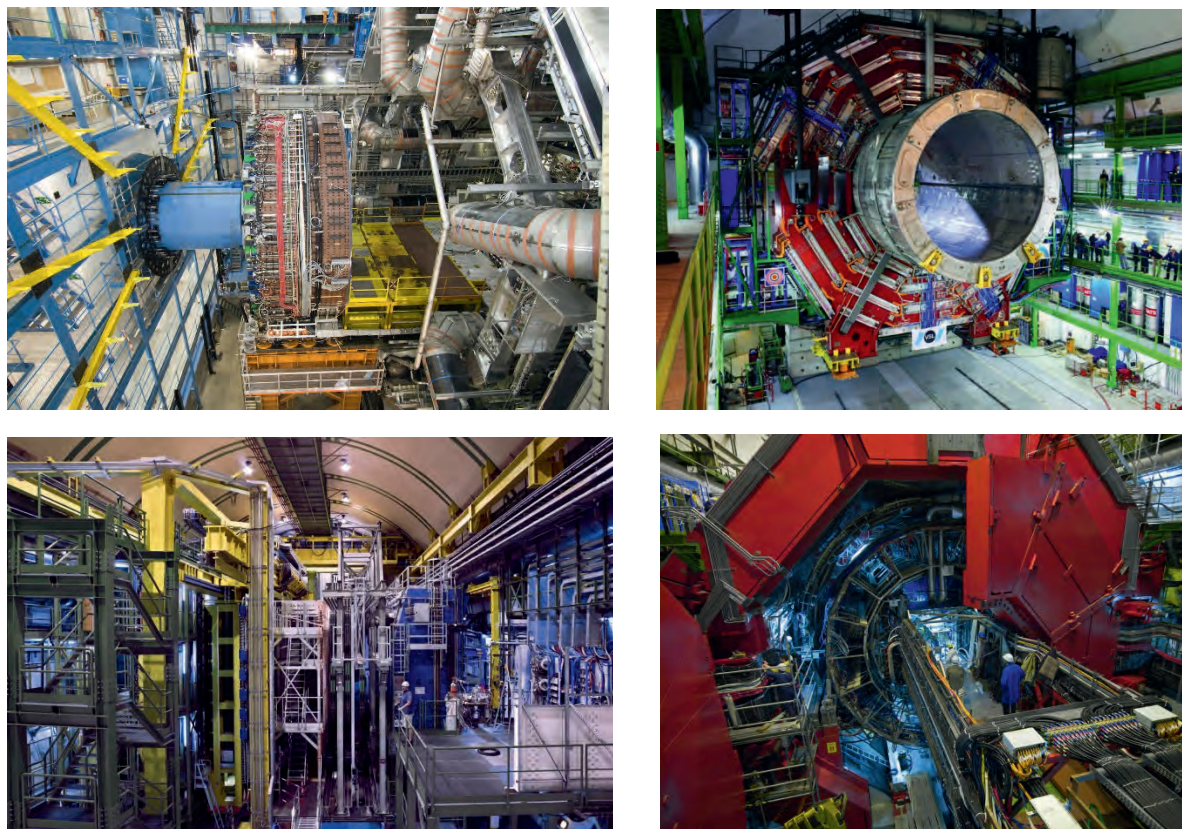


Fig. 3: The four LHC detectors during the installation phase. Upper row left (a) ATLAS, right (b) CMS, lower row left (c) LHCb, and right (d) ALICE.

In the following the four detectors are very briefly introduced separately. The four collaborations have published very detailed and comprehensive technical descriptions of their detectors as finally built and operated during the first years of LHC in [21-24]. The interested reader is invited to consult these major documentations for detailed information.

3.3.1 *ATLAS* [21]

The design of the ATLAS detector is based on a large toroid magnet system for the muon spectrometer complemented with a small superconducting solenoid around the inner tracking cylinder, centred at the interaction point. The novel and challenging superconducting air-core toroid magnet system, contains about 80 km of superconductor cable in eight separate barrel coils (each $25 \times 5 \text{ m}^2$ in a ‘racetrack’ shape) and two matching end-cap toroids. A field of ~ 0.5 Tesla is generated over a large volume. The toroids are complemented with a thin solenoid (2.4 m diameter, 5.3 m length), which provides an axial magnetic field of 2 Tesla. The momentum of the muons can be precisely measured as they travel unperturbed by material for more than 5 m in the air-core toroid field. About 1200 large muon drift tube chambers of various shapes, with a total area of 5000 m^2 , measure the impact positions with an accuracy of better than 0.1 mm. Another set of about 4200 fast chambers are used to provide a muon trigger (resistive plate chambers in the barrel, thin gap chambers in the end-caps).

In the field-free region between solenoid and toroids there is a highly granular em calorimeter using a novel Lead - liquid Argon (LAr) sampling calorimeter complemented by full-coverage hadronic sampling calorimeters. For the latter a plastic scintillator – iron sampling technique, also with a novel geometry, is used in the barrel part of the experiment, whereas LAr calorimeters cover the end-cap regions near the beam axis where particle fluxes, and thereby radiation exposures, are highest. The em and hadronic calorimeters have 200000 and 10000 cells respectively.

The reconstruction of all charged particles, including those of displaced secondary vertices, is done in the inner tracking detector, which combines highly granular pixel ($50\ \mu\text{m} \times 400\ \mu\text{m}$, total 80 million channels) and microstrip ($13\ \text{cm} \times 80\ \mu\text{m}$, total 6 million channels) silicon semiconductor sensors close to the beam axis, and a ‘straw tube’ gaseous detector (4 mm diameter, 350000 channels) that provides about 35 signal hits per track. The latter also helps in the identification of electrons using information from the effect of transition radiation.

Figure 3a shows one end of the cylindrical barrel detector after about 4 years of in-situ installation work in the underground cavern, 1.5 years before completion. The ends of four of the barrel toroid coils are visible, illustrating the eightfold symmetry of the structure. The relatively lightweight overall structure of the detector adds up to 7000 tonnes, spanning over a large volume of 22 m diameter with a length of 46 m.

3.3.2 *CMS [22]*

The CMS detector features prominently a state-of-the-art superconducting high-field solenoid of 4 Tesla. This single magnet solution serves both the high precision inner tracking chambers as well as the external muon detector, which is instrumented with large gaseous drift chambers in the barrel and cathode strip chambers in the end-caps complemented by resistive plate trigger chambers, embedded in the return yoke. This configuration allowed one to achieve a compact overall design, limiting the diameter to 15 m. The magnet yoke makes up for most of the total detector weight of 12500 tonnes.

Muon detection, and their most accurate measurement, was a priority criterion for the CMS design, followed by precision measurements for photons and electrons, achieved with a new type of radiation hard em crystal calorimeter, the largest ever built. The challenging, but very successful development and manufacture of the 75848 lead tungstate scintillating crystals, in the final set-up pointing to the interaction point, took more than a decade.

The solution to charged particle tracking was to opt for a small number of precise position measurements for each charged particle trajectory (13 layers with a position resolution of $\sim 15\ \mu\text{m}$ per measurement) leading to a large number of cells distributed inside a cylindrical volume 5.8 m long and 2.5 m in diameter: 66 million $100 \times 150\ \mu\text{m}^2$ silicon pixels and 9.3 million silicon microstrips ranging from about $10\ \text{cm} \times 80\ \mu\text{m}$ to $20\ \text{cm} \times 180\ \mu\text{m}$. With $198\ \text{m}^2$ of silicon detector area the CMS tracker is by far the largest silicon tracker ever built.

Finally the hadron calorimeter, comprising ~ 3000 small solid angle projective towers covering almost the full solid angle, is built from alternate plates of $\sim 5\ \text{cm}$ brass absorber and $\sim 4\ \text{mm}$ thick scintillator plates that sample the shower energy. The scintillation light is detected by photodetectors (hybrid photodiodes) that can operate in the strong magnetic field, as the calorimeters are placed inside the solenoid coil.

The iron yoke of the CMS detector is sectioned into five barrel wheels and three endcap disks at each end. The sectioning enabled the detector to be assembled and tested in a large surface hall while the underground cavern was being prepared. The sections, weighing between 350 and 2000 tonnes, were then lowered into the cavern (Fig. 3b) between October 2006 and January 2008, using a dedicated gantry crane system equipped with strand jacks: a pioneering use of this technology to simplify the underground assembly of large experiments.

3.3.3 *LHCb [23]*

The LHCb detector concept exploits the dominant rate of beauty production towards the beam directions, and for practical reasons concentrates just on one of the two sides. Away from the LHC collision region the layout therefore resembles a fixed target spectrometer, but with very unique features.

A silicon strip vertex locator (VELO) detector can be positioned during stable beams very close to the interaction region and beam line in order to measure accurately primary and secondary vertices, selecting events with b-quarks. Particle identification to cleanly identify the various final states is achieved by two ring image Cerenkov detectors (RICH), whereas the momentum measurements are based on a large-aperture warm dipole magnet generating an integrated field path of 4 Tesla-meters for trajectories going through all tracking stations (silicon strips and straw tube drift chambers) of the spectrometer. Calorimetry is provided for by sampling scintillator lead (em) and coarser scintillator iron devices. The muon detector behind the calorimeters, an absorber with iron plates of a total of 20 interaction lengths, sampled by four chamber layers, completes the LHCb spectrometer. Figure 3c shows a picture of the fully installed LHCb detector in LHC Point-8.

3.3.4 ALICE [24]

The ALICE detector has to cope with extremely high multiplicity events, characteristic of heavy ion collisions, including charged particle measurements at an as low as possible momentum threshold. Furthermore, particle identification is needed for many of the specific heavy ion physics signatures.

The ALICE Collaboration has reused the former huge L3 warm solenoid magnet providing a field of 0.5 Tesla over a large central volume. Within the magnet, with its 10000 tonnes heavy iron yoke, is located, as central tracking detector, the world's largest Time Projection Chamber (TPC) with a field cage of 5.6 m diameter and 5.4 m length, which provides precision tracking as well as particle identification by dE/dx ionization information. The innermost region, inside the TPC, facing the most extreme particle density region around the collision point, is covered by an optimized inner tracking system with silicon pixel and silicon drift detectors followed by double-sided silicon strip detectors. Several detector systems dedicated to particle identification over various limited solid angles cover the outside radius of the TPC: transition radiation and Cherenkov radiation detectors, and a state-of-the-art time-of-flight system (TOF). High resolution em calorimetry for photon measurements is implemented with lead tungstate scintillating crystals (similar characteristics as for CMS).

A muon spectrometer starting with a massive 4 m long sophisticated hadron absorber cone, and featuring a classical dipole magnet, covers on one side the solid angle from 2 to 9 degrees with respect to the beam direction. A front view of the ALICE detector nearing installation completion is shown in Fig. 3d.

3.4 Triggering and Computing

A particular challenge for ATLAS and CMS are the very high collision rates in the LHC, necessary for the Higgs search and studies, given the small production cross section combined with the need to investigate final states with very small branching fractions. In the first three years of operation the LHC reached a peak instantaneous luminosity of $7 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ with a 50 ns bunch spacing, which meant that the detectors had to simultaneously cope with up to ~ 50 overlapping (pile-up) events in a given bunch crossing. In the years ahead, the instantaneous luminosity is still expected to rise two- to three-fold.

It is technically not possible to store all data for all events, therefore a trigger system is used to reject large numbers of events and retain only the interesting ones from crossings with potential physics processes of interest. This is done in real time by sophisticated integrated trigger and data acquisition systems, involving custom-made fast electronics in a first stage and large computing farms in subsequent stages before the data is transferred to mass storage for further analyses. The initial data rate from up to 40 MHz bunch crossings with multiple pile-up events is thereby reduced to a few hundreds of Hz for offline analysis. A description of these systems is far beyond the scope of these lectures, see [21,22] for details, as well as [23,24] for the specific data selection and data flow challenges for LHCb and ALICE.

The LHC experiments generate huge amounts of data (tens of petabytes of data per year) requiring a fully distributed computing model. The worldwide LHC Computing Grid (wLCG) was developed to deal with this task [17,25]. With its hierarchical architecture of tiered centres it allows any user anywhere access to any data recorded or produced in the analyses steps during the lifetime of the experiments. The centre at CERN receives the raw data, carries out prompt reconstruction, almost in real time, and exports the raw and reconstructed data to the Tier-1 centres and also to Tier-2 centres for physics analysis. The Tier-0 must keep pace with the event rate of several hundred Hz of typically 1 MB of raw data per event from each experiment. The large Tier-1 centres provide also long-term storage of raw data and reconstructed data outside of CERN (as a second copy). They carry out, for example, second-pass reconstruction, when better calibration constants are available. The large number of events simulated by Monte Carlo methods and necessary for quantifying the expectations are produced mainly in Tier-2 centres.

3.5 Comment on Testing and Commissioning

The Individual detector components (e.g. chambers) were built and assembled in a distributed way all around the globe in the numerous participating institutes and were typically first tested at their production sites, then after delivered to CERN, and finally again after their installation in the underground caverns. The collaborations also invested enormous efforts in testing representative samples of the detectors in test beams at CERN and other accelerator laboratories around the world. These test beam campaigns not only verified that performance criteria were met over the several years of production of detector components, but were also used to prepare the calibration and alignment data for LHC operation. Very important were the so-called large combined test beam set-ups, which represented whole ‘slices’ of the different detector layers of the final detectors.

During the progressing installation the experiments made extensive use of the constant flow of cosmic rays impinging on Earth providing a reasonable flux of muons even at a depth of 100 m underground, typically a few hundred per second traversing the detectors. These muons were used to check the whole chain from hardware to analysis programs of the experiments, and to align the detector elements and calibrate their response prior to the pp collisions. In particular, after the LHC incident on 19th September 2008 the experiments used the 15 months LHC down time, before the first collisions on 23rd November 2009, to run the full detectors in very extensive cosmic ray campaigns, collecting many hundreds of millions of muon events. These runs allowed the experiments to be ready for physics operation, with already accurately pre-calibrated and pre-aligned detectors, by the time of the first pp collisions.

An excellent account of this huge and essential commissioning work has been given in previous lectures to this school series by A. Hoecker, and the reader is highly recommended to consult [26].

4 The Discovery of the Higgs Boson

4.1 Standard Model Measurements to Demonstrate the Performance

Observing, and measuring accurately, at the LHC collision energies, the production of known particles of the SM, was always considered to be a necessary stepping stone towards exploring the full potential of the LHC with its promise of new physics, firstly of the discovery of the Higgs boson. The SM processes, such as W and Z production, are often referred to as ‘standard candles’ for the experiments. However, there is much more value to measuring SM processes than this: never before could the SM physics be studied at a hadron collider with such sophisticated and highly accurate detectors, ultimately allowing tests of detailed predictions of the SM with unprecedented precision and minimal instrumental systematic errors.

An example of a very early measurement is shown in Fig. 4, produced only after a month or so after first high-energy collisions in spring 2010. ATLAS and CMS observed in such di-muon invariant mass distributions a ‘summary’ of decades of particle physics, with remarkable mass resolution.

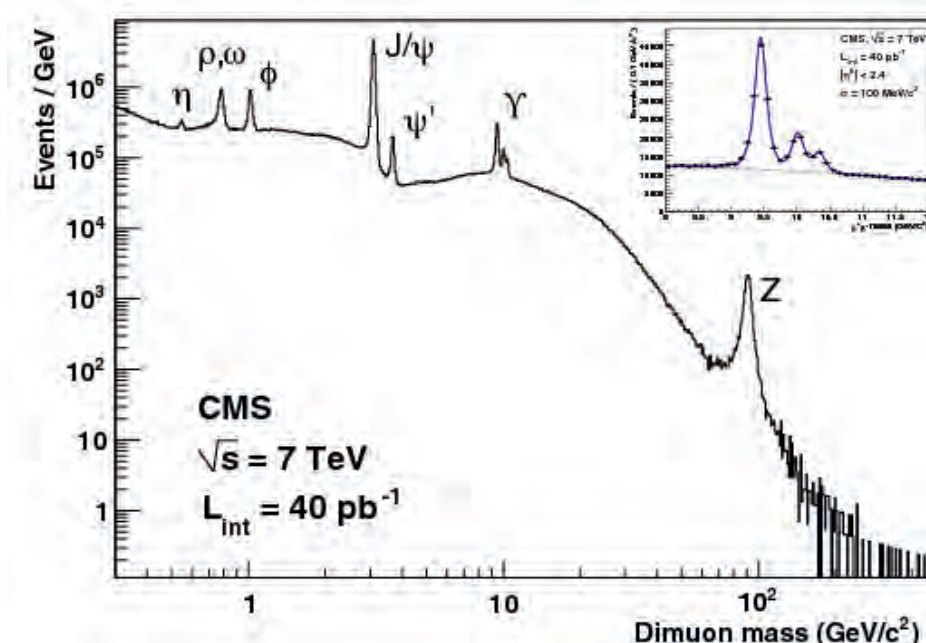


Fig. 4: The distribution of the invariant mass for di-muon events, shown here from CMS, displays the various well-known resonant states of the SM. The inset illustrates the excellent mass resolution for the three states of the Y family.

In the following paragraphs a few examples will be shown from a very extensive body of publications and publically released conference contributions, which are all available at [15] for ATLAS and [16] for CMS, where detailed specific references can be found.

The charged and neutral Intermediate Vector Bosons (IVB) W and Z are the major benchmark measurements at the LHC for demonstrating the excellent detector performance, as well as for testing model predictions to a high degree of accuracy. The Z decays into electron and muon pairs can be extracted almost free of any backgrounds, as shown in Fig. 5.

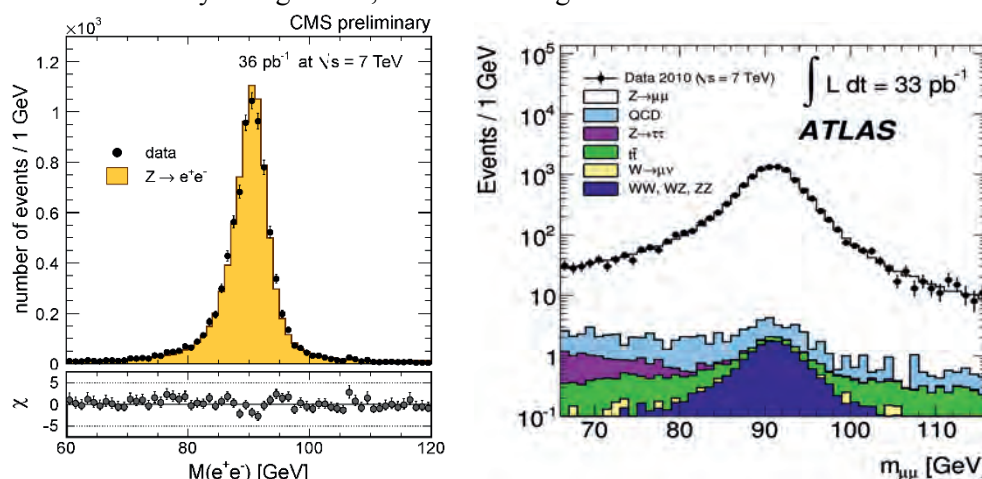


Fig. 5: The CMS electron-pair mass distribution on a linear (left) and the ATLAS muon-pair mass distribution on a logarithmic (right) vertical scale, in the Z mass region. The estimated small background contributions are indicated, as well as the expected signal shape from MC simulations.

The classical W decay signatures into an electron or muon and the associated neutrino are an excellent test for the E_T^{miss} performance of the detector due to the undetected neutrino. E_T^{miss} is inferred from the measured energy imbalance in the transverse projection of all observed signals w.r.t. the beam axis. The E_T^{miss} spectrum for events with a well-identified muon candidate is shown in Fig. 6a, and shows a clear W signal over the expected SM background sources. After applying a selection of events with $E_T^{\text{miss}} > 25$ GeV only a small residual background remains present under the W signal, as indicated in the transverse mass distribution given in Fig. 6b.

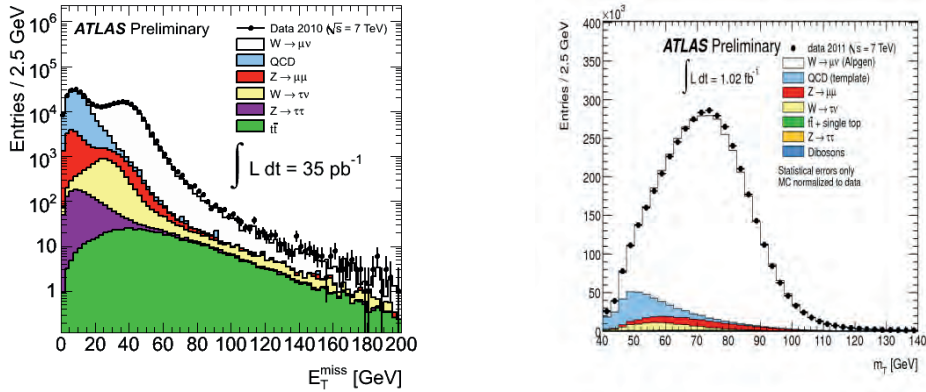


Fig. 6: (a, left) Missing transverse energy distribution for events with a muon candidate. (b, right) Transverse mass distribution for W to muon decays. The expected background contributions are indicated as well (examples from ATLAS).

The good agreement between the measured and expected cross-sections times the leptonic decay branching ratios (which is the expected rate for W and Z bosons to be produced and then decay to leptons) is illustrated in Fig. 7. With the presently available data samples the measurements are expected to already strongly constrain the theoretical model parameters. Figure 7a shows the cross-section measurements and predictions as a function of the collision energy, whereas in (b) the W and Z cross-section results are displayed in a 2-dimensional plot including their correlated error ellipse, and compared to predictions with various parton distribution functions (describing the quark and gluon momentum distributions inside the protons). Detailed measurements of properties for IVB production and decay at the LHC have been published already and are being refined now with the full Run-1 data samples. They include, for example, the lepton charge asymmetry measurements for W decays, which were an important signature of the electro-weak nature of the W at the time of their discovery some 30 years ago.

