# Proceedings of the 2018 European School of High-Energy Physics
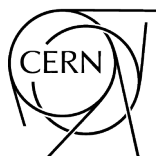
Maratea, Italy, 20 June–3 July 2018

Editors:

M. Mulders

C. Duhr

This report should be cited as:

Proceedings of the 2018 European School of High-Energy Physics, Maratea, Italy, 20 June–3 July 2018,
edited by M. Mulders and C. Duhr CERN-2019-006 (CERN, Geneva, 2019),
http://doi.org/10.23730/CYRSP-2019-006

A contribution in this report should be cited as:

[Author name(s)], in Proceedings of the 2018 European School of High-Energy Physics, Maratea, Italy,
20 June – 3 July 2018, edited by M. Mulders and C. Duhr, CERN-2019-006 (CERN, Geneva, 2019), pp. [first
page]–[last page], http://doi.org/10.23730/CYRSP-2019-006.[first page]

# Abstract

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

# Preface

The twenty-sixth event in the series of the European School of High-Energy Physics took place in Maratea, Italy, from 20 June to 3 July 2018. It was organized jointly by CERN, Geneva, Switzerland, and JINR, Dubna, Russia, with support from INFN, in particular the Naples Unit, the University of Naples Federico II, in particular the Department of Physics E. Pancini, the University of Basilicata and the Italian Physical Society (SIF). The local organization team was chaired by Luca Lista (INFN Naples). The other members of the local committee were: Maria Grazia Alviggi, Fabrizio Bianchi, Concezio Bozzi, Gianpaolo Carlino, Nicola Cavallo, Francesco Cirotto, Antonia Di Crescenzo, Francesco Fabozzi, Orso Iorio, Mario Merola, Celeste Satriano and Stefania Spagnolo.

A total of 96 students of 37 different nationalities attended the school, mainly from institutes in member states of CERN and/or JINR, but also some from other regions. The participants were generally students in experimental High-Energy Physics in the final years of work towards their PhDs.

The School was hosted at the Grand Hotel Pianeta Maratea, on the gulf of Policastro, south of Naples. According to the tradition of the school, the students shared twin rooms mixing participants of different nationalities.

A total of 30 lectures were complemented by daily discussion sessions led by six discussion leaders. The students displayed their own research work in the form of posters in an evening session in the first week, and the posters stayed on display until the end of the School. The full scientific programme was arranged in the on-site conference facilities. The School also included an element of outreach training, complementing the main scientific programme. This consisted of a two-part course from the Inside Edge media training company. Additionally, students had the opportunity to act out radio interviews under realistic conditions based on a hypothetical scenario.

The students from each discussion group subsequently carried out a collaborative project, preparing a talk on a physics-related topic at a level appropriate for a general audience. The talks were given by student representatives of each group in an evening session in the second week of the School. A jury, chaired by Leonardo Alfonsi (PsiQuadro), judged the presentations; other members of the jury were Andrea De Simone (SISSA), and Fabio Maltoni (Louvain University). We are very grateful to all of these people for their help.

Our thanks go to the local-organization team and, in particular, to Luca Lista, for all of their work and assistance in preparing the School, on both scientific and practical matters, and for their presence throughout the event. Our thanks also go to the efficient and friendly hotel management and staff who assisted the School organizers and the participants in many ways. Very great thanks are due to all of the lecturers and discussion leaders for their active participation in the School and for making the scientific programme so stimulating. The students, who in turn manifested their good spirits during two intense weeks, appreciated listening to and discussing with the teaching staff of world renown.

We would like to express our strong appreciation to Fabiola Gianotti, Director General of CERN, and Victor Matveev, Director of JINR, for their lectures on the scientific programmes of the two organizations and for discussing with the School participants. We are also very happy that Nadia Pastrone, President of the INFN National Scientific Committee on Particle Physics, was able to attend the School as one of the lecturers.

In addition to the rich academic programme, the participants enjoyed numerous sports, leisure and cultural activities in and around Maratea. We were able to try local wines at a pre-dinner wine tasting, offered by the Gioia Al Negro local vineyard. There was a half-day excursion to the local town of Sapri, followed by a visit to the beautiful Villa Nitti for a buffet dinner. A boat trip along the scenic Cilento coast and the protected marine

area of Infreschi was the main attraction of the full day excursion, with opportunities to visit beaches reachable by boat only and a stop in the village of Scario. On the final Saturday afternoon, the students were able to make use of the hotel facilities during free time finishing with a dinner organized by the hotel on the beach. The excursions provided an excellent environment for informal interactions between staff and students.

We are very grateful to the School Administrators, Kate Ross (CERN) and Tatyana Donskova (JINR), for their untiring efforts in the lengthy preparations for and the day-to-day operation of the School. Their continuous care of the participants and their needs during the School was highly appreciated.

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

Nick Ellis
(On behalf of the Organizing Committee)

photo credit: Angela Sorrentino

photo credit: Angela Sorrentino

# People in the photograph

| | | | |
|---|---|---|---|
| 1 | Alberto Orso Maria Iorio | 55 | Anna Benecke |
| 2 | Alessandro Lapertosa | 56 | Vasilisa Lenivenko |
| 3 | Francesco Tramontano | 57 | Natascia Vignaroli |
| 4 | Cheng Chen | 58 | Ekaterina Govorkova |
| 5 | Martijn Mulders | 59 | Renat Sadykov |
| 6 | Basile Vermassen | 60 | Ruth Magdalena Jacobs |
| 7 | Aishik Ghosh | 61 | Fang-Ying Tsai |
| 8 | Ivan Lilienberg | 62 | Eleni Vryonidou |
| 9 | Mohamed Darwish | 63 | Thu Joni Pham |
| 10 | Emmanouil Vourliotis | 64 | Michael Wassmer |
| 11 | Kate Ross | 65 | Simona Gargiulo |
| 12 | Afiq Anuar | 66 | Daniele Marangotto |
| 13 | Federico Betti | 67 | Julian Wollrath |
| 14 | Samuel Belin | 68 | Konie Al Khoury |
| 15 | Lais Soares Lavra | 69 | Anum Qureshi |
| 16 | Francesco Giuseppe Gravili | 70 | Hadar Cohen |
| 17 | Daniel Spitzbart | 71 | Daniel Stefaniak Nielsen |
| 18 | Milene Calvetti | 72 | Francesco Armando Di Bello |
| 19 | Diego Ramirez Garcia | 73 | Violaine Bellee |
| 20 | Francesco La Ruffa | 74 | Merve Sahinsoy |
| 21 | Roman Litvinov | 75 | Jonathan Shlomi |
| 22 | Nicolas Tonon | 76 | Aleksey Pavlov |
| 23 | Etienne Dreyer | 77 | Archil Durglishvili |
| 24 | Olesia Kuchinskaya | 78 | Ece Akilli |
| 25 | Cristian Baldenegro Barrera | 79 | Daniel Savoiu |
| 26 | Tatyana Donskova | 80 | Adomas Jelinskas |
| 27 | Filip Nechansky | 81 | Alena Loesle |
| 28 | Nis Meinert | 82 | Jo Van den Brand |
| 29 | Ralf Gugel | 83 | Daniela Schaefer |
| 30 | Elmira Melyan | 84 | Anton Petrov |
| 31 | Daniele Ruini | 85 | Jared Little |
| 32 | Yasiel Delabat Diaz | 86 | Hessamoddin Kaveh |
| 33 | Marco Lavorgna | 87 | Peter Johannes Falke |
| 34 | Antonio Vagnerini | 88 | Titus Mombacher |
| 35 | Lukas Gerritzen | 89 | Nihal Brahimi |
| 36 | Alessia Saggio | 90 | Daniel Dabrowski |
| 37 | Baptiste Ravina | 91 | Alejandro Alfonso Albero |
| 38 | Rosanna Ignazzi | 92 | Vincent Wai Sum Wong |
| 39 | Harrison Prosper | 93 | Dylan Frizzell |
| 40 | Davide Pagani | 94 | Ziyu Guo |
| 41 | Andrea De Simone | 95 | Sophia Borowka |
| 42 | Fabrizio Trovato | 96 | You-Ying Li |
| 43 | Christina Agapopoulou | 97 | Matteo Defranchis |
| 44 | Joscha Knolle | 98 | Hoang Dai Nghia Nguyen |
| 45 | Antonio Giannini | 99 | Sandra Consuegra Rodriguez |
| 46 | Christoph Andreas Garbers | 100 | Lukas Tropp |
| 47 | Aleksandr Dolmatov | 101 | Alessandra Betti |
| 48 | Tadej Novak | 102 | Thibault Chevalerias |
| 49 | Ilnur Gabdrakhmanov | 103 | Danyer Perez Adan |
| 50 | Anjishnu Bandyopadhyay | 104 | Nick Ellis |
| 51 | Ana Rosario Cueto Gomez | 105 | Filippo Errico |
| 52 | Sonia Carra | 106 | Francesco Cirotto |
| 53 | Eftychia Tzovara | 107 | Luca Lista |
| 54 | Korbinian Schweiger | | |