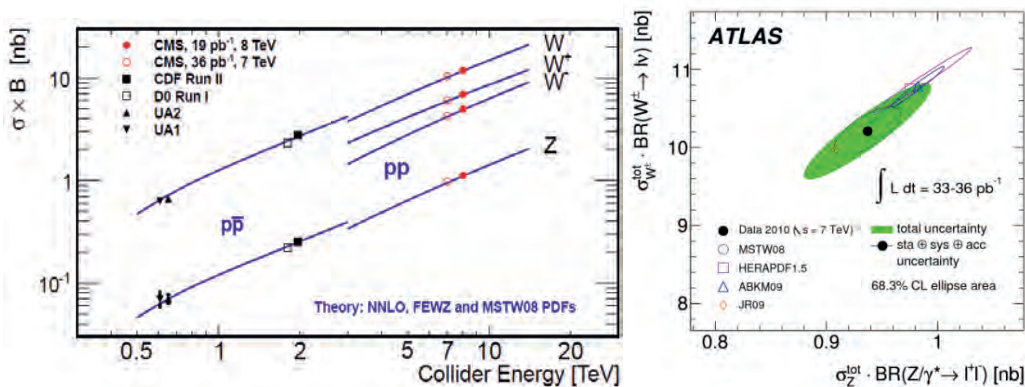


Fig. 7: (a, left) CMS W and Z production cross-sections times leptonic branching ratio as a function of the collision energy, showing also previous measurements at lower energy colliders. (b, right) Correlation of the measured (solid dot) leptonic W and Z cross-section from ATLAS, compared to theoretical expectations with various choices for the parton distribution functions (open symbols).

Hard collisions (characterized by having final state particles with significant transverse energy) at the LHC are dominated by the production of high transverse momentum jets, which are the

collimated sprays of particles from the hadronization of the initially scattered partons (quarks, gluons) in the colliding protons. At work is the strong interaction described by Quantum Chromo Dynamics (QCD). Most commonly two jets emerge at opposite azimuth with balanced transverse momenta, from an initially lowest order parton-parton scattering process. However, higher order QCD corrections alter this picture significantly, and detailed measurements of multi-jet configurations are very important to constrain the QCD descriptions of hadronic processes.

The most impressive results at this stage are the inclusive jet and the di-jet cross-section measurements; an example from ATLAS for them is shown in Fig. 8a. These measurements cover unprecedented kinematical ranges spanning typically over jet transverse momenta from 20 GeV to 2 TeV, in many angular (pseudorapidity) bins up to $|\eta| < 4.4$ (i.e. very close to the beam axis). The cross-sections vary over these ranges by up to 12 orders of magnitude. In general the agreement with perturbative QCD calculations, including next to leading order (NLO) corrections, is well within the systematic uncertainties. This cannot be seen in Fig. 8a directly, only in ratio plots measurement/theory for a given η -interval as shown for CMS data in Fig. 8b. The systematic uncertainties in the ratios are typically less than 20%, which is a great achievement compared to previous such measurements. The systematic uncertainties on the measurements are dominated by the jet energy scale uncertainty (calibration of the detectors for the energy of jets), which, thanks to a considerable effort, has been determined to typically better than 3%.

Jets can also be produced together with W and Z bosons, so-called QCD corrections to the Intermediate Vector Boson production. Many results of these processes have been published. A good understanding of them is particularly important as they are, in many cases, a dominant source of backgrounds to the search for new particles, as well as to the measurements of top quark production discussed next.

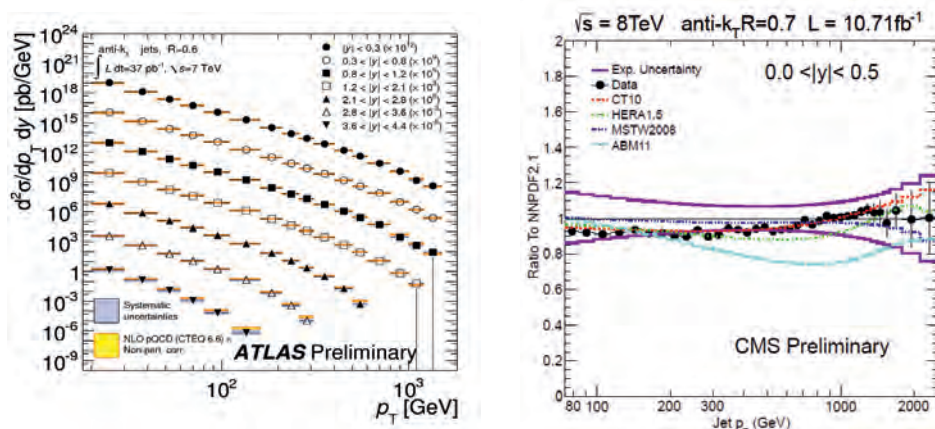


Fig. 8: ATLAS measurements of inclusive jet (a, left) cross-sections, and (b, right) CMS jet measurements compared to NLO perturbative QCD predictions plotted as ratio data/calculation.

The heaviest known particle in the SM is the top quark with its roughly 175 GeV mass. It decays almost exclusively into a W and a bottom quark. The measurement of top quark pair production typically requests that at least one of the W decays leptonically (also needed to trigger the events), and therefore the final states require one or two leptons (electrons or muons), E_T^{miss} , and jets, some of which, coming from the b-quarks, can be tagged by the displaced secondary vertices due to the finite life times of b-hadrons. Whilst it is beyond the scope of these notes to describe the sophisticated analyses employed, the message is that there are clear top pair signals in ATLAS and CMS, both in the single and two-lepton channels, when considering the correct jet topologies. The resulting cross-sections are shown in Fig. 9 which also illustrates the expected large rise of the cross-section with the collision energy increase from 2 TeV at the Tevatron to 7 TeV and 8 TeV at the LHC. Good agreement with NLO QCD calculations is seen within the present few % measurement errors. It can be mentioned that both ATLAS and CMS have also reported first single top measurements (events

with just one top quark) at a rate in good agreement with QCD expectations, as well as detailed studies of top properties like its mass.

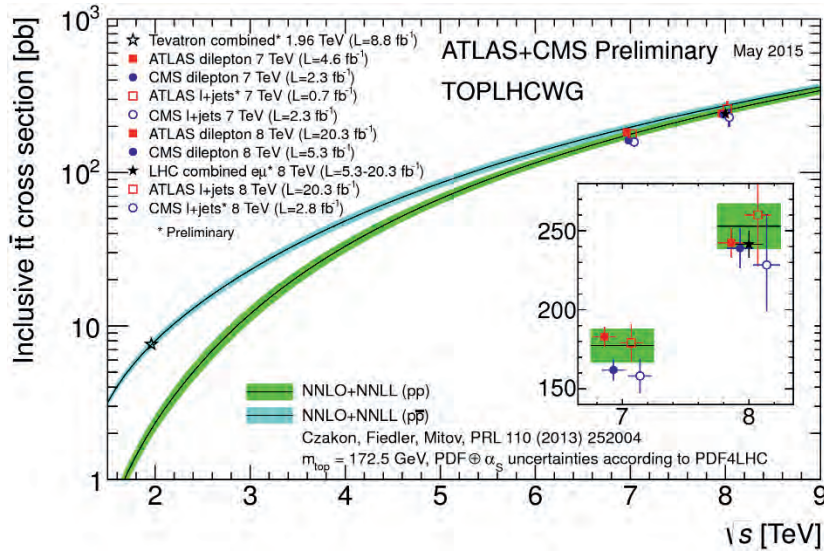


Fig. 9: Top pair production cross-section as a function of the collision energy, showing the Tevatron and LHC measurements.

In summary one can note that the data collected in the first three years of high-energy LHC operation have allowed ATLAS and CMS to make numerous precise measurements of SM processes, including production of bottom and top quarks, W and Z bosons, singly and in pairs. In particular very detailed measurements of QCD processes have been made. A collection of examples of such studies is shown in Fig. 10 where measurements of cross-sections for various selected electroweak and QCD processes are compared with the SM predictions. These very diverse measurements, probing cross-sections over a range of many orders of magnitude, confirm the predictions of the SM within the errors in all cases. Establishing this agreement is essential before any claims for discoveries can be made, i.e. to demonstrate on the one hand that the detector performance is well understood, and on the other hand that known SM processes are correctly observed in the experiments as they often constitute large backgrounds to signatures of new physics, such as those expected for the Higgs boson. The speed with which the wide range of measurements have shown that SM predictions for known physics have been essentially spot-on is a tribute to a large amount of work done by many particle physics theorists along with the results from the other collider experiments at LEP, Tevatron, HERA, and b-factories.

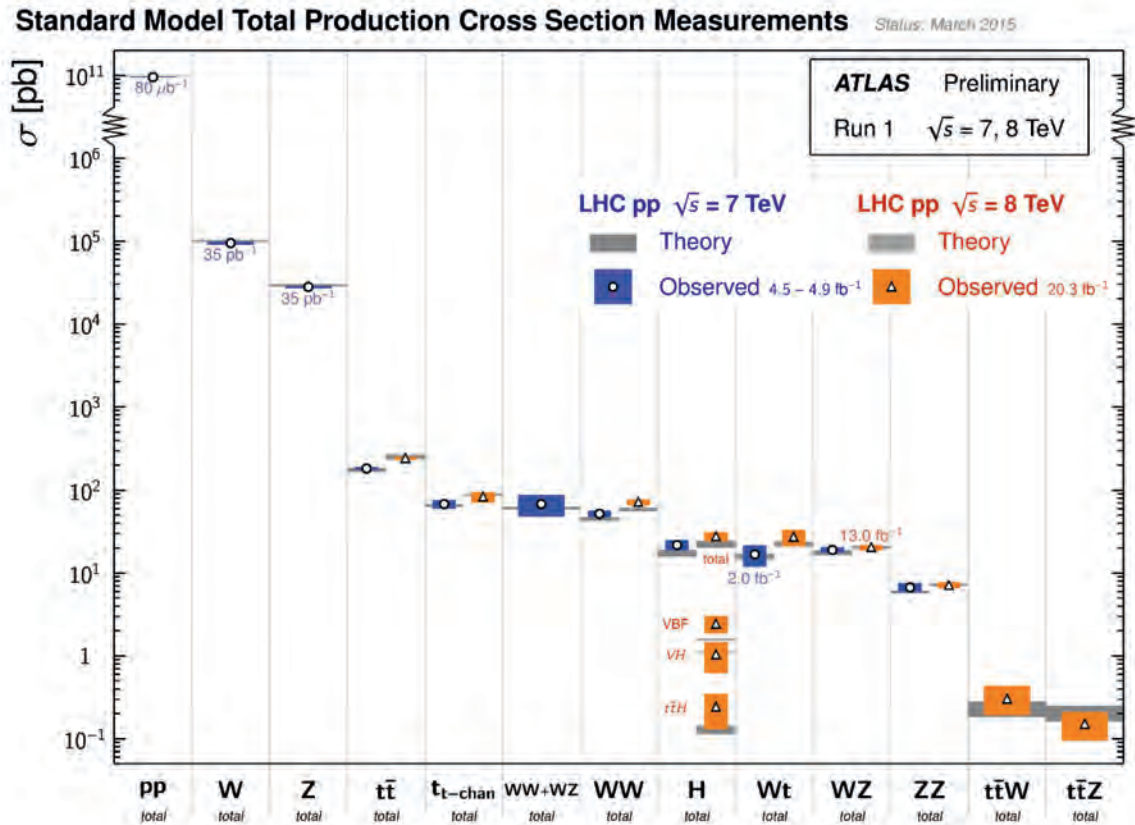


Fig. 10: A comparison of cross-section measurements for electroweak and QCD processes with theoretical predictions from the SM, shown here as example from the ATLAS experiment.

4.2 Discovery and Measurements of the Higgs boson

The discovery of a heavy scalar boson was announced jointly by the ATLAS and CMS Collaborations [1,2] on 4th July 2012 with a partial data sample corresponding to about 10 fb⁻¹ coming to about equal parts from running at 7 TeV collision energy in 2011 and 8 TeV in 2012 until June. The fantastic performance of the LHC during the second half of 2012 allowed the experiments to more than double their data sets. By the end of 2012 (LHC Run-1) the total amount of data that had been examined corresponded to ~5 fb⁻¹ at $\sqrt{s} = 7$ TeV and ~20 fb⁻¹ at $\sqrt{s} = 8$ TeV, equating to the examination of some 2000 trillion proton-proton collisions. Using these data first measurements of the properties of the new boson were also made. The accumulated luminosity evolution over Run-1 is illustrated in Fig. 11 for ATLAS, showing also that the experiment was very efficient in recording stably delivered luminosity as well as maintaining a high fraction (~90%) of data quality ‘good for physics’.

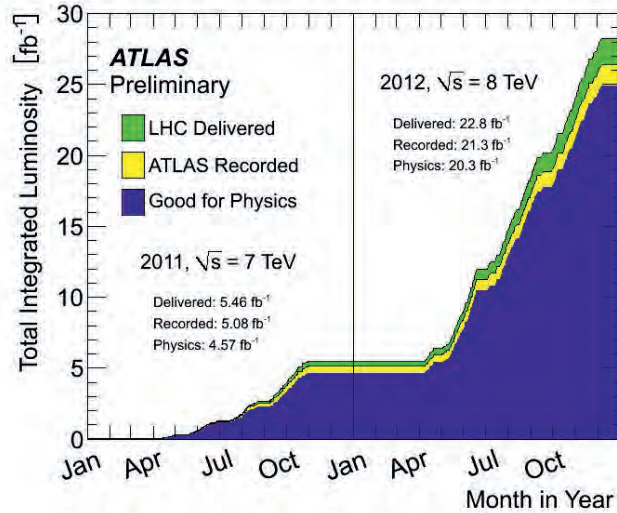


Fig. 11: Integrated luminosity over the high-energy running periods of Run-1 in 2011 and 2012, showing the stably delivered, recorded and finally used data sets for physics (shown is the example for ATLAS, CMS is very similar).

4.2.1 Decays to bosons: the $H \rightarrow \gamma\gamma$, the $H \rightarrow ZZ \rightarrow 4l$ and $H \rightarrow WW \rightarrow 2l2\nu$ decay modes

As examples for the full data sets, the results from the ATLAS experiment are shown for the $H \rightarrow \gamma\gamma$ decay mode (Fig. 12a) and those from the CMS experiment for the $H \rightarrow ZZ \rightarrow 4l$ mode (Fig. 12b). The signal is unmistakable and the significances are summarized in Table 2. The data show a clear excess of events above the expected background around 125 GeV. As for all the Higgs analysis figures shown in the following, the complementary data plots and results from the two experiments can be found in the detailed list of publications available from ATLAS [15] and CMS [16].

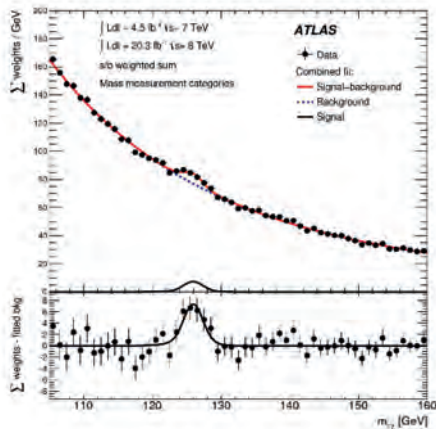


Fig. 12a: Invariant mass distribution of di-photon candidates. The result of a fit to the background described by a polynomial and the sum of signal components is superimposed. The bottom inset displays the residuals of the data with respect to the fitted background component.

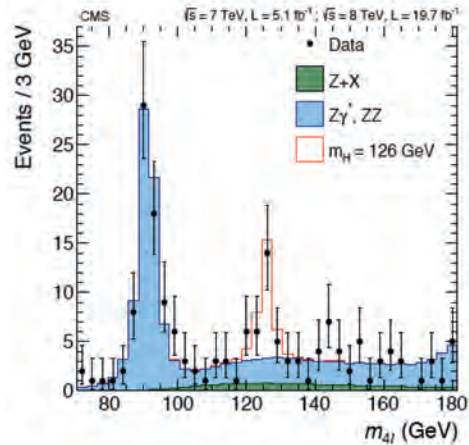


Fig. 12b: The four-lepton invariant mass distribution in the CMS experiment for selected candidates relative to the background expectation. The expected signal contribution is also shown.

The search for $H \rightarrow WW$ is primarily based on the study of the final state in which both W bosons decay leptonically, resulting in a signature with two isolated, oppositely charged, high p_T leptons (electrons or muons) and large E_T^{miss} due to the undetected neutrinos. The signal sensitivity is improved by separating events according to lepton flavour; into e^+e^- , $\mu^+\mu^-$, and $e\mu$ samples, and

according to jet multiplicity into 0-jet and 1-jet samples. The dominant background arises from irreducible non-resonant WW production. Any background arising from Z bosons, with same flavour but opposite sign leptons, is removed by a di-lepton mass cut $(m_Z - 15) < m_{ll} < (m_Z + 15)$ GeV.

The m_{ll} distribution in the 0-jet and $e\mu$ final state is shown for CMS in Fig. 13a. The expected contribution from a SM Higgs boson with $m_H = 125$ GeV is also indicated. The transverse mass, m_T , distribution is shown in Fig. 13b from ATLAS, as well as the background-subtracted distribution. Both show a clear excess of events compatible with a Higgs boson with mass ~ 125 GeV. The observed (expected) significances of the excess with respect to the background-only hypothesis are shown in Table 2.

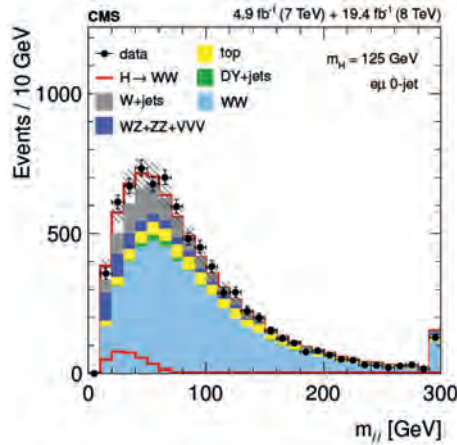


Fig. 13a: Distribution of dilepton mass in the 0-jet, $e\mu$ final state in CMS for a $m_H = 125$ GeV SM Higgs boson decaying via $H \rightarrow WW \rightarrow l\nu l\nu$ and for the main backgrounds

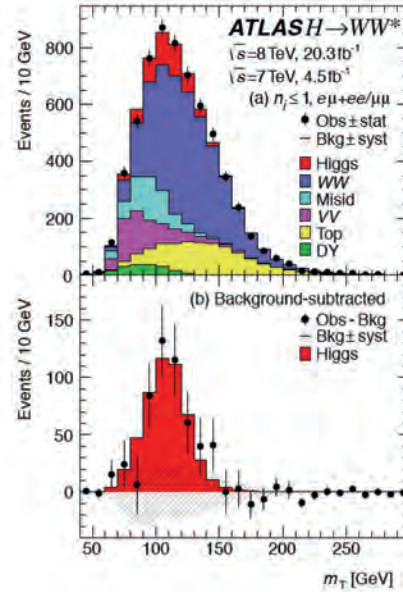


Fig. 13b: The transverse mass distributions for events passing the full selection of the $H \rightarrow WW \rightarrow l\nu l\nu$ analysis in ATLAS summed over all lepton flavours for final states with $N_{jet} \leq 1$. In the lower part the residuals of the data with respect to the estimated background are shown, compared to the expected m_T distribution for a SM Higgs boson.

4.2.2 Decays to fermions: the $H \rightarrow \tau\tau$ and the $H \rightarrow bb$ decay modes

It is important to establish whether this new particle also couples to fermions, and in particular to down-type fermions, since the measurements above mainly constrain the couplings to the up-type top quark. Determination of the couplings to down-type fermions requires direct measurement of the Higgs boson decays to bottom quarks and τ leptons.

The $H \rightarrow \tau\tau$ search is typically performed using the final-state signatures $e\mu$, $\mu\mu$, $e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$, where electrons and muons arise from leptonic τ -decays and τ_h denotes a τ lepton decaying hadronically. Each of these categories is further divided into two exclusive sub-categories based on the number and the type of the jets in the event: (i) events with one forward and one backward jet, consistent with the Vector-Boson-Fusion (VBF) topology, (ii) events with at least one high p_T hadronic jet but not selected in the previous category. In each of these categories, a search is made for an excess in the reconstructed $\tau\tau$ mass distribution. The main irreducible background, $Z \rightarrow \tau\tau$

production, and the largest reducible backgrounds ($W + \text{jets}$, multijet production, $Z \rightarrow ee$) are evaluated from various control samples in data.

The $H \rightarrow b\bar{b}$ decay mode has by far the largest branching ratio ($\sim 54\%$). However since $\sigma_{b\bar{b}}$ (QCD) $\sim 10^7 \times \sigma(H \rightarrow b\bar{b})$ the search concentrates on Higgs boson production in association with a W or Z boson using the following decay modes: $W \rightarrow e\nu/\mu\nu$ and $Z \rightarrow ee/\mu\mu/\nu\nu$. The $Z \rightarrow \nu\nu$ decay is identified by the requirement of a large missing transverse energy. The Higgs boson candidate is reconstructed by requiring two b-tagged jets.

Evidence for a Higgs boson decaying to a $\tau\tau$ lepton pair is reported by the CMS and ATLAS Collaborations. The results are given in Table 2. The CMS results reported in Table 2 include both the $H \rightarrow \tau\tau$ and $H \rightarrow WW$ contributions, considered as signal in the $\tau\tau$ decay-tag analysis. This treatment leads to an increased sensitivity to the presence of a Higgs boson that decays into both $\tau\tau$ and WW .