# Photographs (montage)

The 2018 European School of High-Energy Physics

Maratea, Italy, 20 June - 3 July 2018

# Contents

# Quantum field theory and the Electroweak Standard Model

*A.V. Bednyakov*
BLTP JINR, Dubna, Russia

### Abstract
These lecture notes cover the basics of Quantum Field Theory (QFT) and peculiarities in the construction of the Electroweak (EW) sector of the Standard Model (SM). In addition, the present status, issues, and prospects of the SM are discussed.

### Keywords
Quantum field theory; standard model; electroweak; lectures.

## 1 Introduction

The Standard Model (SM) [1–3] was established in the mid-1970s. Its success is incredible: even after almost half a century, no significant deviations from the SM predictions have been found.

*But what is the SM?*

After the discovery [4,5] of the Higgs boson at the LHC, it is fair to give the following *short* answer [6]:

*The Absolutely Amazing Theory of Almost Everything.*

There are many excellent lectures (e.g., [7–10]) and textbooks (e.g., [11, 12]) that can provide a lot of convincing arguments for such a fancy name. In this course we are not able to cover all the aspects of the SM, but just review some basic facts and underlying principles of the model emphasizing salient features of the latter.

Let us start with a brief overview of the SM particle content (see Fig. 1). One usually distinguishes fermions (half-integer spin) from bosons (integer spin). Traditionally, fermions are associated with "matter", while bosons take the role of "force carriers" that mediate interactions between spin-1/2 particles. In the SM, there are three *generations* involving two types of fermions - *quarks* and *leptons*. In total, we have

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

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

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

There is also a famous spin-0 Higgs boson $h$, which plays an important role in the construction of the SM. It turns out that only gluons and photons ($\gamma$) are assumed to be massless.[1] All other *elementary* particles are massive due to the Higgs mechanism.

---

[1]Initially neutrinos $\nu_l$ were assumed to be massless in the SM but experiments show that it is not the case.
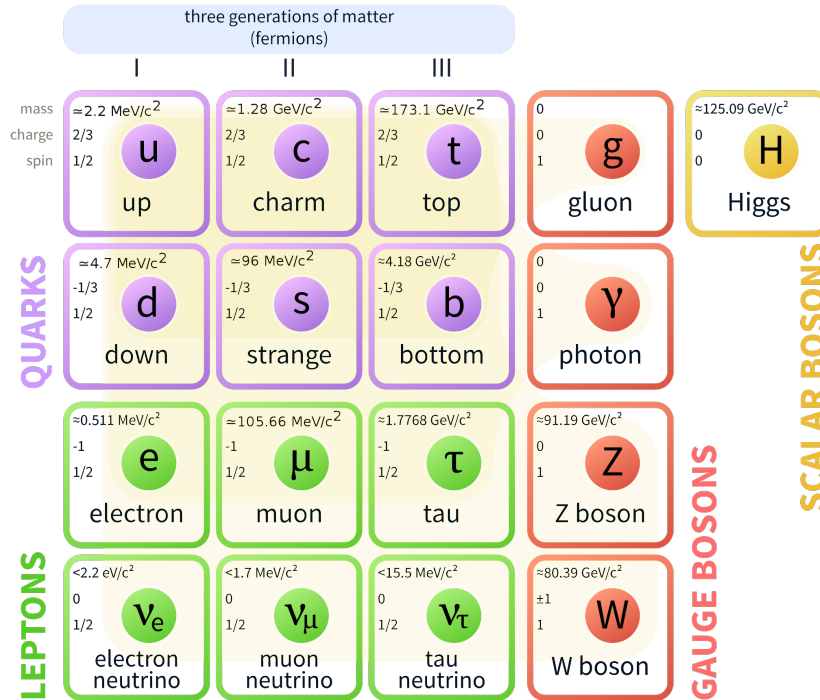
---

# Standard Model of Elementary Particles



**Fig. 1:** Particle content of the Standard Model. Courtesy to Wikipedia.

In the SM the properties of the particle interactions can be read off the SM *Lagrangian* $\mathcal{L}_{SM}$. One can find its compact version on the famous CERN T-shirt. However, there is a lot of structure behind the short expression and it is *Quantum Field Theory* or QFT (see, e.g., textbooks [12–16]) that allows us to derive the full Lagrangian and understand why the T-shirt Lagrangian is unique in a sense.

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

We begin by introducing quantum fields in Sec. 2 as the key objects of the relativistic quantum theory of particles. Then we discuss (global) symmetries in Sec. 3 and emphasize the relation between symmetries and particle properties. We switch from free to interacting fields in Sec. 4 and give a brief overview of techniques used to perform calculations in QFT models. We introduce gauge (or local) symmetries in Sec. 5 and discuss how they are realized in the SM (Sec. 6). The experimental status of the SM can be found in Sec. 7. Final remarks and conclusions are provided in Sec. 8.

## 2 From particles to quantum fields

Before we begin our discussion of quantum fields let us set up our notation. We work in natural units with the speed of light $c = 1$ and the (reduced) Planck constant $\hbar = 1$. In this way, all the quantities in particle physics are expressed in powers of electron-Volts (eV). To recover ordinary units, the following conversion constants can be used:

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

In High-Energy Physics (HEP) we routinely deal with particles traveling at speed $v \lesssim c$. As a consequence, we require that our theory should respect Lorentz *symmetry* that leaves a scalar product[2]

$$px \equiv p_\mu x_\mu = g_{\mu\nu} p_\mu x_\nu = p_0 x_0 - \mathbf{p} \cdot \mathbf{x}, \quad g_{\mu\nu} = \text{diag}(1, -1, -1, -1) \tag{2}$$

of any four-vectors, e.g., space-time coordinates $x_\mu$ and energy-momenta $p_\mu$

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

invariant under rotations and boosts parametrized by $\Lambda_{\mu\nu}$:

$$x_\mu \to x'_\mu = \Lambda_{\mu\nu} x_\nu, \qquad x_\mu x_\mu = x'_\mu x'_\mu \Rightarrow \Lambda_{\mu\alpha} \Lambda_{\mu\beta} = g_{\alpha\beta} \tag{3}$$

It is this requirement that forces us to use QFT as a theory of *relativistic* particles. Relativistic quantum mechanics (RQM) describing a fixed number of particles turns out to be inconsistent. Indeed, from the energy-momentum relation for a free relativistic particle

$$E^2 = \mathbf{p}^2 + m^2 \quad \text{(instead of } E = \frac{\mathbf{p}^2}{2m} \text{ in the non-relativistic case)},$$

and the *correspondence* principle

$$E \to i\frac{\partial}{\partial t}, \qquad \mathbf{p} \to -i\nabla$$

one obtains a relativistic analog of the Shrödinger equation - the Klein-Gordon (KG) equation

$$\left(\partial_t^2 - \nabla^2 + m^2\right) \phi(t, \mathbf{x}) = 0 \quad \text{(instead of } i\partial_t \psi = -\frac{\nabla^2}{2m}\psi) \tag{4}$$

for a wave-function $\phi(t, \mathbf{x}) \equiv \langle \mathbf{x} | \phi(t) \rangle$. It has two plane-wave solutions for any three-dimensional $\mathbf{p}$:

$$\phi_\mathbf{p}(t, \mathbf{x}) = e^{-iEt + \mathbf{px}}, \quad \text{with } E = \pm\omega_p, \quad \omega_p = +\sqrt{\mathbf{p}^2 + m^2}. \tag{5}$$