The CMS measurements in the $H \rightarrow \tau\tau$ and VH with $H \rightarrow b\bar{b}$ searches are mutually consistent, within the precision of the present data, and with the expectation for the production and decay of the SM Higgs boson. CMS has combined these two results, requiring the simultaneous analysis of the data selected by the two individual measurements. Figure 14 shows the scan of the profile likelihood as a function of the signal strength relative to the expectation for the production and decay to fermions ($b\bar{b}$ and $\tau\tau$) of a standard model Higgs boson for $m_H = 125$ GeV. The evidence against the background-only hypothesis is found to have a maximum of 3.8σ for $m_H = 125$ GeV.

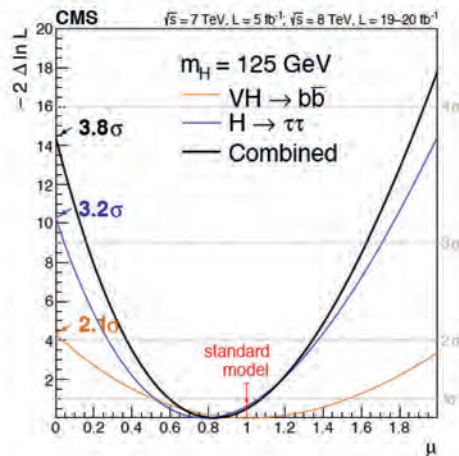


Fig. 14: Scan of the profile likelihood as a function of the signal strength relative to the expectation for the production and decay of a standard model Higgs boson, for $m_H = 125$ GeV

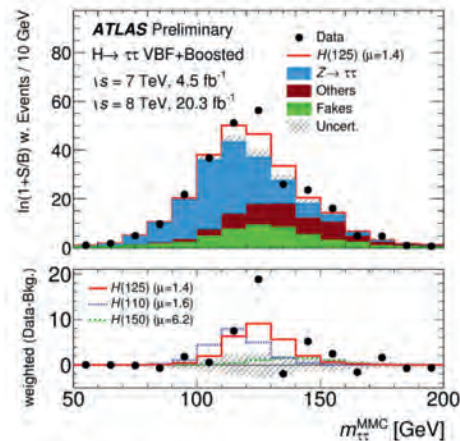


Fig. 15: Observed and expected weighted di-tau mass distributions in ATLAS. The bottom panel shows the difference between weighted data events and weighted background events (points) compared to signal events yields for various masses, with signal strengths set to their best-fit values.

Figure 15 shows the observed and expected $\tau\tau$ mass distributions from the ATLAS experiment, weighing all sub-distributions in each category of each channel by the ratio between the expected signal and background yields for the respective category in a di-tau mass interval containing 68% of the signal. The plot also shows the difference between the observed data and expected background distributions, together with the expected distribution for a SM Higgs boson signal with $m_H = 125$ GeV. The observed (expected) significance of the excess with respect to the background-only hypothesis at this mass is 4.5 (3.4) standard deviations in the ATLAS experiment.

The Tevatron experiments, CDF and D0, have also reported a combined observed significance of 3.0σ [27], where the $H \rightarrow b\bar{b}$ mode is the dominant one. All these results establish the existence of the fermionic decays of the new boson, consistent with the expectation from the SM.

4.2.3 Higgs Boson Properties

4.2.3.1 The mass of the Higgs boson

Both ATLAS and CMS experiments have separately combined their measurements of the mass of the Higgs bosons from the two channels that have the best mass resolution, namely $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow 4l$. The signal in all channels is assumed to be due to a state with a unique mass. The obtained values are from ATLAS $m_H = 125.36 \pm 0.37(\text{stat}) \pm 0.18(\text{syst})$ GeV and from CMS $m_H = 125.02 \pm 0.27(\text{stat}) \pm 0.14(\text{syst})$ GeV, in excellent agreement.

4.2.3.2 Significance of the observed excess

Table 2 summarises the median expected and observed local significances for a SM Higgs boson mass hypothesis of 125 GeV from the individual decay modes in ATLAS and CMS. Both experiments confirm independently the discovery of a new particle with a mass near 125 GeV.

Table 2: The expected and observed local p -values in ATLAS and CMS expressed as the corresponding number of standard deviations of the observed excess from the background-only hypothesis, for $m_H = 125$ GeV, for various decay modes.

Experiment	ATLAS		CMS	
	Expected (σ)	Observed (σ)	Expected (σ)	Observed (σ)
$\gamma\gamma$	4.6	5.2	5.3	5.6
ZZ	6.2	8.1	6.3	6.5
WW	5.8	6.1	5.4	4.7
$b\bar{b}$	2.6	1.4	2.6	2.0
$\tau\tau$	3.4	4.5	3.9	3.8
$\tau\tau + b\bar{b}$ combined	-	-	4.4	3.8

4.2.3.3 Signal strength

To establish whether or not the newly found state is the Higgs boson of the SM, one needs to precisely measure its other properties and attributes. Several tests of compatibility of the observed excesses with those expected from a standard model Higgs boson have been made. In one comparison labelled as the signal strength $\mu = \sigma/\sigma_{\text{SM}}$, the measured production \times decay rate of the signal is compared with the SM expectation, determined for each decay mode individually and for the overall combination of all channels. A signal strength of one would be indicative of a SM Higgs boson.

Both the ATLAS and CMS experiments have measured μ values, by decay mode and by additional tags used to select preferentially events from a particular production mechanism. The best-fit value for the common signal strength μ , obtained in the different sub-combinations and the overall combination of all search channels in the ATLAS and CMS experiments is shown in Fig. 16. The observed μ value is $1.00 \pm 0.09(\text{stat}) \pm 0.08(\text{theory})$ for CMS for a Higgs boson mass of 125.0 GeV and 1.30 ± 0.18 in ATLAS for a Higgs boson mass of 125.5 GeV. In both the experiments the μ -values are consistent with the value expected for the SM Higgs boson ($\mu = 1$). The Tevatron has also measured the value of this signal strength, primarily using the $b\bar{b}$ channel and find it to be 1.44 ± 0.59 [27].

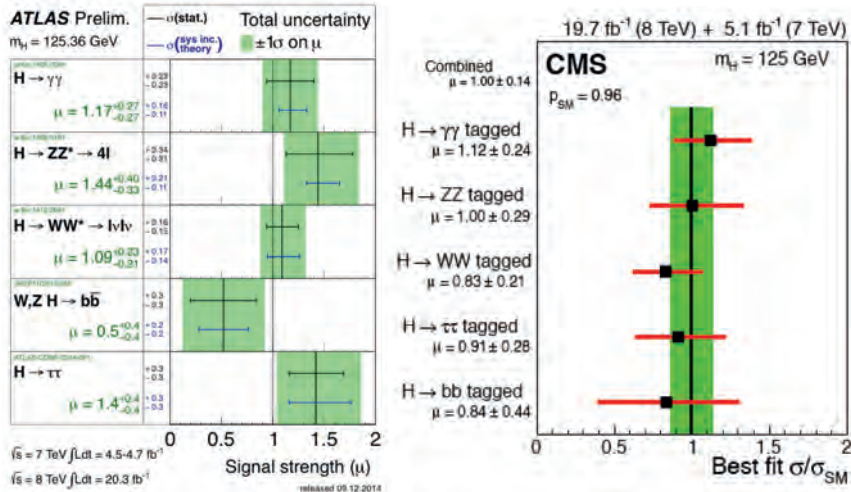


Fig. 16: Values of μ for sub-combinations by decay mode in (left) ATLAS and (right) in CMS

4.2.3.4 Couplings of the Higgs boson

Figure 17 illustrates the dependence of the Higgs boson couplings on the mass of the decay particles (τ , b-quark, W, Z and t-quark). The couplings are plotted in terms of λ or $\sqrt{(g/2v)}$. The line is the expectation from the SM. For the fermions, the values, λ , of the fitted Yukawa couplings Hff are shown, while for vector bosons the square-root of the coupling for the HVV vertex divided by twice the vacuum expectation value of the Higgs boson field ($\sqrt{(g/2v)}$). For a Higgs boson with a mass of 125 GeV decaying to $\mu\mu$ CMS has found that the observed (expected) upper limit on the production rate is 7.4 (6.5 +2.8, -1.9). This corresponds to an upper limit on the branching fraction of 0.0016. The couplings are indeed proportional to mass, as expected for a SM Higgs boson, over a broad mass range, from the τ -lepton mass (about 1.8 GeV) to that of the top quark (mass about one hundred times larger).

4.2.3.5 Spin and parity

Another key to the identity of the new boson is its quantum numbers amongst which is the spin-parity (J^P). The angular distributions of the decay particles can be used to test various spin hypotheses.

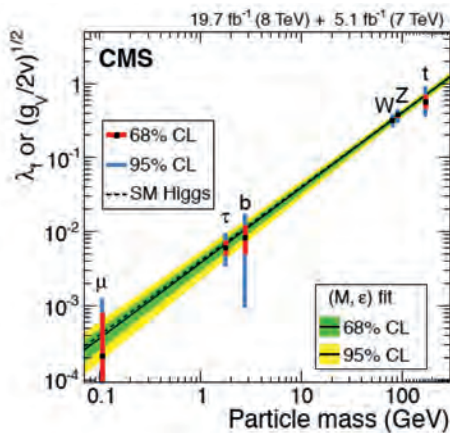


Fig. 17: Summary of the fits from the CMS experiment for deviations in the couplings λ or $\sqrt{(g/2v)}$ as function of particle mass for a Higgs boson with a mass of 125 GeV (see text)

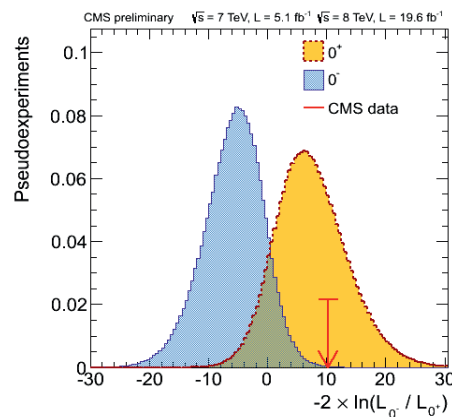


Fig. 18: Distribution of $q = -2\ln(L_{JP}/L_{SM})$ for two signal types, 0^+ (yellow histogram) and 0^- hypothesis (blue histogram) for $m_H = 126$ GeV for a large number of generated experiments. The arrow indicates the observed value.

In the decay mode $H \rightarrow ZZ \rightarrow 4l$ the full final state is reconstructed, including the angular variables sensitive to the spin-parity. The information from the five angles and the two di-lepton pair masses are combined to form boosted decision tree (BDT) discriminants. A decision tree is a set of cuts employed to classify events as “signal-like” or “background-like”.

In the decay mode $H \rightarrow WW \rightarrow l\nu l\nu$, for example, in the ATLAS experiment the discriminants used in the fit are outputs of two different BDTs, trained separately against all backgrounds to identify 0^+ and 2^+ events, respectively. For the BDT the kinematic variables used are the transverse mass m_T , the azimuthal separation of the two leptons, $\Delta\phi_l$, m_{ll} and dilepton p_{ll}^T .

A first study has been presented by CMS in the $ZZ \rightarrow 4l$ channel with the data already disfavouring the pure pseudo-scalar hypothesis (Fig. 18). The CMS experiment has combined the $ZZ \rightarrow 4l$ and $WW \rightarrow l\nu l\nu$ spin analyses. Under the assumption that the observed boson has $J^P=0^+$, the data disfavour the hypothesis of a graviton-like boson with minimal couplings produced in gluon fusion, $J^P=2^+$, with a CLs value of 0.60%.

ATLAS has also presented a combined study of the spin of the Higgs boson candidate using the $H \rightarrow \gamma\gamma$, $H \rightarrow WW \rightarrow l\nu l\nu$ and $H \rightarrow ZZ \rightarrow 4l$ decays to discriminate between the SM assignment of $J^P=0^+$ and a specific model of $J^P=2^+$. The data strongly favour the $J^P=0^+$ hypothesis (see Fig. 19). The specific $J^P=2^+$ hypothesis is excluded with a confidence level above 99.9%, independently of the assumed contributions of gluon fusion and quark-antiquark annihilation processes in the production of the spin-2 particle.

The above-mentioned example analyses show that the spin-parity $J^P=0^+$ hypothesis is strongly favoured by both experiment, with the alternatives $J^P = 0^-, 1^+, 1^-, 2^+$ hypotheses rejected with confidence levels larger than 97.8%.

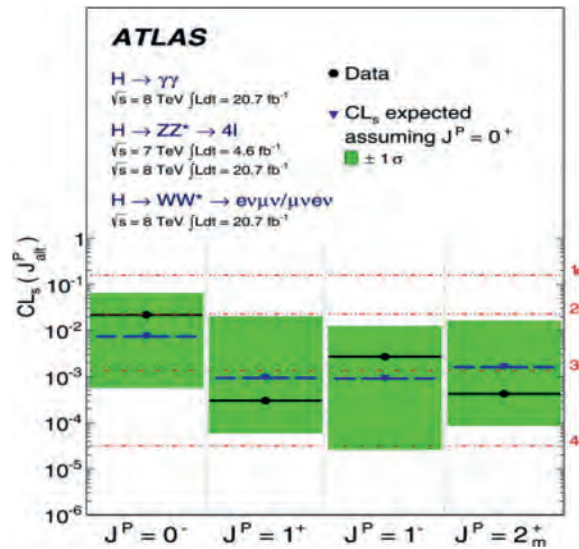


Fig. 19: Expected (blue triangles/dashed lines) and observed (black circles/solid lines) confidence level CL_s for alternative spin-parity hypotheses assuming a $J^P=0^+$ signal. The green band represents the 68% $CL_s(J_{alt}^P)$ expected exclusion range for a signal with assumed $J^P=0^+$.

5 Beyond the Standard Model at the LHC

Besides the quest to elucidate the mechanism of the electro-weak symmetry breaking by searching for the Higgs boson, the major excitement for LHC comes from the great potential to explore uncharted territory of physics Beyond the Standard Model (BSM), thanks to its highest collision energy ever available in a laboratory so far. Since the beginning of the project, the search for Supersymmetry (SUSY) was a strong motivation, and besides the H boson it has been the other main benchmark physics that was guiding the detector designs. However, many other hypothetical new processes can

be searched for, and indeed ATLAS and CMS have already reported in many publications a very broad spectrum of searches for BSM signatures (mass peaks for new particles or kinematical distributions with deviations from the expectations of known physics processes). No such new effect has yet been found, and all of these searches result in highly-improved, stringent exclusion limits, often well beyond the one TeV scale already. Only a few examples are mentioned below.

The most popular searches concern SUSY, which predicts additional fundamental particles. The search for SUSY is motivated in part by the prospect that the lightest stable neutral SUSY particle (LSP) could be an excellent candidate for explaining the Dark Matter (DM) in the Universe. The mysterious existence of DM was postulated by Fritz Zwicky, and rather convincingly evidenced by Vera Rubin, both astronomers, in the 1930s and 1970s respectively.

The SUSY searches at LHC are very complex as they must be sensitive to many (model-dependent) decay chains, implying a large variety of possible final state topologies. A common feature for most of them is the existence of significant missing transverse energy, E_T^{miss} , due to the escaping LSPs (an experimental signature similar to that of the neutrinos in the W decays). Furthermore, the SUSY signatures often include high transverse momentum jets, some tagged as b-jets for third-generation squarks as particularly motivated by naturalness arguments developed in other lectures, and leptons. The expected topologies depend not only on the model parameters, but also on the mass relations between various squarks and gluinos (the SUSY partners of the SM quarks and the gluons). A summary of 95% CL mass exclusion regions from many SUSY searches is shown in Fig. 20 from ATLAS. Very similar results are available from CMS.

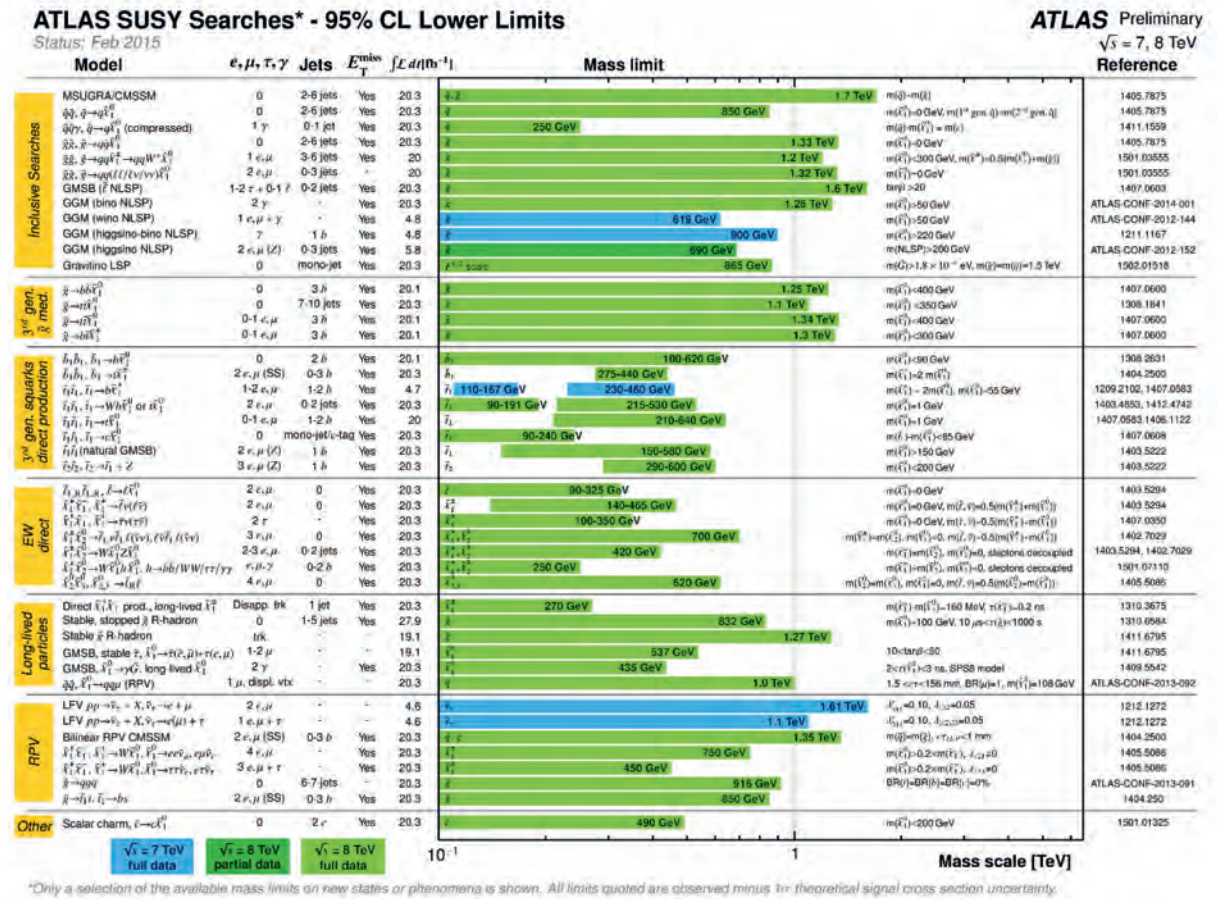


Fig. 20: A summary of 95% CL mass exclusion limits from many SUSY searches as obtained by ATLAS (very similar results are available from CMS as well).

Many other searches aimed at exploring BSM physics have been conducted, at this stage all without finding anywhere an excess of observed event rates over the expected backgrounds from the known SM processes. However, much more stringent limits and constraints could be established than what were available up to now. A non-exhaustive summary of 95% CL limits is displayed in Fig. 21 from CMS, with very similar results being reported by ATLAS.

Outlook

In spring 2015 the LHC will start operation again with Run-2 at the collision energy of 13 TeV and eventually 14 TeV in the coming years. Together with the increase in energy there will be also an increase in the luminosity, bringing the LHC to its full design performance (14 TeV, $2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$). The projected integrated luminosity by 2022 is about 300 fb^{-1} .

The increased energy means larger cross-sections, particularly striking for heavy objects, as can be seen in Fig. 22. It is therefore with great expectations that the experiments are looking forward to collect data in the forthcoming Run-2 and Run-3 periods, covering the initial LHC project planning.

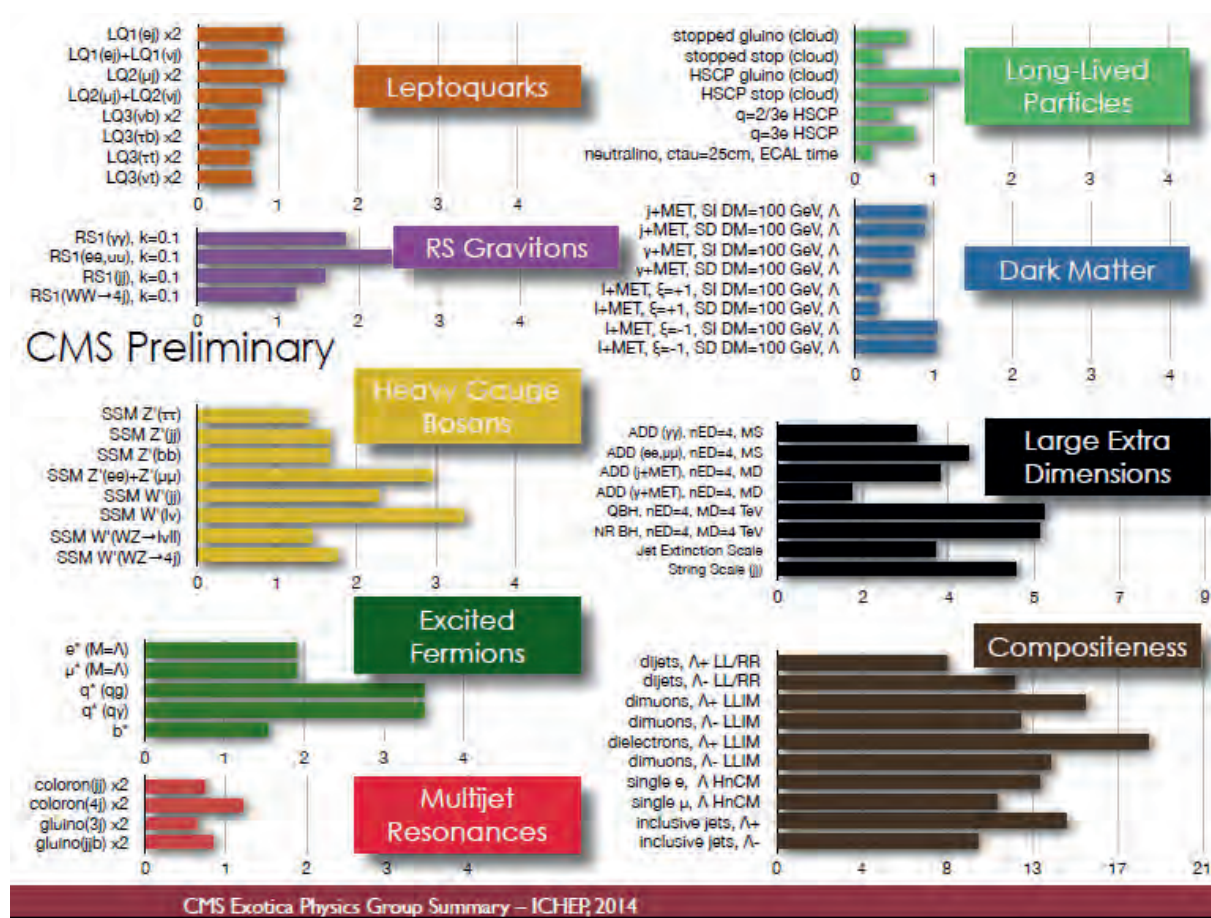


Fig. 21: A summary of 95% CL exclusion limits from many BSM searches other than SUSY as obtained by CMS (very similar results are available from ATLAS as well).

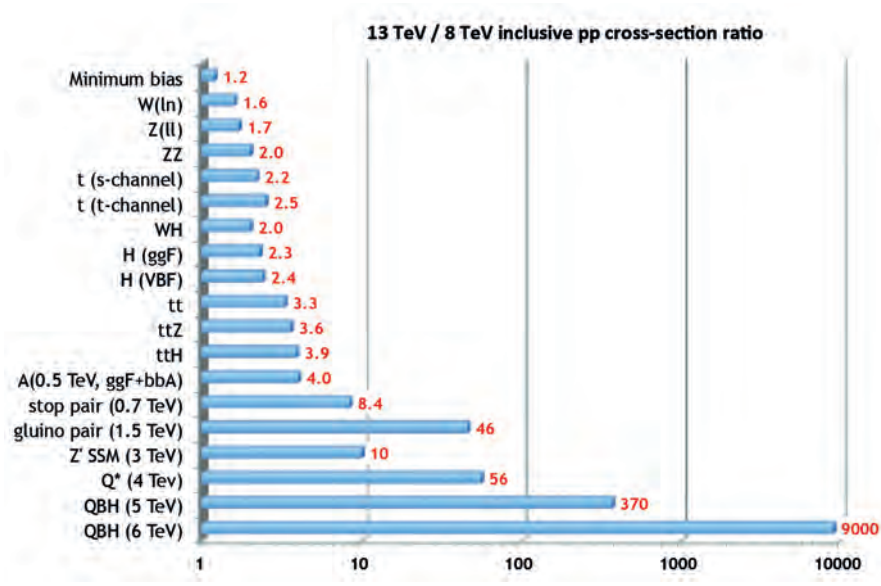


Fig. 22: Examples of inclusive production cross-section ratios for 13 TeV / 8 TeV

5.1 Prospects with the High Luminosity Upgrades (HL-LHC)

The roadmap of physics at the LHC beyond its initial design phase, with typically 300 fb^{-1} integrated luminosity until the early 2020s, has dramatically changed with the discovery of the Higgs-like boson. Not only will there be the unchallenged window for directly observable hypothetical heavy mass particles, messengers of new physics beyond the Standard Model, but also a clear task to investigate in greatest details the properties of the new boson. Needless to say, this basic scenario could well be strongly enriched further if the forthcoming 14 TeV data of the current decade would reveal any new BSM physics, which would then be of course exploited best with the highest available integrated luminosity.

These prospects have strongly motivated to launch a very mayor high luminosity upgrade project planning both for the experiments and the LHC machine, called the HL-LHC, with the goal to integrate a tenfold luminosity (3000 fb^{-1}) by the early 2030s. The importance of this future direction for particle physics has been fully recognized in the Update of the European Strategy for Particle Physics [28] where HL-LHC is singled out as first-priority in the European road map for the decades to come. A few examples are given here based on the studies in this context. The Fig. 23 illustrates the updated anticipated road map for the LHC operation for the coming decades.

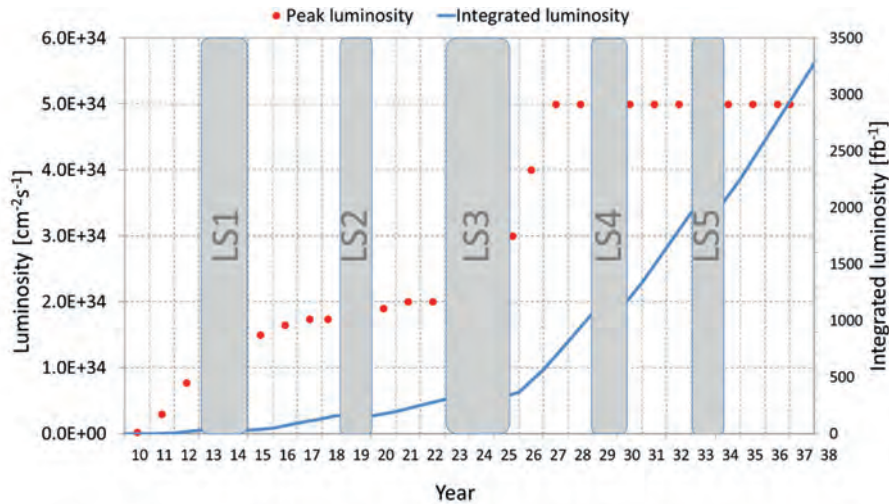


Fig. 23: Updated anticipated road-map for the LHC operation, with the HL-LHC starting after the Long Shut-Down number 3 (LS3) around 2025

The LHC potential for detailed studies of the electroweak symmetry breaking mechanism will be discussed first, namely the precision measurements of the Higgs couplings, the Higgs self-coupling, and vector boson scattering at high energy. In a second part a few examples of extending the reach into exploratory BSM physics will be given, including SUSY and searches for massive heavy resonances. The ATLAS and CMS Collaborations have presented a wealth of evaluations for the physics reach with the anticipated luminosity of 3000 fb^{-1} for the HL-LHC era [29,30]. These estimates, given here per single experiment, are based on a very substantial simulation effort taking into account realistic pile-up conditions. Both Collaborations work on very substantial detector upgrade projects that will maintain similar detector performances as at present, and which are needed in any case for critical components to allow operation beyond the initial LHC design era. Note that similarly LHCb and ALICE have engaged into very major upgrade projects as well.

5.1.1 Measurements of Higgs boson couplings

While measurements of the Higgs boson couplings have already begun by ATLAS and CMS, these will remain a central topic within the approved LHC programme. The luminosity of the HL-LHC will provide further substantially improved statistical precision for all established channels. However, most importantly, it will also allow one to study crucial rare Higgs boson production and decay modes.

Two examples for families of channels that will only become accessible in a quantitative way with the HL-LHC are mentioned here for illustration:

- WH / ZH, $H \rightarrow \gamma\gamma$ and ttH, $H \rightarrow \gamma\gamma$. These channels have a low signal rate at the LHC, but one can expect to observe more than 100 events at the HL-LHC. The ttH initial state gives the cleanest signal with a signal-to-background ratio (S/B) of $\sim 20\%$. It also provides a measurement of the top-Yukawa coupling, which is not easily accessible elsewhere. Figure 24a shows the expected signal.
- $H \rightarrow \mu\mu$. The S/B of this low-rate channel is only $\sim 0.2\%$ but the narrow peak allows one to extract a more than 6σ significant signal for an inclusive measurement, see Fig. 24b. The exclusive ttH, $H \rightarrow \mu\mu$ would yield a clean (S/B > 1) sample of 30 events providing information on both top- and μ -Yukawa couplings.

An overview of the expected measurement precisions on the signal rate in each channel is given in Fig. 25 from ATLAS, but very similar results are available from CMS as well, comparing 300 and 3000 fb^{-1} . It should be stressed that only a limited selection of channels (initial and final states) were studied so far, and further improvements can be expected with future studies.

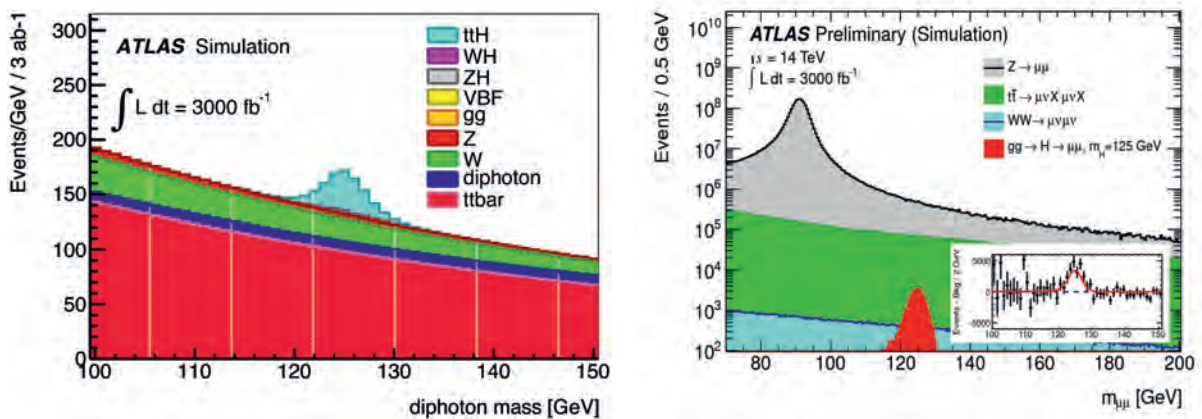


Fig. 24: Examples of expected invariant mass distributions for 3000 fb^{-1} , (a) for ttH, $H \rightarrow \gamma\gamma$ selected with 1 lepton, and (b) inclusive $H \rightarrow \mu\mu$.

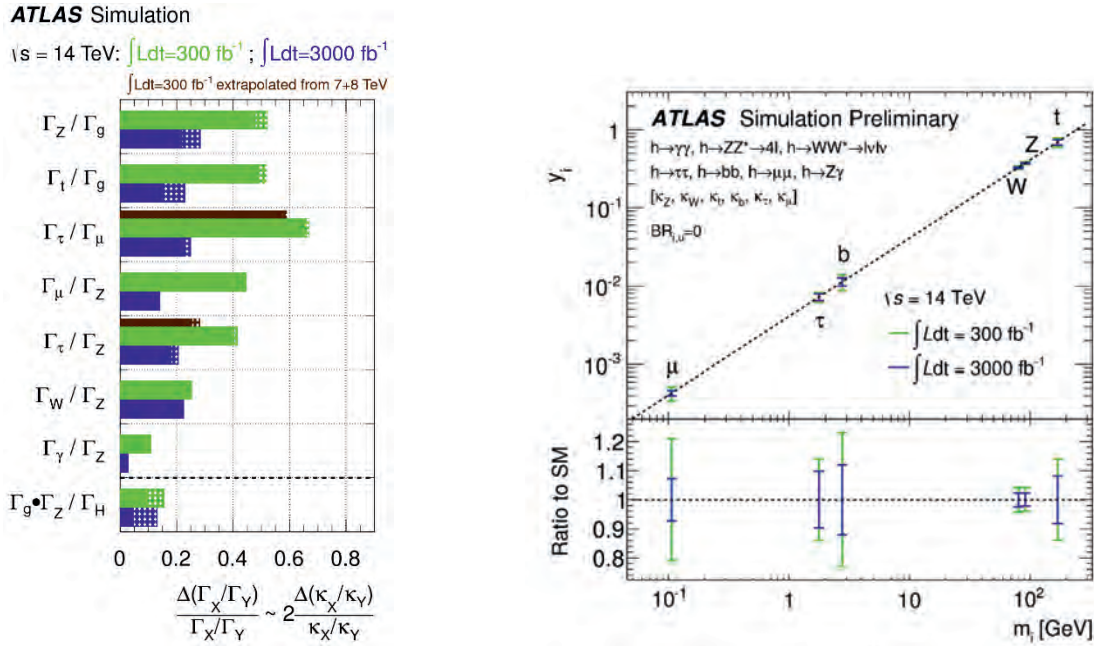


Fig. 25: (a) Expected precisions on ratios of Higgs boson partial widths. The bars give the expected relative uncertainty for a SM Higgs with mass 125 GeV (dashed are current theory uncertainty from QCD scale and PDFs). The thin bar for $\tau\tau$ show extrapolations from current analysis to 300 fb^{-1} , instead of the dedicated studies for VBF channels. (b) Expected precisions for the couplings (see also Fig. 17).

5.1.2 Observation of the Higgs self-coupling

In order to fully determine the parameters of the SM and to establish the EW symmetry breaking mechanism, the measurement of the Higgs self-coupling is important. A direct analysis of the Higgs trilinear self-coupling λ_{HHH} can be done via the detection of Higgs boson pair production, through interference effects with the dominant pair production at LHC by gluon-gluon fusion. Initial sensitivity studies have been performed only on two channels so far, $HH \rightarrow b\bar{b}\gamma\gamma$ and $bbWW$, for their clean signature and high branching ratio, respectively. Only the $b\bar{b}\gamma\gamma$ final state has been found to be accessible for 3000 fb^{-1} , yielding a 3σ observation per experiment. Additionally, promising channels like $bb\tau\tau$ in the final state are under investigation. The expectation is that a 30% measurement on λ_{HHH} can be achieved by combining the HL-LHC measurements.

5.1.3 Vector boson scattering

If the new boson discovered at LHC is fully confirmed to be the SM Higgs, then unitarity of scattering amplitudes in longitudinal Vector Boson Scattering (VBS) should be preserved at high energy. It is important to confirm this prediction experimentally. It would also be important to look for new physics contributing to the regularization of the cross section or else enhancing it. For example, Technicolour or little Higgs models, postulate TeV scale resonances to become observable.

At the LHC the VBS are tagged with two forward jets on either side, the remnants of the quarks that have emitted the vector bosons involved in the scattering process. Studies of several channels have been reported for different VB decay final states for $WW+2\text{jets}$, $WZ+2\text{jets}$, and $ZZ+2\text{jets}$ events. As an example the clean channel $ZZ+2\text{jets} \rightarrow 4 \text{ charged leptons} + 2\text{jets}$, which would allow one to fully reconstruct a hypothetical 1 TeV mass ZZ resonance peak over the SM VBS events and non-VBS di-boson background has been reported in [29].

5.1.4 Exploratory Beyond Standard Model physics at HL-LHC

Exploratory physics reach for BSM has always been a great motivation for the LHC, and that remains true more than ever also for the HL-LHC. Many quantitative studies exist, and have been refined now with sophisticated simulations by ATLAS and CMS with their realistic detector understanding, gained by the current LHC running in already very challenging pile-up conditions.

Considering first Supersymmetry (SUSY) searches, the new studies have confirmed that the mass reach in the generic searches for gluinos and squarks of the first two generations will be extended from typically 2.6 TeV to 3.2 TeV when adding the HL-LHC data. These results remain essentially unchanged for lightest supersymmetry particle (LSP) masses up to 1/3 of the mass of the strongly produced sparticles.

Naturalness arguments suggest the top squark to be light, preferably below 1 TeV. At 14 TeV the direct stop pair production cross section for 600 GeV (1 TeV) stops is 240 fb (10 fb). An increase in the luminosity from 300 to 3000 fb⁻¹ increases therefore the sensitivity significantly for heavy stop in the interesting region or, if stop candidates are found, will enable to measure their properties. As an illustrative example of a new detailed study the Fig. 26 summarizes the results in the stop-LSP plane for two decay chains. Both the 5σ discovery range and the 95 % CL exclusion limits are shown. The cross sections for electroweak gaugino searches are small at the LHC, and the discovery potential will get strongly enhanced by the ten-fold luminosity increase. For example, the discovery potential for associated production of charginos and neutralinos extends to scenarios with chargino masses of about 800 GeV for neutralino masses below 300 GeV.

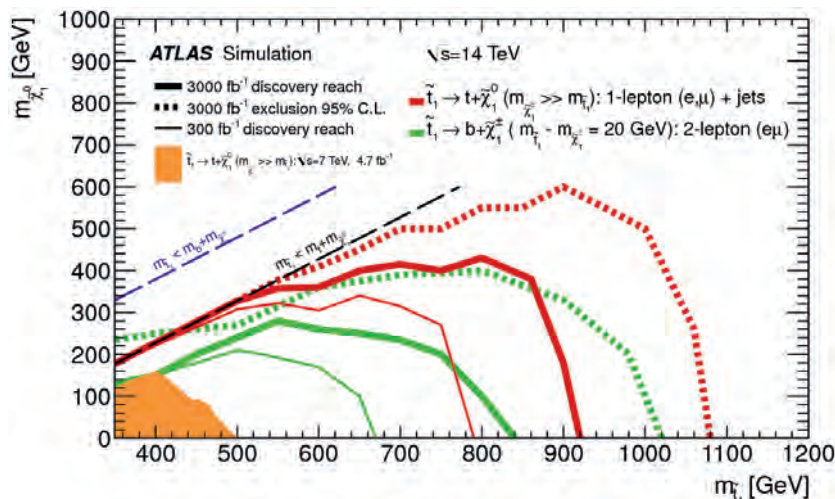


Fig. 26: 5σ discovery reach and 95% CL exclusion limits in the stop-LSP mass plane for two decay channels, as indicated, for direct stop pair production.

A broad variety of resonances and other exotic signatures are sought for at the LHC. The reach for direct observations extends deep into the TeV mass scale, as a typical example one can quote the straight-forward searches for new sequential standard model like Z' decaying into charged lepton pairs. The mass reach of typically 6.5 TeV with 300 fb⁻¹ will increase to 7.8 TeV with 3000 fb⁻¹. This improved reach of about 20% is very typical for many other searches.

A notable area of exotic physics that will benefit particularly from an HL-LHC phase is the sector of final states with top quarks. Strongly and weakly produced top-antitop resonances have been studied as an interesting benchmark. For example, strongly-produced Kaluza-Klein gluons in extradimension models could result in broad top-antitop resonance signals. The mass reach for them will increase very significantly from 4.3 TeV at 300 fb⁻¹ to 6.7 TeV with 3000 fb⁻¹.

5.2 Prospects beyond the LHC

In the lectures an outlook for ambitious, future facilities beyond the LHC project was given. A rich variety of ambitious project dreams are pursued in the community to explore further the High Energy Frontier for many decades to come. They include hadron colliders as well as e^+e^- colliders (linear colliders and circular storage rings), a field too vast to describe here in this limited write-up. Studies are evolving fast, and the students are encouraged to consult the updated web information available for the CERN Future Circular Collider (FCC) studies [31], the CERN Compact Linear Collider (CLIC) studies [32], the International Linear Collider (ILC) studies [33] and the Chinese collider ring complex (CEPC-SPPC) studies [34].

With the LHC the journey into new physics territory at the high energy frontier has only just begun, and rarely before have we enjoyed such an exciting time in particle physics with great promises for discoveries. But the long LHC story, still only at the beginning of its exploitation, has also told us that timely plans and courageous decisions on a global scale have to be made by the world community of particle physicists, in order to ‘plant the right seed’ for the future of our field.

Acknowledgement

The 8th CERN Latin-American School of High-Energy Physics has been a most enjoyable time for me with great and motivating interactions with excellent students. The school was perfectly organized and I thank its organizers, Nick Ellis, Martijn Mulders and Edgar Fernando Carrera Jarrin, for having invited me to give these lectures. All the travel and stay at Ibarra has been made a pleasure by Kate Ross, whose friendly assistance is warmly acknowledged.

For this write-up I have significantly benefitted from summary articles that I had the pleasure to co-author with long-standing colleagues in the field of hadron colliders, namely Tejinder S. Virdee, Paul Grannis and Giorgio Brianti.

References

- [1] ATLAS Collaboration, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, *Phys. Lett.* **B 716** (2012) 1.
- [2] CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, *Phys. Lett.* **B 716** (2012) 30.
- [3] S.L. Glashow, *Nucl. Phys.* **22** (1961) 579.
- [4] S. Weinberg, *Phys. Rev. Lett.* **19** (1967) 1264.
- [5] A. Salam, Proceedings of the eighth Nobel symposium, ed. N. Svartholm, p. 367. Almqvist & Wiskell, 1968.
- [6] F. Englert and R. Brout, *Phys. Rev. Lett.* **13** (1964) 321.
- [7] P.W. Higgs, *Phys. Lett.* **12** (1964) 132.
- [8] P.W. Higgs, *Phys. Rev. Lett.* **13** (1964) 508.
- [9] G.S. Guralnik, C.R. Hagen, T.W.B. Kibble, *Phys. Rev.* **155** (1967) 1554.
- [10] P.W. Higgs, *Phys. Rev.* **145** (1966) 1156.
- [11] T.W.B. Kibble, *Phys. Rev.* **155** (1967) 1554.
- [12] M. Della Negra, P. Jenni, T. S. Virdee, Journey in the Search for the Higgs Boson: The ATLAS and CMS Experiments at the Large Hadron Collider, *Science* Vol. **338** no. 6114 pp. 1560-1568 (2012), <http://www.sciencemag.org/content/338/6114/1560.full.html>.
- [13] P. Jenni and T. S. Virdee, The Discovery of the Higgs Boson at the LHC, in *60 Years of CERN Experiments*, Eds. H. Schopper and L. Di Lella (World Scientific, 2015), 978-981-4644-14-3.