One can see that the spectrum (5) is not bounded from below. Another manifestation of this problem is the fact that for a *general* wave-packet solution

$$\phi(t, \mathbf{x}) = \frac{1}{(2\pi)^{3/2}} \int \frac{d\mathbf{p}}{\sqrt{2\omega_p}} \left[ a(\mathbf{p}) e^{-i\omega_p t + i\mathbf{px}} + b(\mathbf{p}) e^{+i\omega_p t - i\mathbf{px}} \right] \tag{6}$$

we are not able to introduce a *positive-definite* probability density $\rho$

$$\rho \equiv j_0 = i\left(\phi^* \partial_t \phi - \phi \partial_t \phi^*\right) \Rightarrow 2E \text{ for } \phi \propto e^{-iEt}, \tag{7}$$

required to interpret $\phi$ as a wave-function of a single particle. Of course, one can try to impose the positive-energy condition $[b(\mathbf{p}) \equiv 0]$ but it is not stable under *interactions*. A single-particle interpretation fails to account for the appearance of negative-energy modes and we need a new formalism to deal with such situations. Moreover, in RQM space coordinates play a role of dynamical variables and are represented by operators, while time is an evolution parameter. Obviously, a *consistent* relativistic theory should treat space and time on equal footing.

In order to circumvent these difficulties, one can re-interpret $\phi(\mathbf{x}, t)$ satisfying (4) as a *quantum* field, i.e., an *operator*[3] $\hat{\phi}(\mathbf{x}, t)$. The space coordinates $\mathbf{x}$ can be treated as a *label* for infinitely many

---

[2]Summation over repeated indices is implied.

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

dynamical variables and we are free to choose a system of reference, in which we evolve these variables. As a consequence, a single field can account for an infinite number of particles, which are treated as field *excitations*. In the QFT notation the solution of the KG equation($p_0 = \omega_p$) can be rewritten

$$\phi(x) = \frac{1}{(2\pi)^{3/2}} \int \frac{d\mathbf{p}}{\sqrt{2\omega_p}} \left[ a_{\mathbf{p}}^- e^{-ipx} + b_{\mathbf{p}}^+ e^{+ipx} \right] \tag{8}$$

as a linear combination of *operators* $a_{\mathbf{p}}^\pm$ and $b_{\mathbf{p}}^\pm$ obeying

$$a_{\mathbf{p}}^- a_{\mathbf{p}'}^+ - a_{\mathbf{p}'}^+ a_{\mathbf{p}}^- \equiv \left[ a_{\mathbf{p}}^-, a_{\mathbf{p}'}^+ \right] = \delta^3(\mathbf{p} - \mathbf{p}'), \quad \left[ b_{\mathbf{p}}^-, b_{\mathbf{p}'}^+ \right] = \delta^3(\mathbf{p} - \mathbf{p}'). \tag{9}$$

All other commutators are zero, e.g., $\left[ a_{\mathbf{p}}^\pm, a_{\mathbf{p}'}^\pm \right] = 0$. The operators satisfy $a_{\mathbf{p}}^\pm = (a_{\mathbf{p}}^\mp)^\dagger$ and $b_{\mathbf{p}}^\pm = (b_{\mathbf{p}}^\mp)^\dagger$, and for $a_{\mathbf{p}}^\pm \equiv b_{\mathbf{p}}^\pm$ the field is hermitian $\phi^\dagger(x) = \phi(x)$.