- [14] P. Grannis and P. Jenni, The Evolution of Hadron-Collider Experiments, *Phys. Today* **66(6)** (2013) 38.
- [15] ATLAS publications: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic>
- [16] CMS publications: <http://cms-results.web.cern.ch/cms-results/public-results/publications/>
- [17] *The Large Hadron Collider, a Marvel of Technology*, Ed. L. Evans, EPFL Press, 2009.
- [18] G. Brianti and P. Jenni, The Large Hadron Collider, in *Technology Meets Research, 60 Years of Technology Achievements at CERN*, Eds. C. Benvenuti et al. (World Scientific, 2015).
- [19] C. Llewellyn Smith, *Nature* **448** (2007) 281.
- [20] *ECFA-LEP Working Group, 1979 Progress Report*, Ed. A. Zichichi, ECFA/79/39 (1979).
- [21] ATLAS Collaboration, The ATLAS Experiment at the LHC, *JINST* **3**, S08003 (2008).
- [22] CMS Collaboration, The CMS Experiment at the LHC, *JINST* **3**, S08004 (2008).
- [23] LHCb Collaboration, The LHCb Detector at the LHC, *JINST* **3**, S08005 (2008).
- [24] ALICE Collaboration, The ALICE Experiment at the LHC, *JINST* **3**, S08002 (2008).
- [25] WLCG comprehensive information: <http://wlcg.web.cern.ch/>
- [26] A. Hoecker, Commissioning and early physics analysis with the ATLAS and CMS experiments, in *5th CERN – Latin-American School of High Energy Physics*, 2009, CERN-2010-001, 449, arXiv:1010.5976.
- [27] CDF and D0 Collaborations, Higgs boson studies at the Tevatron, *Phys. Rev.* **D88** (2013) 052014.
- [28] The European Strategy for Particle Physics Update 2013, *CERN-Council-S/106* (2013), <http://cds.cern.ch/record/1567258/files/esc-e-106.pdf>
- [29] ATLAS Collaboration, Physics at a High-Luminosity LHC with ATLAS, *ATL-PHYS-PUB-2013-007* (2013), arXiv:1307.7292v2[hep-ex].
- [30] CMS Collaboration, Projected Performance of an Upgraded CMS Detector at the LHC and HL-LHC, *CMS NOTE-13-002* (2013), arXiv:1307.7135v2[hep-ex].
- [31] CERN Future Circular Collider (FCC) web site: <https://fcc.web.cern.ch/Pages/default.aspx>
- [32] CERN Compact Linear Collider (CLIC) web site: <http://clic-study.web.cern.ch/>
- [33] International Linear Collider (ILC) web site: <https://www.linearcollider.org/ILC>
- [34] Chinese Circular Electron Positron Collider web site: <http://cepc.ihep.ac.cn/>

Practical Statistics

L. Lyons

Blackett Lab., Imperial College, London, UK and Particle Physics, Oxford, UK

Abstract

Accelerators and detectors are expensive, both in terms of money and human effort. It is thus important to invest effort in performing a good statistical analysis of the data, in order to extract the best information from it. This series of five lectures deals with practical aspects of statistical issues that arise in typical High Energy Physics analyses.

Keywords

Statistics; lectures ; data analysis method; statistical analysis; frequentist; Bayesian.

1 Outline

This series of five lectures deals with practical aspects of statistical issues that arise in typical High Energy Physics analyses. The topics are:

- Introduction. This is largely a reminder of topics which you should have encountered as undergraduates. Some of them are looked at in novel ways, and will hopefully provide new insights.
- Least Squares and Likelihoods. We deal with two different methods for parameter determination. Least Squares is also useful for Goodness of Fit testing, while likelihood ratios play a crucial role in choosing between two hypotheses.
- Bayes and Frequentism. These are two fundamental and very different approaches to statistical searches. They disagree even in their views on ‘What is probability?’
- Searches for New Physics. Many statistical issues arise in searches for New Physics. These may result in discovery claims, or alternatively in exclusion of theoretical models in some region of their parameter space (e.g. mass ranges).
- Learning to love the covariance matrix. This is relevant for dealing with the possible correlations between uncertainties on two or more quantities. The covariance matrix takes care of all these correlations, so that you do not have to worry about each situation separately. This was an unscheduled lecture which was included at the request of several students.

Lectures 3 to 5 are not included in these proceedings but can be found elsewhere [1–3].

The material in these lectures follows loosely that in my book [4], together with some significant updates (see ref. [5]).

2 Introduction to Lecture 1

The first lecture, covered in Sections 2 to 11, is a recapitulation of material that should already be familiar, but hopefully with some new emphases. We start with a discussion of ‘What is Statistics?’ and a comparison of ‘Statistics’ and ‘Probability’. Next the importance of calculating uncertainties is emphasised, as well as the difference between random and systematic uncertainties.

The following sections are about combinations. The first is about how to combine different individual contributions to a particular experimental result; the second is the combination of two or more separate experimental determinations of the same physical quantity.

The final topics are the Binomial, Poisson and Gaussian probability distributions. Understanding of these is important for many statistical analyses.

3 What is Statistics?

Statistics is used to provide quantitative results that give summaries of available data. In High Energy Physics, there are several different types of statistical activities that are used:

- Parameter Determination:
We analyse the data in order to extract the best value(s) of one or more parameters in a model. This could be, for example, the gradient and intercept of a straight line fit to the data; or the mass of the Higgs boson, as deduced using its decay products. In all cases, as well as obtaining the best values of the parameter(s), their uncertainties and possible correlations must be specified.
- Goodness of Fit:
We are comparing a single theory with the data, in order to see if they are compatible. If the theory contains free parameters, their best values need to be used to check the Goodness of Fit. If the quality of the fit is unsatisfactory, the best values of the parameters are probably meaningless.
- Hypothesis Testing:
Here we are comparing the data with two different theories, to see which provides a better description. For example, we may be very interested in knowing whether a model involving the production of a supersymmetric particle is better than one without it.
- Decision Making:
As the result of the information we have available, we want to decide what further action to take. For example, we may have some evidence that our data shows hints of an exciting discovery, and need to decide whether we should collect more data. This was the situation faced by the CERN management in 2000, when there were perhaps hints of a Higgs boson in data collected at the LEP Collider.
Such decisions usually require a ‘cost function’ for the various possible outcomes, as well as assessments of their relative probabilities. In the example just quoted, numerical values were needed for the cost of missing an important discovery if the experiment was not continued; and on the other hand of running the LEP Collider for another year and for delaying the start of building the Large Hadron Collider.
Decision Making is not considered further in these lectures.

4 Probability and Statistics

Probability theory involves starting with a model, and using it to make predictions about possible outcomes of an experiment where randomness plays a role; it involves precise mathematics, and in general there is only one correct solution about the probabilities of the different outcomes. Statistics involves the opposite procedure of using the observed data in order to make statements about the relevant theory or model. This is usually not a precise process and there may be different approaches which yield different answers, none of which being necessarily invalid.

The example of throwing dice (see Table 1) illustrates the relationship of Probability Theory and Statistics for some of the statistical procedures.

5 Why uncertainties?

Without an estimate of the uncertainty of a parameter, its central value is essentially useless. This is illustrated by Table 2. The three lines of the Table refer to different possible results; all have the same central value of the ratio of the experimental result divided by the theoretical prediction, but each has a different uncertainty on this ratio. The conclusions about whether the data supports the theory are very different, depending on the magnitude of the uncertainty, even though the central values are the same for each of the three situations. It is thus crucial to estimate uncertainties accurately, and also correlations when measuring two or more parameters.

Table 1: Probability and Statistics: Throwing dice

Probability	Statistics	Procedure
Given $p(5) = 1/6$, what is $\text{prob}(20 \text{ 5s in } 100 \text{ trials})$?	Given 20 5s in 100 trials, what is $p(5)$? and its uncertainty?	Parameter Determination
If unbiased, what is $\text{prob}(n \text{ evens in } 100 \text{ trials})$?	Given 60 evens in 100 trials, is it unbiased?	Goodness of Fit
	Or is $\text{prob}(\text{evens}) = 2/3$?	Hypothesis Testing
THEORY \rightarrow DATA	DATA \rightarrow THEORY	

Table 2: Experiment testing General Relativity.

Experiment/Theory	Uncertainty	Conclusion
0.970	± 0.05	Consistent with 1.0
0.970	± 0.006	Inconsistent with 1.0
0.970	± 0.7	Do a better experiment

6 Random and systematic uncertainties

Random or statistical uncertainties result from the limited accuracy of measurements, or from the fluctuations that arise in counting experiments where the Poisson distribution is relevant (see Section 10). If the experiment is repeated, the results will vary somewhat, and the spread of the answers provides (not necessarily the best) estimate of the statistical uncertainty.

Systematic uncertainties can also arise in the measuring process. The quantities we measure may be shifted from the true values. For example, our measuring device may be miscalibrated, or the number of events we count may be not only from the desired signal, but also from various background sources. Such effects would bias our result, and we should correct for them, for example by performing some calibration measurement. The systematic uncertainty arises from the remaining uncertainty in our corrections. Systematics can cause a similar shift in a repeated series of experiments, and so, in contrast to statistical uncertainties, they may not be detectable by looking for a spread in the results.

For example consider a pendulum experiment designed to measure the acceleration due to gravity g at sea level in a given location:

$$g = 4\pi^2 L / \tau^2 \tag{1}$$

where L is the length of the pendulum, $\tau = T/N$ is its period, and T is the time for N oscillations. The uncertainties we have mentioned so far are the statistical ones on L and T ¹. There may also be systematic uncertainties on these variables.

Unfortunately there are further possible systematics not associated with the measured quantities, and which thus require more careful consideration. For example, the derivation of eqn. (1) assumes that:

- our pendulum is simple i.e. the string is massless, and has a massive bob of infinitesimal size;
- the support of the pendulum is rigid;
- the oscillations are of very small amplitude (so that $\sin \theta \approx \theta$); and
- they are undamped.

¹Note that although N involves counting the number of swings, we do not have to allow for Poisson fluctuations, since there are no random fluctuations involved.

None of these will be exact in practice, and so corrections must be estimated for them. The uncertainties in these corrections are systematics.

Furthermore, there may be theoretical uncertainties. For example, we may want the value of g at sea level, but the measurements were performed on top of a mountain. We thus need to apply a correction, which depends on our elevation and on the local geology. There might be two or more different theoretical correction factors, and again this will contribute a systematic uncertainty.

6.1 Presenting the results

A common way of presenting the result of a measurement y is as $y \pm \sigma_{stat} \pm \sigma_{syst}$, where the statistical and systematic uncertainties are shown separately. Alternatively, it may be presented as $y \pm \sigma$, where the total uncertainty is usually given by $\sigma^2 = \sigma_{stat}^2 + \sigma_{syst}^2$.

The other extreme is to give a list of all the individual systematics separately (usually in a Table, rather than in the Abstract or Conclusions). The motivations for this are that:

- systematics are sometimes caused by uncertainties in other people’s measurements of some relevant quantity. If subsequently this measurement is updated, it will be possible to reduce the systematic uncertainty appropriately; and
- our measurement may be combined with others to produce a ‘World average’, or it may be used together with another result to calculate something else. In both these cases, correlations between the different experimental measurements are needed, and so the individual sources are required.

For example, it may be interesting to compare the sea-level values of g at the same location several years apart. In that case, although there might be significant uncertainties from the correction of the measurements to sea-level, they are a fully correlated, and so will cancel in their difference

7 Combining uncertainties

In this section, we consider how to estimate the uncertainty σ_z in a quantity of interest z , which is defined in terms of measured quantities x, y, \dots by a known function $z(x, y, \dots)$. The uncertainties on the measured quantities are known and assumed to be uncorrelated. The recipe for σ_z depends on the functional form of z .

7.1 Linear forms

As a very simple example, consider

$$z = x - y \tag{2}$$

From this, we obtain

$$\delta z = \delta x - \delta y \tag{3}$$

where δz is the change in z that would be produced by specific changes in x and y . But eqn. 3 refers to specific offsets, rather than the uncertainties σ_z , etc, which are the RMS values of the offsets i.e. $\overline{\delta z^2}$, etc. Thus we need to square eqn. 3, which yields

$$\delta z^2 = \delta x^2 + \delta y^2 - 2\delta x\delta y, \tag{4}$$

and to average over a whole series of measurements. We then obtain the correct formula for combining the uncertainties:

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2, \tag{5}$$

provided we ignore the last term in eqn 4. The justification for this is that the average value of $\delta x \delta y$ is zero, provided the uncertainties on x and on y are uncorrelated.²

For the general linear form

$$z = k_1 x + k_2 y + \dots \quad (6)$$

where k_1, k_2, \dots are constants, the uncertainty on z is given by

$$\sigma_z = k_1 \sigma_x \& k_2 \sigma_y \& \dots, \quad (7)$$

where the symbol $\&$ is used to mean ‘combine using Pythagoras’ Theorem’. For the special case of $z = x - y$, as is expected this gives the result of eqn. 5 for σ_z .

For this case of z being a **linear** function of the measurements, it is the **absolute** uncertainties that are relevant for determining σ_z . It is important **not** to use **fractional** uncertainties. Thus if you want to determine your height by making independent measurements of the distances of the top of your head and the bottom of your feet from the centre of the earth, each with an accuracy of 1 part in 1000, you will not determine your height to anything like 1 part in 1000.

7.2 Products and quotients

The general form here is

$$z = x^\alpha y^\beta \dots, \quad (8)$$

where the powers α, β , etc. are constants. This includes forms such as $x^2, y^3/x, \sqrt{x}/y$, etc. The formula for combining the uncertainties is

$$\sigma_z/z = \alpha \sigma_x/x \& \beta \sigma_y/y \& \dots \quad (9)$$

That is, the **fractional** uncertainty on z is derived from the **fractional** uncertainties on the measurements.

Because this result was derived by taking the first term of a Taylor expansion for δz , it will be a good approximation only for small uncertainties. If the uncertainties are large, more sophisticated approaches are required for determining the uncertainty in z . This also applies to the next section, but is irrelevant for the linear cases discussed above, as all terms in the Taylor series beyond those involving first derivatives are zero.

7.3 All other functions

Finally we deal with any functional form $z = z(x_1, x_2, x_3, \dots)$. Our prescription of writing down the first term in the Taylor series expansion for δz , squaring and averaging gives

$$\sigma_z = \frac{\partial z}{\partial x_1} \sigma_1 \& \frac{\partial z}{\partial x_2} \sigma_2 \& \dots \quad (10)$$

where the σ_i are the uncertainties on x_i , again assumed uncorrelated.

A slightly easier method to apply is to use a numerical approach for calculating the partial derivatives. We evaluate

$$\begin{aligned} z_0 &= z(x_1, x_2, x_3, \dots) \\ z_1 &= z(x_1 + \sigma_1, x_2, x_3, \dots) \\ z_2 &= z(x_1, x_2 + \sigma_2, x_3, \dots) \\ z_3 &= z(x_1, x_2, x_3 + \sigma_3, \dots) \\ &\text{etc.} \end{aligned} \quad (11)$$

and then

$$\sigma_z^2 = \Sigma (z_i - z_0)^2 \quad (12)$$

²Note that it is the **uncertainties** which are required to be uncorrelated. Thus for a simple pendulum, L and τ are correlated by eqn 1, but the uncertainties on the measured length and period are uncorrelated.

8 Combining experiments

Sometimes different experiments will measure the same physical quantity. It is then reasonable to ask what is our best information available when these experiments are combined. It is a general rule that it is better to use the **DATA** for the experiments and then perform a combined analysis, rather than simply combine the **RESULTS**. However, combining the results is a simpler procedure, and access to the original data is not always possible.

For a series of unbiased, uncorrelated measurements x_i of the same physical quantity, the combined value $\hat{x} \pm \hat{\sigma}$ is given by weighting each measurement by w_i , which is proportional to the inverse of the square of its uncertainty i.e.

$$\hat{x} = \Sigma w_i x_i, \quad w_i = (1/\sigma_i^2)/\Sigma(1/\sigma_j^2) \quad (13)$$

with the uncertainty $\hat{\sigma}$ on the combined value being given by

$$1/\hat{\sigma}^2 = \Sigma 1/\sigma_i^2 \quad (14)$$

This ensures that the uncertainty on the combination is at least as small as the smallest uncertainty of the individual measurements. It should be remembered that the combined uncertainty takes no account of whether or not the individual measurements are consistent with each other.

In an informal sense, $1/\sigma_i^2$ is the information content of a measurement. Then each x_i is weighted proportionally to its information content. Also the equation for $\hat{\sigma}^2$ says that the information content of the combination is the sum of the information contents of the individual measurements.

An example demonstrates that care is needed in applying the formulae. Consider counting the number of high energy cosmic rays being recorded by a large counter system for two consecutive one-week periods, with the number of counts being 100 ± 10 and 1 ± 1 ³. (See section 10 for the choice of uncertainties). Unthinking application of the formulae for the combined result give the ridiculous 2 ± 1 . What has gone wrong?

The answer is that we are supposed to use the **true** accuracies of the individual measurements to assign the weights. Here we have used the **estimated** accuracies. Because the estimated uncertainty depends on the estimated rate, a downward fluctuation in the measurement results in an underestimated uncertainty, an overestimated weight, and a downward bias in the combination. In our example, the combination should assume that the true rate was the same in the two measurements which used the same detector and which lasted the same time as each other, and hence their true accuracies are (unknown but) equal. So the two measurements should each be given a weight of 0.5, which yields the sensible combined result of 50.5 ± 5 counts.

8.1 BLUE

A method of combining correlated results is the ‘**B**est **L**inear **U**nbiased **E**stimate’ (**BLUE**). We look for the best linear unbiased combination

$$x_{BLUE} = \Sigma w_i x_i, \quad (15)$$

where the weights are chosen to give the smallest uncertainty σ_{BLUE} on x_{BLUE} . Also for the combination to be unbiased, the weights must add up to unity. They are thus determined by minimising $\Sigma \Sigma w_i w_j E_{ij}^{-1}$, subject to the constraint $\Sigma w_i = 1$; here E is the covariance matrix for the correlated measurements.

³It is vital to be aware that it is a crime (punishable by a forcible transfer to doing a doctorate on Astrology) to combine such discrepant measurements. It seems likely that someone turned off the detector between the two runs; or there was a large background in the first measurement which was eliminated for the second; etc. The only reason for my using such discrepant numbers is to produce a dramatically stupid result. The effect would have been present with measurements like 100 ± 10 and 81 ± 9 .

The *BLUE* procedure just described is equivalent to the χ^2 approach for checking whether a correlated set of measurements are consistent with a common value. The advantage of *BLUE* is that it provides the weights for each measurement in the combination. It thus enables us to calculate the contribution of various sources of uncertainty in the individual measurements to the uncertainty on the combined result.

8.2 Why weighted averaging can be better than simple averaging

Consider a remote island whose inhabitants are very conservative, and no-one leaves or arrives except for some anthropologists who wish to determine the number of married people there. Because the islanders are very traditional, it is necessary to send two teams of anthropologists, one consisting of males to interview the men, and the other of females for the women. There are too many islanders to interview them all, so each team interviews a sample and then extrapolates. The first team estimates the number of married men as $10,000 \pm 300$. The second, who unfortunately have less funding and so can interview only a smaller sample, have a larger statistical uncertainty; they estimate $9,000 \pm 900$ married women. Then how many married people are there on the island?

The simple approach is to add the numbers of married men and women, to give $19,000 \pm 950$ married people. But if we use some theoretical input, maybe we can improve the accuracy of our estimate. So if we assume that the islanders are monogamous, the numbers of married men and women should be equal, as they are both estimates of the number of married couples. The weighted average is $9,900 \pm 285$ married couples and hence $19,800 \pm 570$ married people.

The contrast in these results is not so much the difference in the estimates, but that incorporating the assumption of monogamy and hence using the weighted average gives a smaller uncertainty on the answer. Of course, if our assumption is incorrect, this answer will be biased.

A Particle Physics example incorporating the same idea of theoretical input reducing the uncertainty of a measurement can be found in the ‘Kinematic Fitting’ section of Lecture 2.

9 Binomial distribution

This and the next sections on the Poisson and Gaussian distributions are probability theory, in that they make statements about the probabilities of different outcomes, assuming that the theoretical distribution is known. However, the results are important for Statistics, where we use data in order to make statements about theory.

The binomial distribution applies when we have a set of N independent trials, in each of which a ‘success’ occurs with probability p . Then the probability $P(s; N, p)$ of s successes in the N trials is obviously

$$P(s; N, p) = \frac{N!}{s!(N-s)!} p^s (1-p)^{N-s}. \quad (16)$$

An example of a Binomial distribution would be the number of times we have a 6 in 20 throws of a die; or the distribution of the number of successfully reconstructed tracks in a sample of 100, when the probability for reconstructing each of them is 0.98

The expected number of successes $\langle s \rangle$ is $\sum s \times P(s; N, p)$, which after some algebra turns out to be (not surprisingly) Np . The variance σ_s^2 of the distribution in s is obviously given by $Np(1-p)$. Note that, while for the Poisson distribution the mean and variance are equal, this is not so in general for the Binomial - it is approximately so at small p .

As an example several Binomial distributions with fixed number of trials N but varying probabilities of success p are shown in Fig. 1.

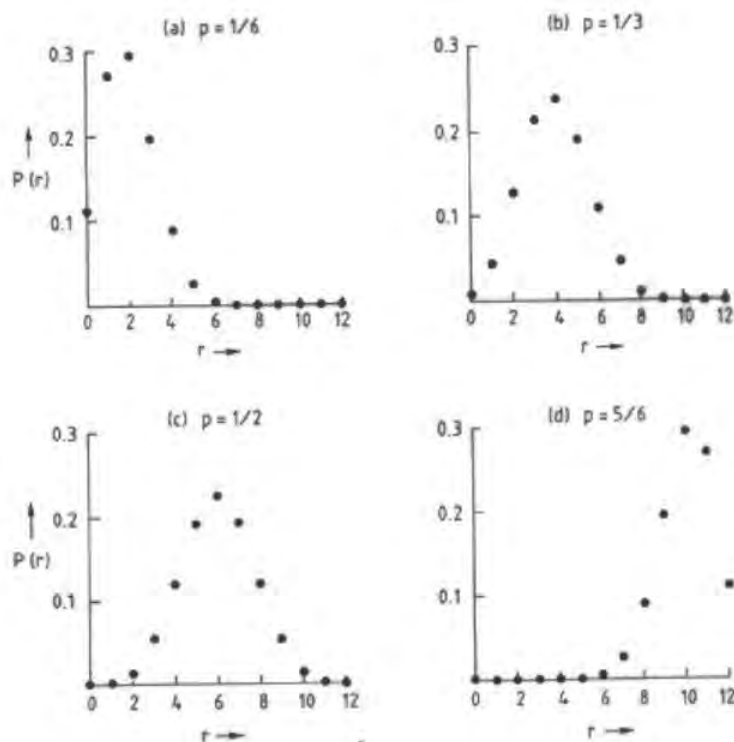


Fig. 1: The probabilities $P(r)$ according to the binomial distribution, for r successes out of 12 independent trials, when the probability p of success in an individual trial is as specified. As the expected number of successes is $12p$, the peak of the distribution moves to the right as p increases. The variance of the distribution is $12p(1-p)$ and hence is largest for $p = 1/2$. Since the chance of success when $p = 1/6$ is the same as that for failure when $p = 5/6$, diagrams (a) and (d) are mirror images of each other. Similarly for $p = 1/2$ (see (c)) the distribution is symmetric about $r = 6$ successes.

10 Poisson distribution

The Poisson distribution (see Fig. 2) applies to situations where we are counting a series of observations which are occurring randomly and independently during a fixed time interval t , where the underlying rate r is constant. The observed number n will fluctuate when the experiment is repeated, and can in principle take any integer value from zero to infinity. The Poisson probability of observing n decays is given by

$$P_n = e^{-rt}(rt)^n/n! \quad (17)$$

It applies to the number of decays observed from a large number N of radioactive nuclei, when the observation time t is small compared to the lifetime τ . It will not apply if t is much larger than τ , or if the detection system has a dead time, so that after observing a decay the detector cannot observe another decay for a period T_{dead} .

Another example is the number of counts in any specific bin of a histogram when the data is accumulated over a fixed time.

The average number of observations is given by

$$\langle n \rangle = \sum n P_n = rt \quad (18)$$

If we write the expected number as μ , the Poisson probability becomes

$$P_n = e^{-\mu} \mu^n / n! \quad (19)$$

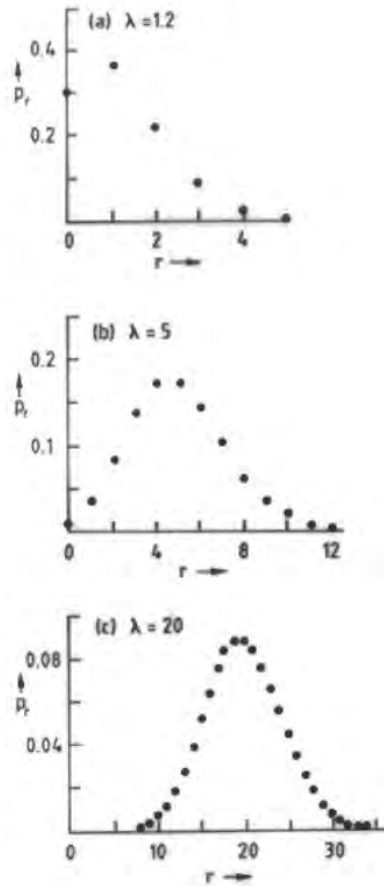


Fig. 2: Poisson distributions for various values of the Poisson parameter λ : (a) $\lambda = 1.2$ (b) $\lambda = 5.0$ (c) $\lambda = 20.0$. P_r is the probability for observing r events. For each λ , the mean value of r is λ and the RMS width is $\sqrt{\lambda}$. As λ increases above about 10, the distribution becomes more like a Gaussian.

It is also relatively easy to show that the variance

$$\sigma^2 = \Sigma(n - \mu)^2 P_n = \mu \tag{20}$$

This leads to the well-known $n \pm \sqrt{n}$ approximation for the value of the Poisson parameter when we have n counts. This approximation is, however, particularly bad when there are zero observed events; then 0 ± 0 incorrectly suggests that the Poisson parameter can be only zero.

Poisson probabilities can be regarded as the limit of Binomial ones as the number of trials N tends to infinity and the Binomial probability of success p tends to zero, but the product Np remains constant at μ .

When the Poisson mean becomes large, the distribution of observed counts approximates to a Gaussian (although the Gaussian is a continuous distribution extending down to $-\infty$, while a Poisson observable can only take on non-negative integral values). This approximation is useful for the χ^2 method for parameter estimation and goodness of fit (see Lecture 2).

10.1 Relation of Poisson and Binomial Distributions

An interesting example of the relationship between the Poisson and Binomial distributions is exhibited by the following example.

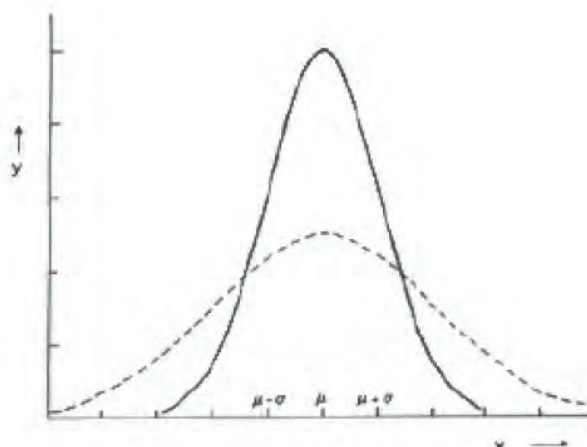


Fig. 3: Gaussian distributions. Both are centred at $x = \mu$, but the dashed curve is twice as wide as the solid one. Because they have the same normalisation the maximum of the solid curve is twice as high as that of the dashed one. The scale on the horizontal axis refers to the solid curve.

Imagine that the number of people attending a series of lectures is Poisson distributed with a constant mean ν , and that the fraction of them who are male is p . Then the overall probability P of having N people of whom M are male and $F = N - M$ are female is given by the product of the Poisson probability P_{pois} for N and the binomial probability P_{bin} for M of the N people being male. i.e.

$$P = P_{pois}P_{bin} = \frac{e^{-\nu}\nu^N}{N!} \times \frac{N!}{M!F!}p^M(1-p)^F \quad (21)$$

This can be rearranged as

$$P = \frac{e^{-\nu p}(\nu p)^M}{M!} \times \frac{e^{-\nu(1-p)}(\nu(1-p))^F}{F!} \quad (22)$$

This is the product of two Poissons, one with Poisson parameter νp , the expected number of males, and the other with parameter $\nu(1-p)$, the expected number of females. Thus with a Poisson-varying total number of observations, divided into two categories (here male and female), we can regard this as Poissonian in the total number and Binomial in the separate categories, or as two independent Poissons, one for each category. Other situations to which this applies could be radioactive nuclei, with decays detected in the forward or backward hemispheres; cosmic ray showers, initiated by protons or by heavier nuclei; patients arriving at a hospital emergency centre, who survive or who die; etc.

10.2 For your thought

The first few Poisson probabilities $P(n; \mu)$ are

$$P(0) = e^{-\mu}, \quad P(1) = \mu e^{-\mu}, \quad P(2) = (\mu^2/2!) e^{-\mu}, \quad \text{etc.} \quad (23)$$

Thus for small μ , $P(1)$ and $P(2)$ are approximately μ and $\mu^2/2$ respectively. But if the probability of one rare event happening is μ , why is the probability for 2 independent rare events not equal to μ^2 ?

11 Gaussian distribution

The Gaussian or normal distribution (shown in Fig. 3) is of widespread usage in data analysis. Under suitable conditions, in a repeated series of measurements x with accuracy σ when the true value of the

quantity is μ , the distribution of x is given by a Gaussian⁴. A mathematical motivation is given by the Central Limit Theorem, which states that the sum of a large number of variables with (almost) any distributions is approximately Gaussian.

For the Gaussian, the probability density $y(x)$ of an observation x is given by

$$y(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (24)$$

where the parameters μ and σ are respectively the centre and width of the distribution. The factor $1/(\sqrt{2\pi}\sigma)$ is required to normalise the area under the curve, so that $y(x)$ can be directly interpreted as a probability density.

There are several properties of σ :

- The mean value of x is μ , and the standard deviation of its distribution is σ . Since the usual symbol for standard deviation is σ , this leads to the formula $\sigma = \sigma$ (which is not so trivial as it seems, since the two σ s have different meanings). This explains the curious factor of 2 in the denominator of the exponential, since without it, the two types of σ would not be equal.
- The value of y at the $\mu \pm \sigma$ is equal to the peak height multiplied by $e^{-0.5} = 0.61$. If we are prepared to overlook the difference between 0.61 and 0.5, σ is the half-width of the distribution at ‘half’ the peak height.
- The fractional area in the range $x = \mu - \sigma$ to $\mu + \sigma$ is 0.68. Thus for a series of unbiased, independent Gaussian distributed measurements, about 2/3 are expected to lie within σ of the true value.
- The peak height of y at $x = \mu$ is $1/(\sqrt{2\pi}\sigma)$. It is reasonable that this is proportional to $1/\sigma$ as the width is proportional to σ , so σ cancels out in the product of the height and width, as is required for a distribution normalised to unity.

For deciding whether an experimental measurement is consistent with a theory, more useful than the Gaussian distribution itself is its tail area beyond r , a number of standard deviations from the central value (see Fig. 4). This gives the probability of obtaining a result as extreme as ours or more so as a consequence of statistical fluctuations, assuming that the theory is correct (and that our measurement is unbiased, it is Gaussian distributed, etc.). If this probability is small, the measurement and the theory may be inconsistent.

Figure 4 has two different vertical scales, the left one for the probability of a fluctuation in a specific direction, and the right side for a fluctuation in either direction. Which to use depends on the particular situation. For example if we were performing a neutrino oscillation disappearance experiment, we would be looking for a reduction in the number of events as compared with the no-oscillation scenario, and hence would be interested in just the single-sided tail. In contrast searching for any deviation from the Standard Model expectation, maybe the two-sided tails would be more relevant.

12 Introduction to Lecture 2

This lecture deals with two different methods for determining parameters, least squares and likelihood, when a functional form is fitted to our data. A simple example would be straight line fitting, where the parameters are the intercept and gradient of the line. However the methods are much more general than this. Also there are other methods of extracting parameters; these include the more fundamental Bayesian and Frequentist methods, which are dealt with in Lecture 3 .

The least squares method also provides a measure of Goodness of Fit for the agreement between the theory with the best values of the parameters, and the data; this is dealt with in section 14. The

⁴However, it is often the case that such a distribution has heavier tails than the Gaussian.

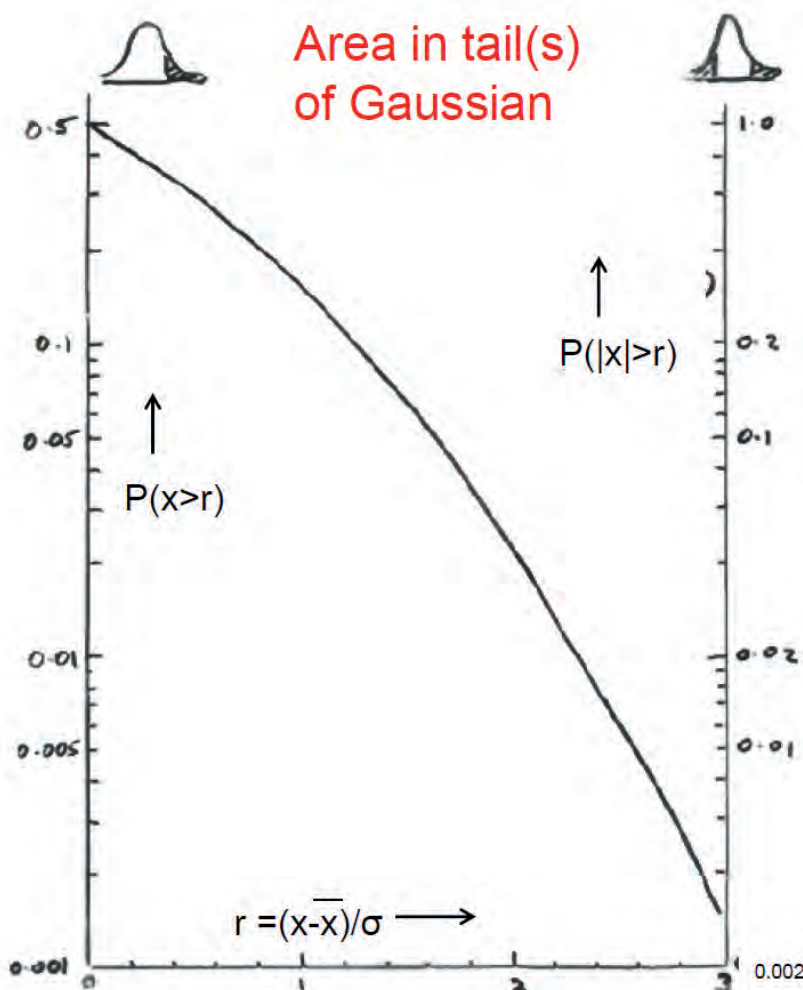


Fig. 4: The fractional area in the tail(s) of a Gaussian distribution i.e. the area with f above some specified value r , where f is the distance from the mean, measured in units of the standard deviation σ . The scale on the left refers to the one-sided tail, while that on the right is for both tails. Thus for $r = 0$, the fractional areas are $1/2$ and 1 respectively.

likelihood technique plays an important role in the Bayes approach, and likelihood ratios are relevant for choosing between two hypotheses; this is covered in Lecture 4.

13 Least squares: Basic idea

As a specific example, we will consider fitting a straight line $y = a + bx$ to some data, which consist of a series on n data points, each of which specifies $(x_i, y_i \pm \sigma_i)$ i.e. at precisely known x_i , the y co-ordinate is measured with an uncertainty σ_i . The σ_i are assumed to be uncorrelated. The more general case could involve

- a more complicated functional form than linear;
- multidimensional x and/or y ;
- correlations among the σ_i ; and
- uncertainties on the x_i values.

In Particle Physics, we often deal with a histogram of some physical quantity x (e.g. mass, angle, transverse momentum, etc.), in which case y is simply the number of counts for that x bin. Another possibility is that y and x are both physical quantities e.g. we have a two-dimensional plot showing the recession velocities of galaxies as a function their distance.

There are two statistical issues: Are our data consistent with the theory i.e. a straight line? And what are the best estimates of the parameters, the intercept and the gradient? The former is a Goodness of Fit issue, while the latter is Parameter Determination. The Goodness of Fit is more fundamental, in that if the data are not consistent with the hypothesis, the parameter values are meaningless. However, we will first consider Parameter Determination, since checking the quality of the fit requires us to use the best straight line.

The data statistic used for both questions is S , the weighted sum of squared discrepancies⁵

$$S = \sum (y_i^{th} - y_i^{obs})^2 / \sigma_i^2 = \sum (a + bx_i - y_i^{obs})^2 / \sigma_i^2 \quad (25)$$

where $y_i^{th} = a + bx_i$ is the predicted value of y at x_i , and y_i^{obs} is the observed value. In the expression for S , we regard the data $(x_i, y_i \pm \sigma_i)$ as being fixed, and the parameters a and b as being variable. If for specific values of a and b the predicted values of y and the corresponding observed ones are all close (as measured in terms of the uncertainties σ), then S will be ‘small’, while significant discrepancies result in large S . Thus, according to the least squares method, the best values of the parameters are those that minimise S , and the width of the S distribution determines their uncertainties. For a good fit, the value of S_{min} should be ‘small’. A more quantitative discussion of ‘small’ appears below.

To determine the best values of a and b , we need to set the first derivatives of S with respect to a and b both equal to zero. This leads to two simultaneous linear equations for a and b ⁶ which are readily solved, to yield

$$\begin{aligned} a &= \frac{\langle x^2 \rangle \langle y \rangle - \langle xy \rangle \langle x \rangle}{\langle x^2 \rangle - \langle x \rangle^2} \\ b &= \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2} \end{aligned} \quad (26)$$

where $\langle f \rangle = \sum (f_i / \sigma_i^2) / \sum (1 / \sigma_i^2)$ i.e it is the weighted average of the quantity inside the brackets. If the positions of the data points are such that $\langle x \rangle = 0$, then $a = \langle y \rangle$, i.e. the height of the best fit line at the weighted centre of gravity of the data points is just the weighted average of the y values.

It is also essential to calculate the uncertainties σ_a and σ_b on the parameters and their correlation coefficient $\rho = cov / (\sigma_x \sigma_y)$, where cov is their covariance. The elements of the inverse covariance matrix M are given by

$$\begin{aligned} M_{aa} &= \frac{1}{2} \frac{\partial^2 S}{\partial a^2} = \sum (1 / \sigma_i^2) \\ M_{ab} &= \frac{1}{2} \frac{\partial^2 S}{\partial a \partial b} = \sum (x_i / \sigma_i^2) \\ M_{bb} &= \frac{1}{2} \frac{\partial^2 S}{\partial b^2} = \sum (x_i^2 / \sigma_i^2) \end{aligned} \quad (27)$$

The covariance matrix is obtained by inverting M . Since the covariance is proportional to $-\langle x \rangle$, if the data are centred around $x = 0$, the uncertainties on a and b will be uncorrelated. That is one reason why track parameters are usually specified at the centre of the track, rather than at its starting point.

⁵Many people refer to this as χ^2 . I prefer S , because otherwise a discussion about whether or not χ^2 follows the mathematical χ^2 distribution sounds confusing.

⁶The derivatives are linear in the parameters, because the functional form is linear in them. This would also be true for more complicated situations such as a higher order polynomial (Yes, with respect to the coefficients, a 10^{th} order polynomial is linear), a series of inverse powers, Fourier series, etc.

13.1 Correlated uncertainties on data

So far we have considered that the uncertainties on the data are uncorrelated, but this is not always the case; correlations can arise from some common systematic. Then instead of the first equation of (25), we use

$$S = \Sigma \Sigma (y_i^{th} - y_i^{obs}) E_{ij} (y_j^{th} - y_j^{obs}) \quad (28)$$

where the double summation is over i and j , and E is the inverse covariance matrix⁷ for the uncertainties on the y_i . For the special case of uncorrelated uncertainties, the only non-zero elements of E are the diagonal ones $E_{ii} = 1/\sigma_i^2$ and then eqn. (28) reduces to (25).

This new equation for S can then be minimised to give the best values of the parameters, and S_{min} can be used in a Goodness of Fit test. As before, if y^{th} is linear in the parameters, their best estimates can be obtained by solving simultaneous linear equations, without the need for a minimisation programme.

14 Least squares for Goodness of Fit

14.1 The chi-squared distribution

It turns out that, if we repeated our experiment a large number of times, and certain conditions are satisfied, then S_{min} will follow a χ^2 distribution with $\nu = n - p$ degrees of freedom, where n is the number of data points, p is the number of free parameters in the fit, and S_{min} is the value of S for the best values of the free parameters. For example, a straight line with free intercept and gradient fitted to 12 data points would have $\nu = 10$.

The conditions for this to be true include:

- the theory is correct;
- the data are unbiased and asymptotic;
- the y_i are Gaussian distributed about their true values;
- the estimates for σ_i are correct; etc.

Useful properties to know about the mathematical χ^2 distribution are that their mean is ν and their variance is 2ν . Thus if a global fit to a lot of data has $S_{min} = 2200$ and there are 2000 degrees of freedom, we can immediately estimate that this is equivalent to a fluctuation of 3.2σ .

More useful than plots of χ^2 distributions are those of the fractional tail area beyond a particular value of χ^2 (see figs. 5 and 6 respectively). The χ^2 goodness of fit test consists of

- For the given theoretical form, find the best values of its free parameters, and hence S_{min} ;
- Determine ν from n and p ; and
- Use S_{min} and ν to obtain the tail probability p ⁸.

Then p is the probability that, if the theory is correct, by random fluctuations we would have obtained a value of S_{min} at least as large as the observed one. If this probability is smaller than some pre-defined level α , we reject the hypothesis that the model provides a good description of the data.

14.2 When $\nu \neq n - p$

If we add an extra parameter into our theoretical description, even if it is not really needed, we expect the value of S_{min} to decrease slightly. (This contrasts with including a parameter which is really relevant,

⁷We use the symbol E for the inverse covariance matrix of the measured variables y , and M for that of the output parameters (e.g. a and b for the straight line fit).

⁸If the conditions for S_{min} to follow a χ^2 distribution are satisfied, this simply involves using the tail probability of a χ^2 distribution. In other cases, it may be necessary to use Monte Carlo simulation to obtain the distribution of S_{min} ; this could be tedious.

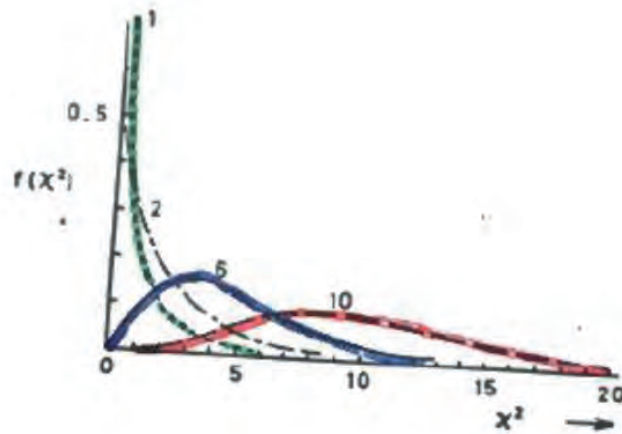


Fig. 5: Mathematical distributions of χ^2 , for different numbers of degrees of freedom ν (shown beside each curve). As ν increases, so do the mean and variance of the distribution.

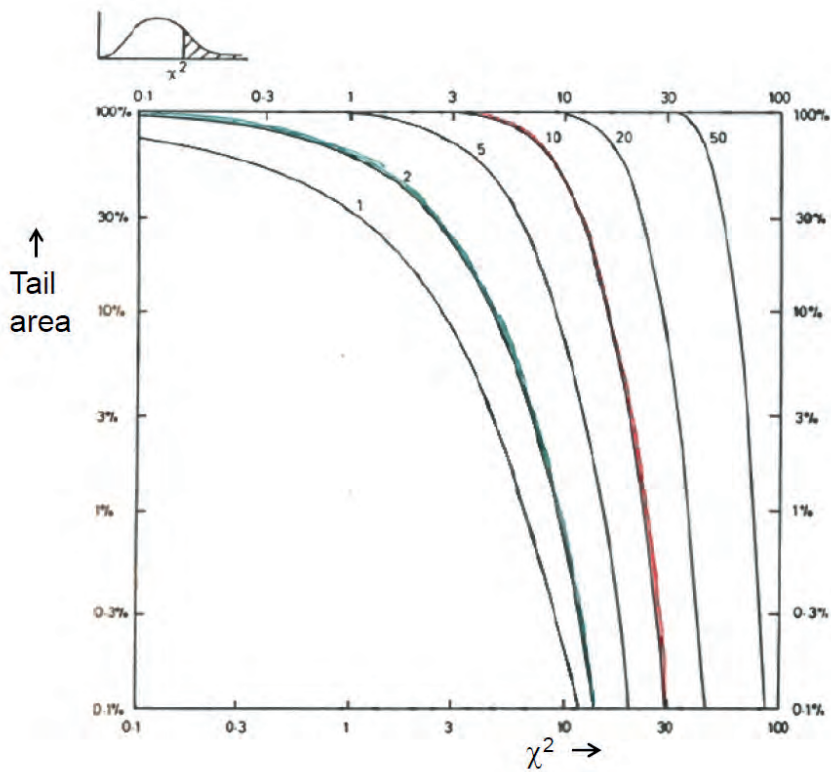


Fig. 6: The percentage area in the upper tails of χ^2 distributions, for various numbers of degrees of freedom, shown by each curve. Both scales are logarithmic. These curves bear the same relationship to those of figure 5 as does fig. 4 to the Gaussian of fig. 3, both in Lecture 1.

which can result in a dramatic reduction in S_{min} .) In determining p -values, this is allowed for by the reduction of ν . On average, a parameter which is not needed reduces S_{min} by 1. But consider the following examples.

14.2.1 Small oscillatory term

Imaging we are fitting a histogram of a variable ϕ by a distribution of the form

$$\frac{dy}{d\phi} = N[1 + 10^{-6}\cos(\phi - \phi_0)], \quad (29)$$

where the two parameters are the normalisation N and the phase ϕ_0 . Because of the factor 10^{-6} in front of the cosine term, ϕ_0 will have a miniscule effect on the prediction, and so including this as a parameter has negligible effect on S_{min} ; ϕ_0 is effectively not a free parameter.

14.2.2 Neutrino oscillations

For a scenario of two oscillating neutrino flavours, the probability P of a neutrino of energy E to remain the same flavour after a flight length L is

$$P = 1 - A\sin^2(\delta m^2 L/E) \quad (30)$$

where the two parameters are δm^2 , the difference in the mass-squareds of the two neutrino flavours, and $A = \sin^2 2\theta$ with θ being the mixing angle. However, since for small angles α , $\sin\alpha \approx \alpha$, for small $\delta m^2 L/E$ the probability P of eqn 30 is approximately $1 - A(\delta m^2 L/E)^2$. Thus the two parameters occur only as the product $A(\delta m^2)^2$, and cannot be determined separately. Thus in that regime we have effectively just a single parameter.

In both the above examples, an enormous amount of data would enable us to distinguish the small effects produced by the second parameter; hence the requirement for asymptotic conditions.

14.3 Errors of First and Second Kind

In deciding in a Goodness of Fit test whether or not to reject the null hypothesis H_0 (e.g. that the data points lie on a straight line), there are two sorts of mistake we might make:

- Error of the First Kind. This is when we reject H_0 when it is in fact true. The fraction of cases in which this happens should equal α , the cut on the p -value.
- Error of the Second Kind. This is when we do not reject H_0 , even though some other hypothesis is true. The rate at which this happens depends on how similar H_0 and the alternative hypothesis are, the relative frequencies of the two hypotheses being true, etc.

As α increases the rates of Errors of the First and Second kinds go up and down respectively. These Errors correspond to a loss of efficiency and to an increase of contamination respectively.

14.4 Other Goodness of Fit tests

The χ^2 method is by no means the only one for testing Goodness of Fit. Indeed whole books have been written on the subject [6]. Here we mention just one other, the Kolmogorov-Smirnov method (K-S), which has the advantage of working with individual observations. It thus can be used with fewer observations than are required for the binned histograms in the χ^2 approach.

A cumulative plot is produced of the fraction of events as a function of the variable of interest x . An example is shown in Fig. 7. This shows the fraction of data events with x smaller than any particular

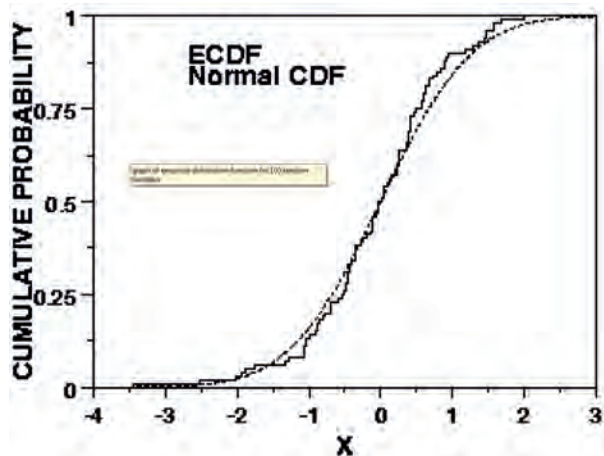


Fig. 7: Cumulative distributions for the Kolmogorov-Smirnov goodness of fit method. The stepped distribution shows the fraction of events in a data sample, while the continuous curve is that expected for a Gaussian with mean zero and unit variance. The method uses the maximum vertical separation d between the two distributions, and the number of observations, to obtain the probability of obtaining a value of d at least as large as the observed one. A small probability implies that it is unlikely that the data sample comes from the assumed distribution.

value. It is thus a stepped plot, with the fraction going from zero at the extreme left, to unity on the right hand side. Also on the plot is a curve showing the expected cumulative fraction for some theory. The K-S method makes use of the largest (as a function of x) vertical discrepancy d between the data plot and the theoretical curve. Assuming the theory is true and given the number of observations N , the probability p_{KS} of obtaining d at least as large as the observed value can be calculated. The beauty of the K-S method is that this probability is independent of the details of the theory. As in the χ^2 approach, the K-S probability gives a numerical way of checking the compatibility of theory and data. If p_{KS} is small, we are likely to reject the theory as being a good description of the data.

Some features of the K-S method are:

- The main advantage is that it can use a small number of observations.
- The calculation of the K-S probability depends on there being no adjustable parameters in the theory. If there are, it will be necessary for you to determine the expected distribution for d , presumably by Monte Carlo simulation.
- It does not extend naturally to data of more than one dimension, because of there being no unique way of producing an ordering in several dimensions.
- It is not very sensitive to deviations in the tails of distributions, which is where searches for new physics are often concentrated e.g. high mass or transverse momentum. Fortunately variants of K-S exist, which put more emphasis on discrepancies in the tails.
- Instead of comparing a data cumulative distribution with a theoretical curve, it can alternatively be compared with another distribution. This can be from a simulation of a theory, or with another data set. The latter could be to check that two data sets are compatible. The calculation of the K-S probability now requires the maximum discrepancy d , and the numbers of events N_1 and N_2 in each of the two distributions being compared.

15 Kinematic Fitting

Earlier we had the example of estimating the number of married people on an island, and saw that introducing theoretical information could improve the accuracy of our answer. Here we use the same idea

in the context of estimating the momenta and directions of objects produced in a high energy interaction. The theory we use is that energy and momentum are conserved between the initial state collision and the observed objects in the reaction.

The reaction can be either at a collider or with a stationary target. We denote it by $a+b \rightarrow c+d+e$, but the number of final state objects can be arbitrary. We assume for the time being the energy and momenta of all the objects are measured⁹.

The technique is to consider all possible configurations of the particles' kinematic variables that conserve momentum and energy, and to choose that configuration that is closest to the measured variables. The degree of closeness is defined by the weighted sum of squares of the discrepancies S , taking the uncertainties and correlations into account. If the uncertainties on the kinematic quantities m_i were uncorrelated,

$$S = \Sigma(f_i - m_i)^2 / \sigma_i^2 \quad (31)$$

where the summation is over the 4 components for all the objects in the reaction, m_i are the measured values and f_i are the corresponding fitting quantities. Because of correlations, however, this becomes

$$S = \Sigma \Sigma (f_i - m_i) E_{ij} (f_j - m_j) \quad (32)$$

where there is now a double summation over the components, and E_{ij} is the $(i, j)^{th}$ component of the inverse covariance matrix for the measured quantities¹⁰. The procedure then consists in varying f in order to minimise S , subject to the energy and momentum constraints. This usually involves Lagrange Multipliers. The result of this procedure is to produce a set of fitted values of all the kinematic quantities, which will have smaller uncertainties than the measured ones. This is an example of incorporating theory to improve the results. Thus if the objects are jets, their directions are usually quite well determined, but their energies less so. The fitting procedure enables the accurately determined jet directions to help reduce the uncertainties on the jet energies.

The fitting procedure also provides S_{min} , which is a measure of how well the best f_i agree with the m_i . In the case described, the distribution of S_{min} is approximately χ^2 with 4 degrees of freedom (because of the 4 constraints).

If S_{min} is too large, then our assumed hypothesis for the reaction may be incorrect; for example, there might have been an extra object produced in the collision that was undetected (e.g. a neutrino, or a charged particle which passed through an uninstrumented region of our detector).

Since we have 4 constraint equations, we can also allow for up to 4 missing kinematic quantities. Examples include an undetected neutrino in the final state (3 unmeasured momentum components), a wide-band neutrino beam of known direction (1 missing variable), etc. With m missing variables in an interaction involving a single vertex, S_{min} should have a χ^2 distribution with $4 - m$ degrees of freedom.

Kinematic fitting can be extended to more complicated event topologies including production and decay vertices, reactions involving particles of well known mass which decay promptly (e.g. $\psi \rightarrow \mu^+ \mu^-$), etc.

15.1 Example of a simplified kinematic fit

Consider a non-relativistic elastic scattering of two equal mass objects, for example a slow anti-proton hitting a stationary proton. For simplicity, the measured angles $\theta_1^m \pm \sigma$ and $\theta_2^m \pm \sigma$ that the outgoing particles make with the direction of travel of the incident anti-proton are assumed to have the same uncorrelated uncertainties σ . As a result of energy and momentum conservation, the angles must satisfy the constraint

$$\theta_1^t + \theta_2^t = \pi/2 \quad (33)$$

⁹For objects like charged particles whose momenta are determined from their trajectories in a magnetic field, the energy is determined from the momentum by using the relevant particle mass.

¹⁰The main correlations are among the 4 components of a single object, rather than between different objects.

where the superscript t denotes the true value. There are 3 further constraints but for simplicity we shall ignore them.

To find our best estimates of θ_1^t and θ_2^t , we must minimise

$$S = (\theta_1^t - \theta_1^m)^2/\sigma^2 + (\theta_2^t - \theta_2^m)^2/\sigma^2 \quad (34)$$

subject to the constraint 33. By using Lagrange Multipliers or by eliminating θ_2^t and then minimising S , this yields

$$\begin{aligned} \theta_1^t &= \theta_1^m + 0.5 * (\pi/2 - \theta_1^m - \theta_2^m) \\ \theta_2^t &= \theta_2^m + 0.5 * (\pi/2 - \theta_1^m - \theta_2^m) \end{aligned} \quad (35)$$

That is, the best estimate of each true value is obtained by adding to the corresponding measured value half the amount by which the measured values fail to satisfy the constraint 33.

The uncertainties on the fitted estimates of the angles are easily obtained by propagation of the uncertainties σ on the measured angles via eqns. 35, and are both equal to $\sigma/\sqrt{2}$.

We thus have an example of the promised outcome that kinematic fitting improves the accuracy of our measurements. The factor of $\sqrt{2}$ improvement can easily be understood in that we have two independent estimates of θ_1^t , the first being the original measurement θ_1^m , and the other coming from the measurement θ_2^m via the constraint 33. However, even with uncorrelated uncertainties on the measured angles, the fitted ones would be anti-correlated.

16 THE paradox

I refer to this as ‘THE’ paradox as, in various forms, it is the basis of the most frequently asked question.

You have a histogram of 100 bins containing some data, and use this to determine the best value μ_0 of a parameter μ by the χ^2 method. It turns out that $S_{min} = 87$, which is reasonable as the expected value for a χ^2 with 99 degrees of freedom is 99 ± 14 . A theorist asks whether his predicted value μ_{th} is consistent with your data, so you calculate $S(\mu_{th}) = 112$. The theorist is happy because this is within the expected range. But you point out that the uncertainty in μ is calculated by finding where S increases by 1 unit from its minimum. Since 112 is 25 units larger than 87, this is equivalent to a 5 standard deviation discrepancy, and so you rule out the theorist’s value of μ .

Deciding which viewpoint is correct is left as an exercise for the reader.

17 Likelihood

The likelihood function is very widely used in many statistics applications. In this Section, we consider it just for Parameter Determination. An important feature of the likelihood approach is that it can be used with **unbinned** data, and hence can be applied in situations where there are not enough individual observations to construct a histogram for the χ^2 approach.

We start by assuming that we wish to fit our data x , using a model $f(x; \mu)$ which has one or more free parameters μ , whose value(s) we need to determine. The function f is known as the ‘probability distribution’ (*pdf*) and specifies the probability (or probability density, for the data having continuous as opposed to discrete values) for obtaining different values of the data, when the parameter(s) are specified. Without this it is impossible to apply the likelihood (or many other) approaches. For example x could be observations of a variable of interest within some range, and f could be any function such as a straight line, with gradient and intercept as parameters. But we will start with an angular distribution

$$y(\cos \theta; \beta) = \frac{dp}{d \cos \theta} = N(1 + \beta \cos^2 \theta) \quad (36)$$

Here θ is the angle at which a particle is observed, $dp/d \cos \theta$ is the *pdf* specifying the probability density for observing a decay at any $\cos \theta$, β is the parameter we want to determine, and N is the crucial

normalisation factor which ensures that the probability of observing a given decay at any $\cos \theta$ in the whole range from -1 to $+1$ is unity. In this case $N = 1/(2(1 + \beta/3))$. The data consists of N decays, with their individual observations $\cos \theta_i$.

Assuming temporarily that the value of the parameter β is specified, the probability density y_1 of observing the first decay at $\cos \theta_1$ is

$$y_1 = N(1 + \beta \cos^2 \theta_1) = 0.5(1 + \beta \cos^2 \theta_1)/(1 + \beta/3), \quad (37)$$

and similarly for the rest of the N observations. Since the individual observations are independent, the overall probability $P(\beta)$ of observing the complete data set of N events is given by the product of the individual probabilities

$$P(\beta) = \prod y_i = \prod 0.5(1 + \beta \cos^2 \theta_i)/(1 + \beta/3) \quad (38)$$

We imagine that this is computed for all values of the parameter β ; then this is known as the likelihood function $L(\beta)$.

The likelihood method then takes as the estimate of β that value which maximises the likelihood. That is, it is the value which maximises (with respect to β) the probability density of observing the given data set. Conversely we rule out values of β for which $L(\beta)$ is very small. The uncertainty on β is related to the width of the $L(\beta)$ distribution (see later).

It is often convenient to consider the logarithm of the likelihood

$$l = \ln L = \sum \ln y_i \quad (39)$$

One reason for this is that, for a large number of observations some fraction could have small y_i . Then the likelihood, involving the product of the y_i , could be very small and may underflow the computer's range for real numbers. In contrast, l involves a sum rather than a product, and $\ln y_i$ rather than y_i , and so produces a gentler number.

17.1 Likelihood and pdf

The procedure for constructing the likelihood is first to write down the *pdf*, and then to insert into that expression the observed data values in order to evaluate their product, which is the likelihood. Thus both the *pdf* and the likelihood involve the data x and the parameter(s) μ . The difference is that the *pdf* is a function of x for fixed values of μ , while the likelihood is a function of μ given the fixed observed data x_{obs} .

Thus for a Poisson distribution, the probability of observing n events when the rate μ is specified is

$$P(n; \mu) = e^{-\mu} \mu^n / n! \quad (40)$$

and is a function of n , while the likelihood is

$$L(\mu; n) = e^{-\mu} \mu^n / n! \quad (41)$$

and is a function of μ for the fixed observed number n .

17.2 Intuitive example: Location and width of peak

We consider a situation where we are studying a resonant state which would result in a bump in the mass distribution of its decay particles. We assume that the bump can be parametrised as a simple Breit-Wigner

$$y(m; M_0, \Gamma) = \frac{\Gamma/(2\pi)}{(m - M_0)^2 + (\Gamma/2)^2} \quad (42)$$

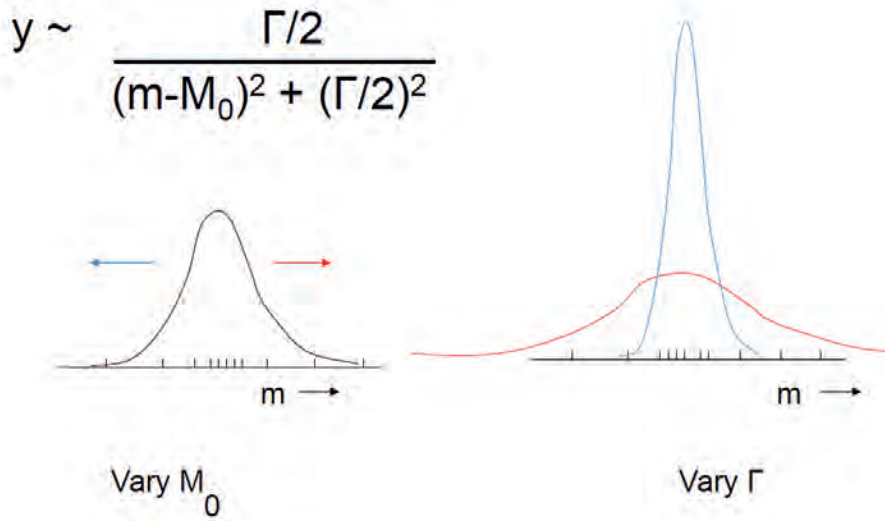


Fig. 8: A visual demonstration of how the maximum likelihood method gives sensible values for the parameters, the position and width of a resonance. The bars along the m -axis represent the experimental measurements of a set of mass values m_i , which are to be fitted by a simple Breit-Wigner resonance shape. In (a), the width Γ of the resonance is kept fixed, and the mass parameter M_0 is varied. This has the effect of sliding the curve to the left or right along the m -axis, without changing its shape or height. To calculate the likelihood for a given position of the curve we multiply all the $y(m_i)$ values; i.e. the height of the curve at each observed mass. The best value of M_0 is thus equivalent to finding the best location of the curve in order to maximise this product. Clearly we need to locate the peak near where most of the data values are. In (b), we regard M_0 as constant, but vary the width. The effect of the normalisation condition then means that the wider curve will have lower peak height and vice versa. The narrow curve suffers because of the very small y values for the extreme observed mass values, while wide curves do not benefit so much from the concentration of masses around the central value. The best value of Γ is the result of a compromise between these two effects.

where y is the probability density of obtaining a mass m if the location and width the state are M_0 and Γ , the parameters we want to determine. It is essential that y is normalised, i.e. its integral over all physical values of m is unity; hence the normalisation factor of $\Gamma/(2\pi)$. The data consists of n observations of m , as shown in fig. 8.

Assume for the moment that we know M_0 and Γ . Then the probability density for observing the i^{th} event with mass m_i is

$$y_i(M_0, \Gamma) = \frac{\Gamma/(2\pi)}{(m_i - M_0)^2 + (\Gamma/2)^2} \quad (43)$$

Since the events are independent, the probability density for observing the whole data sample is

$$y_{\text{all}}(M_0, \Gamma) = \prod \frac{\Gamma/(2\pi)}{(m_i - M_0)^2 + (\Gamma/2)^2} \quad (44)$$

and this is known as the likelihood $L(M_0, \Gamma)$. Then the best values for the parameters are taken as the combination that maximises the probability density for the whole data sample i.e. $L(M_0, \Gamma)$. Parameter values for which L is very small compared to its maximum value are rejected, and the uncertainties on the parameters are related to the width of the distribution of L ; we will be more specific later.

The curve in fig. 8(left) shows the expected probability distribution for fixed parameter values. The way L is calculated involves multiplying the heights of the curve at all the observed m_i values. If we now consider varying M_0 , this moves the curve bodily to the left or right without changing its shape

or normalisation. So to determine the best value of M_0 , we need to find where to locate the curve so that the product of the heights is a maximum; it is plausible that the peak will be located where the majority of events are to be found.

Now we will consider how the optimum value of Γ is obtained. A small Γ results in a narrow curve, so the masses in the tail will make an even smaller contribution to the product in eqn. 44, and hence reduce the likelihood. But a large Γ is not good, because not only is the width larger, but because of the normalisation condition, the peak height is reduced, and so the observations in the peak region make a smaller contribution to the likelihood. The optimal Γ involves a trade-off between these two effects.

Of course, in finding the optimal of values of the two parameters, in general it is necessary to find the maximum of the likelihood as a function of the two parameters, rather than maximising with respect to just one, and then with respect to the other and then stopping (see section 17.5).

17.3 Uncertainty on parameter

With a large amount of data, the likelihood as a function of a parameter μ is often approximately Gaussian. In that case, l is an upturned parabola. Then the following definitions of σ_μ , the uncertainty on μ_{best} , yield identical answers:

- The RMS of the likelihood distribution.
- $[-\frac{d^2l}{d\mu^2}]^{-1/2}$. If you remember that the second derivative of the log likelihood function is involved because it controls the width of the l distribution, a mnemonic helps you remember the formula for σ_μ : Since σ_μ must have the same units as μ , the second derivative must appear to the power $-1/2$. But because the log of the likelihood has a maximum, the second derivative is negative, so the minus sign is necessary before we take the square root.
- It is the distance in μ from the maximum in order to decrease l by half a unit from its maximum value. i.e.

$$l(\mu_{best} + \sigma_\mu) = l_{max} - 0.5 \quad (45)$$

In situations where the likelihood is not Gaussian in shape, these three definitions no longer agree. The third one is most commonly used in that case. Now the upper and lower ends of the intervals can be asymmetric with respect to the central value. It is a mistake to believe that this method provides intervals which have a 68% chance of containing the true value of the parameter¹¹.

Symmetric uncertainties are easier to work with than asymmetric ones. It is thus sometimes better to quote the uncertainty on a function of the first variable you think of. For example, for a charged particle in a magnetic field, the reciprocal of the momentum has a nearly symmetric uncertainty. Especially for high momentum tracks, the upper uncertainty on the momentum can be much larger than the lower one e.g. $1.0^{+1.5}_{-0.4}$ TeV.

17.4 Coverage

An important feature of any statistical method for estimating a range for some parameter μ at a specified confidence level α is its coverage C . If the procedure is applied many times, these ranges will vary because of statistical fluctuations in the observed data. Then C is defined as the fraction of ranges which contain the true value μ_{true} ; it can vary with μ_{true} .

It is very important to realise that coverage is a property of the **statistical procedure** and does not apply to your particular measurement. An ideal plot of coverage as a function of μ would have C constant at its nominal value α . For a Poisson counting experiment, figure 9 shows C as a function of the Poisson parameter μ , when the observed number of counts n is used to determine a range for μ via

¹¹Unfortunately, this incorrect statement occurs in my book [4]. It is corrected in a separate update [5].

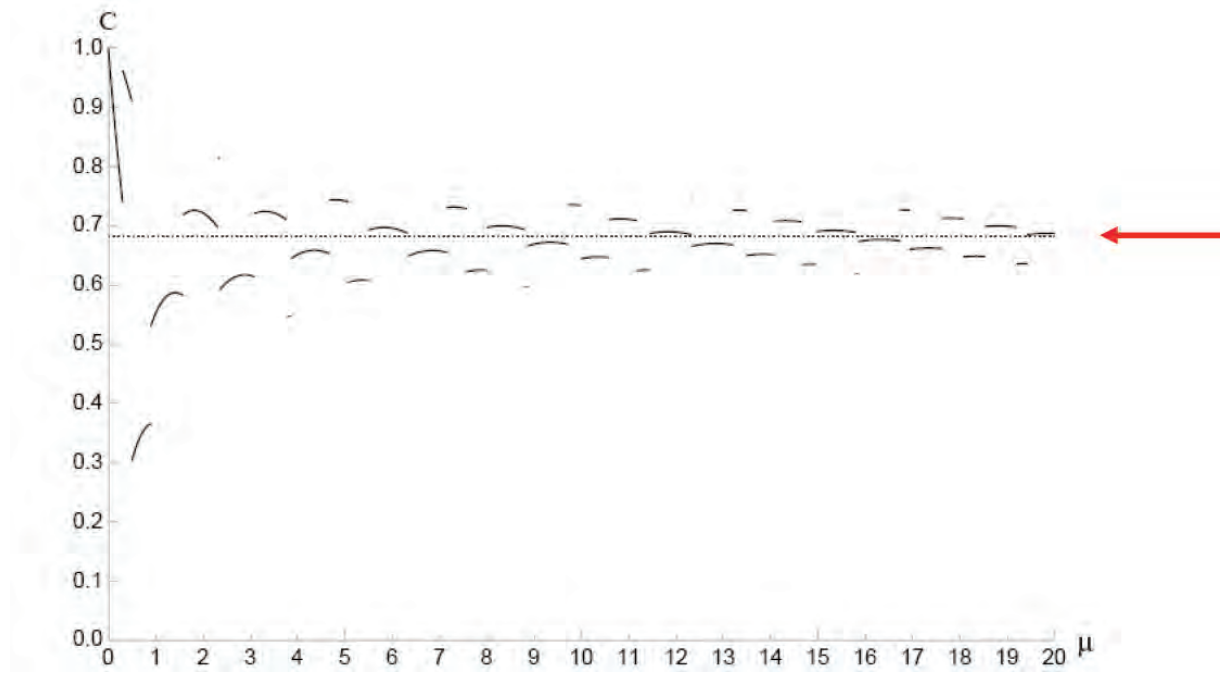


Fig. 9: Coverage C for Poisson parameter intervals, as determined by the $\Delta(\log(L)) = 0.5$ rule. Repeated trials (all using the same Poisson parameter μ) yield different values of n , each resulting in its own range μ_l to μ_u for μ ; then C is the fraction of trials that give ranges which include the chosen value of μ for the trials. The coverage C varies with μ , and has discontinuities because the data n can take on only discrete integer values. For large μ , C seems to approach the expected 0.68, shown by the arrow, but for small μ , the coverage takes on values between 30% and 100%.

the change in log-likelihood being 0.5. The coverage is far from constant at small μ . If C is smaller than α , this is known as undercoverage. Certainly frequentists would regard this as unfortunate; it means that people reading an article containing parameters determined this way are likely to place more than justified reliance on the quoted range. Methods using the Neyman construction to determine parameter ranges by construction do not have undercoverage.

Coverage involves a statement about $Prob[\mu_l \leq \mu_{true} \leq \mu_u]$. This is to be interpreted as a probability statement about how often the ranges μ_l to μ_u contain the (unknown but constant) true value μ_{true} . This is a frequentist statement; Bayesians do not want to consider the ensemble of possible results if the measurement procedure were to be repeated. Thus Bayesians would regard the statement about $Prob[\mu_l \leq \mu_{true} \leq \mu_u]$ as describing what fraction of their estimated posterior probability density for μ_{true} would be between the fixed values μ_l and μ_u , derived from their actual measurement.

17.5 More than one parameter

For the case of just one parameter μ , the likelihood best estimate $\hat{\mu}$ is given by the value of μ which maximises L . Its uncertainty σ_μ is determined either from

$$1/\sigma_\mu^2 = -d^2 \ln L/d\mu^2; \tag{46}$$

or by finding how far $\hat{\mu}$ would have to be changed in order to reduce $\ln L$ by 0.5.

When we have two or more parameters β_i the rule for finding the best estimates $\hat{\beta}_i$ is still to maximise L . For the uncertainties and their correlations, the generalisation of equation 46 is to construct

the inverse covariance matrix \mathbf{M} , whose elements are given by

$$M_{ij} = -\frac{\partial^2 \ln L}{\partial \beta_i \partial \beta_j} \quad (47)$$

Then the inverse of \mathbf{M} is the covariance matrix, whose diagonal elements are the variances of β_i , and whose off-diagonal ones are the covariances.

Alternatively (and more common in practice), the uncertainty on a specific β_j can be obtained by using the profile likelihood $L_{prof}(\beta_j)$. This is the likelihood as a function of the specific β_j , where for each value of β_j , L has been remaximised with respect to all the other β . Then $L_{prof}(\beta_j)$ is used with the ‘reduce $\ln L_{prof} = 0.5$ ’ rule to obtain the uncertainty on β_j . This is equivalent to determining the contour in β -space where $\ln L = \ln L_{max} - 0.5$, and finding the values $\beta_{j,1}$ and $\beta_{j,2}$ on the contour which are furthest from $\hat{\beta}_j$. Then the (probably asymmetric) upper and lower uncertainties on β_j are given by $\beta_{j,2} - \hat{\beta}_j$ and $\hat{\beta}_j - \beta_{j,1}$ respectively.

Because these are likelihood methods of obtaining the intervals, these estimates of uncertainties provide only **nominal** regions of 68% coverage for each parameter; the **actual** coverage can differ from this. Furthermore, the region within the contour described in the previous paragraph for the multidimensional β space will have less than 68% nominal coverage. To achieve that, the ‘0.5’ in the rule for how much $\ln L$ has to be reduced from its maximum must be replaced by a larger number, whose value depends on the dimensionality of β .

18 Worked example: Lifetime determination

Here we consider an experiment which has resulted in N observed decay times t_i of a particle whose lifetime τ we want to determine. The probability density for observing a decay at time t is

$$p(t; \tau) = (1/\tau) e^{-t/\tau} \quad (48)$$

Note the essential normalisation factor $1/\tau$; without this the likelihood method does not work.

It should be realised that realistic situations are more complicated than this. For example, we ignore the possibility of backgrounds, time resolution which smears the expected values of t , acceptance or efficiency effects which vary with t , etc., but this enables us to estimate τ and its uncertainty σ_τ analytically. In real practical cases, it is almost always necessary to calculate the likelihood as a function of τ numerically.

From equation 48 we calculate the log-likelihood as

$$\ln L(\tau) = \ln[\prod (1/\tau) e^{-t_i/\tau}] = \sum (-\ln \tau - t_i/\tau) \quad (49)$$

Differentiating $\ln L(\tau)$ with respect to τ and setting the derivative to zero then yields

$$\tau = \sum t_i / N \quad (50)$$

This equation has an appealing feature, as it can be read as “The mean lifetime is equal to the mean lifetime”, which sounds as if it must be true. However, what it really says is not quite so trivial: “Our best estimate of the lifetime parameter τ is equal to the mean of the N observed decay times in our experiment.”

We next calculate σ_τ from the second derivative of $\ln L$, and obtain

$$\sigma_\tau = \tau / \sqrt{N} \quad (51)$$

This exhibits a common feature that the uncertainty of our parameter estimate decreases as $1/\sqrt{N}$ as we collect more and more data. However, a potential problem arises from the fact that our estimated

uncertainty is proportional to our estimate of the parameter. This is relevant if we are trying to combine different experimental results on the lifetime of a particle. For combining procedures which weight each result by $1/\sigma^2$, a measurement where the fluctuations in the observed times result in a low estimate of τ will tend to be over-weighted (compare the section on ‘Combining Experiments’ in Lecture 1), and so the weighted average would be biased downwards. This shows that it is better to combine different experiments at the data level, rather than simply trying to use their results.

One final point to note about our simplified example is that the likelihood $L(\tau)$ depends on the observations only via the **sum** of the times $\sum t_i$ i.e. their **distribution** is irrelevant. Thus the likelihood distributions for two experiments having the same number of events and the same sum of observed decay times, but with one having the decay times consistent with an exponential distribution and the other having something completely different (e.g. all decays occur at the same time), would have identical likelihood functions. This provides an example of the fact that the unbinned likelihood function does not in general provide useful information on Goodness of Fit.

19 Conclusions

Just as it is impossible to learn to play the violin without ever picking it up and spending hours actually using it, it is important to realise that one does not learn how to apply Statistics merely by listening to lectures. It is really important to work through examples and actual analyses, and to discover more about the topics.

There are many resources that are available to help you. First there are textbooks written by Particle Physicists [8], which address the statistical problems that occur in Particle Physics, and which use a language which is easier for other Particle Physicists to understand.

The large experimental collaborations have Statistics Committees, whose web-sites contain lots of useful statistical information. That of CDF [9] is most accessible to Physicists from other experiments.

The Particle Data Book [10] contains short sections on Probability, Statistics and Monte Carlo simulation. These are concise, and are useful reminders of things you already know. It is harder to use them instead of lengthier articles and textbooks in order to understand a new topic.

If in the course of an analysis you come upon some interesting statistical problem that you do not immediately know how to solve, you might be tempted to invent your own method of how to overcome the problem. This can amount to reinventing the wheel. It is a good idea to try to see if statisticians (or even Particle Physicists) have already dealt with this topic, as it is far preferable to use their circular wheels, rather than your own hexagonal one.

Finally I wish you the best of luck with the statistical analyses of your data.

References

- [1] 'Bayes and Frequentism: The return of an old controversy', Louis Lyons, under Recommendations section on CDF statistics committee page <https://www-cdf.fnal.gov/physics/statistics/> (2002)
- [2] 'Statistical issues in searches for New Physics', Louis Lyons, under Notes section on CDF statistics committee page <https://www-cdf.fnal.gov/physics/statistics/> (2002)
- [3] 'Learning to Love the Error Matrix'. A video and the slides of an earlier version of the lecture can be found at http://vmsstreamer1.fnal.gov/VMS_Site_03/Lectures/AcademicLectures/presentations/040803Lyons.pdf.
- [4] Louis Lyons, 'Statistics for nuclear and particle physicists', Cambridge University Press (1986).
- [5] Louis Lyons, 'Statistics for Nuclear and Particle Physicists: an Update', <http://www-cdf.fnal.gov/physics/statistics/notes/Errata2.pdf> (2009).
- [6] Ralph B. D'Agostino and Michael A. Stephens, 'Goodness of Fit Techniques', CRC Press (1986).
- [7] Joel G. Heinrich, 'Coverage of Error Bars for Poisson Data', http://www-cdf.fnal.gov/physics/statistics/notes/cdf6438_coverage.pdf.
- [8] R.J. Barlow 'Statistics' (Wiley,1989);
 O. Behnke et al. (editors), 'Data Analysis in High Energy Physics: a Practical Guide to Statistical Methods' (Wiley. 2013);
 G. Cowan, 'Statistical Data Analysis' (OUP, 1998);
 F. James, 'Statistical Methods in Experimental Physics', (World Scientific, 2006);
 L. Lista, 'Statistical Methods for Data Analysis in Particle Physics', (Springer, 2015);
 L. Lyons, 'Statistics for Nuclear and Particle Physicists' (CUP, 1986);
 B. Roe, 'Probability and Statistics in Experimental Physics' (Springer, 1992)
- [9] CDF Statistics Committee web-site, <https://www-cdf.fnal.gov/physics/statistics/>
- [10] G. Cowan in Particle Data Group - 2016 Review', <http://pdg.lbl.gov/2016/reviews/rpp2016-rev-statistics.pdf>;
<http://pdg.lbl.gov/2016/reviews/rpp2016-rev-probability.pdf>; and
<http://pdg.lbl.gov/2016/reviews/rpp2016-rev-monte-carlo-techniques.pdf>

Organizing Committee

M. Aguilar (CIEMAT, Spain)
L. Álvarez-Gaumé (CERN)
E. Carrera (USFQ, Ecuador)
C. Dib (UTFSM, Chile)
M.T. Dova (UNLP, Argentina)
J. Ellis (King's College London, UK and CERN)
N. Ellis (Schools Director, CERN (Chair))
A. Gago (PUCP, Peru)
M. Gandelman (UFRJ, Brazil)
P. Garcia (CIEMAT, Spain)
M. Losada (UAN, Colombia)
M. Mulders (Schools Deputy-Director, CERN)
G. Perez (CERN and Weizmann Institute, Israel and CERN)
K. Ross (Schools Administrator, CERN)
G. Zanderighi (CERN)
A. Zepeda (Cinvestav, Mexico)

Local Organizing Committee

A. Ayala (USFQ, Ecuador)
E. Ayala (EPN, Ecuador)
E. Carrera (USFQ, Ecuador (Chair))
J. Grefa (USFQ, Ecuador)
S. Guitarra (USFQ, Ecuador)
C. Mantilla (EPN, Ecuador)
N. Vasquez (EPN, Ecuador)

Lecturers

X. Bertou (Centro Atómico Bariloche, CNEA/CONICET, Argentina)
D. de Florian (UBA, Argentina)
E.S. Fraga (UFRJ, Brazil)
B. Grinstein (UCSD, USA)
C. Grojean (ICREA/IFAE, Spain)
P. Hernandez (U. Valencia, Spain)
J. Iliopoulos (ENS Paris, France)
P. Jenni (U. Freiburg, Germany and CERN)
K. Kunze (U. Salamanca, Spain)
L. Lyons (Imperial College London and U. Oxford, UK)
R. Rosenfeld (IFT-UNESP, Brazil)

Discussion Leaders

L. Da Rold (Centro Atómico Bariloche, CONICET/IB, Argentina)
F. Febres Cordero (Simon Bolivar U., Venezuela)
B. Hoeneisen (USFQ, Ecuador)
J. Jones (PUCP, Peru)
M.E. Tejeda Yeomans (U. de Sonora, Mexico)

Students

Fábio ALVES
Gabriela BAILAS
Andres BAQUERO
Arianna BATISTA CAMEJO
Rocio BERMUDEZ
Armando BERMÚDEZ MARTÍNEZ
Cesar BERNARDES
Sylvain BLUNIER
Christopher BOCK
Jaime CALDERÓN
Liliet CALERO DIAZ
Blanca Cecilia CAÑAS ORDUZ
Juan Pablo CARLOMAGNO
Jose Alonso CARPIO DUMLER
Jorge David CASTAÑO YEPES
Luisa Fernanda CHAPARRO SIERRA
Alvaro CHAVARRIA
Henrique Jose CORREIA ZANOLI
Melissa Maria CRUZ TORRES
Jose Camilo DÍAZ BUSTAMANTE
Félix Napoleón DÍAZ DESPOSORIO
Daniela DOMÍNGUEZ DAMIANI
Daniel EVANGELHO VIEIRA
Rebecca Jane FALLA
Nerina Andrea FIDANZA ROMANI
Jorge FRAGA
Liam David Joseph FULTON
Francisca GARAY WALLS
Oscar GARCÍA
Joaquin GREFA
Barak Raimond GRUBERG CAZON
Daniel Fernando GUERRERO IBARRA
Adolfo GUEVARA
Gerardo HERNÁNDEZ-TOMÉ
David HERVAS AGUILAR
Fabricio Andrés JIMÉNEZ MORALES
Pawel KLIMEK
Ilya KOMAROV
Julio Cesar LEAL REINA
Xin Ran LIU
Robinson LONGAS
Jorge Andres LOPEZ LOPEZ
Cristina MANTILLA SUAREZ
Luis MANZANILLAS
Javier MAZZITELLI
Miqueias MELO DE ALMEIDA
Mauro MENDIZÁBAL
Roberto Anibal MORALES VALBUENA
Thiago MUHLBEIER
Jorge NOGUEIRA
Arturo NUNEZ
Sebastian Andres OLIVARES PINO
Luis OTINIANO
Santiago Rafael PAREDES SAENZ
Eduardo QUINTANA MIRANDA
Raquel QUISHPE
Gustavo Alonso RAMÍREZ HIDALGO
Omar Alberto ROLDAN GARCIA
David ROMERO ABAD
Manfredi RONZANI
Yamila ROTSTEIN HABARNAU
Christian SARMIENTO
Elisabeth SCHOPF
Oscar TABORDA
Sebastian TAPIA ARAYA
Julio TORRES
Stephany VARGAS
Yee Chinn YAP

Posters

Poster title	Presenter
Non-Standard Supersymmetry Scenarios at the LHC	ALVES, F.
Associated production of γ bosons with charm quark jets at the LHC in presence of intrinsic charm in the proton	BAILAS, G.
Study of J/ψ polarization with the ALICE experiment at the LHC	BATISTA CAMEJO, A.; ALICE COLLABORATION
Parton distribution amplitudes of light mesons in a contact interaction	BERMUDEZ, R.; COBOS-MARTINEZ, J.J.; TEJEDA-YEOMANS, M.E.
Search for narrow high-mass resonances decaying to Z and Higgs boson	BERNARDES, C.
Meson phenomenology and phase transitions in nonlocal chiral quark models	CARLOMAGNO, J.P.; GÓMEZ DUMM, D.; SCOCCOLA, N.N.
Neutrino-UHECR correlations considering magnetic deflections	CARPIO, J.A.; GAGO, A.M.
Search for a Higgs boson decaying to $\gamma\gamma^* \rightarrow \gamma\ell^+\ell^-$ in pp collisions at $\sqrt{s}=8$ TeV at the CMS Experiment	CHAPARRO, L. F.; CMS COLLABORATION
DAMIC - Dark Matter in CCDs	CHAVARRIA, A.; DAMIC COLLABORATION
Measurements of CP Violation in the phase space of charmless B decays	CRUZ, M.; LHCb COLLABORATION

Poster title	Presenter
Chargino Pair Same Sign Production via Vector Boson Fusion in ATLAS	KLIMEK, P.
Study of the rare B_s^0 and B^0 decays into the $\pi^+\pi^-\mu^+\mu^-$ final state	KOMAROV, I.; DORIGO M.; LHCb COLLABORATION
UK low-background infrastructure for delivering SuperNEMO	LIU, X.
Electroweak precision test and lepton flavor violation constraints on the inert Zee model	LONGAS, R.; PORTILLO, D.; RESTREPO, D.; ZAPATA, O.
The STEREO experiment: Search for sterile neutrinos at ILL	MANZILLAS, L.
CP Violation in three-body B^\pm decay with final state interaction and CPT constraint	NOGUEIRA, J.; FREDERICO, T.; LOURENÇO, O.; BODIAGA, I.
Effective mass and gyromagnetic factor of fermions in a magnetic field	ROLDAN, O.; FRAGA, E.; KODAMA, T.
Search for a high-mass resonance decaying into a pair of Z vector bosons in the MET-plus-Jet final state	ROMERO, D.
Large scale distribution of ultra high energy cosmic rays detected at the Pierre Auger observatory	TABORDA, O.; PIERRE AUGER COLLABORATION