The operator (8) needs some space to act on and in QFT we consider the *Fock* space. It consists of a *vacuum* $|0\rangle$, which is *annihilated* by $a_{\mathbf{p}}^-$ (and $b_{\mathbf{p}}^-$) for every $\mathbf{p}$

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

and field excitations. The latter are *created* from the vacuum by acting with $a_{\mathbf{k}}^+$ (and/or $b_{\mathbf{k}}^+$), e.g.,

$$|f_1\rangle = \int d\mathbf{k} \cdot f_1(\mathbf{k}) a_{\mathbf{k}}^+ |0\rangle, \qquad \text{1-particle state;} \tag{10}$$

$$|f_2\rangle = \int d\mathbf{k}_1 d\mathbf{k}_2 \cdot f_2(\mathbf{k}_1, \mathbf{k}_2) a_{\mathbf{k_1}}^+ a_{\mathbf{k_2}}^+ |0\rangle \qquad \text{2-particle state,} \tag{11}$$

$$\ldots$$

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

The two set of operators $a^\pm$ and $b^\pm$ correspond to particles and antiparticles. From the commutation relations we deduce that $a_{\mathbf{p}}^+ a_{\mathbf{k}}^+ = a_{\mathbf{k}}^+ a_{\mathbf{p}}^+$, so particles are *not distinguishable* by construction.

The commutation relations (10) should remind us about a bunch of independent *quantum harmonic oscillators*. Indeed, the corresponding Hamiltonian

$$\hat{\mathcal{H}}_{osc} = \sum_j \frac{1}{2}(\hat{p}_j^2 + \omega_j^2 \hat{x}_j^2) = \sum_j \frac{\omega_j}{2} \left( a_j^+ a_j^- + a_j^- a_j^+ \right) = \sum_j \omega_j \left( \hat{n}_j + \frac{1}{2} \right) \tag{12}$$

can be expressed in terms of *ladder* operators $\sqrt{2\omega_j} a_j^\pm = (\omega_j \hat{x}_j \mp i\hat{p}_j)$ (no summation), which satisfy $[a_j^-, a_k^+] = \delta_{jk}$ similar to Eq.(10). For convenience we re-order operators entering into $\hat{\mathcal{H}}_{osc}$ and introduce $\hat{n}_j = a_j^+ a_j^-$ that counts energy quanta $\hat{n}_j |n_j\rangle = n_j |n_j\rangle$. A direct consequence of the re-ordering is the fact that the lowest possible state (vacuum $|0\rangle$) has non-zero energy, which is equal to the sum of zero-point energies $\sum_j \omega_j/2$ of all oscillators.

We can make the analogy between a (free) field and harmonic oscillators more pronounced if we put our field in a box of size $L$. In this case, the energy $\omega_p$ and momentum $\mathbf{p}$ are quantized

$$\mathbf{p} \to \mathbf{p_j} = (2\pi/L)\mathbf{j}, \quad \omega_p \to \omega_{\mathbf{j}} = \sqrt{(2\pi/L)^2 \mathbf{j}^2 + m^2}, \quad \mathbf{j} = (j_1, j_2, j_3), j_i \in \mathbb{Z}.$$

The corresponding $\hat{\mathcal{H}}_{osc}$ (12) can be used to deduce the (QFT) Hamiltonian (by taking the limit $L \to \infty$):

$$\hat{\mathcal{H}}_{part} = \lim_{L\to\infty} \underbrace{\left[ \left( \frac{2\pi}{L} \right)^3 \sum_{\mathbf{j}} \right]}_{\int d\mathbf{p}} \omega_{\mathbf{j}} \left[ \underbrace{\left( \frac{L}{2\pi} \right)^{\frac{3}{2}} a_{\mathbf{j}}^+}_{a_{\mathbf{p}}^+} \underbrace{\left( \frac{L}{2\pi} \right)^{\frac{3}{2}} a_{\mathbf{j}}^-}_{a_{\mathbf{p}}^-} + \frac{1}{2} \underbrace{\left( \frac{L}{2\pi} \right)^3}_{\delta(\mathbf{0})} \right].$$

Since our field (8) involves two kinds of ladder operators, we have

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_{part} + \hat{\mathcal{H}}_{antipart} = \int d\mathbf{p}\, \omega_p \left[ n_{\mathbf{p}} + \bar{n}_{\mathbf{p}} \right] + \int d\mathbf{p}\, \omega_p \delta(\mathbf{0}) \tag{13}$$

with $\bar{n}_{\mathbf{p}} \equiv b_{\mathbf{p}}^+ b_{\mathbf{p}}^-$ and $n_{\mathbf{p}} \equiv a_{\mathbf{p}}^+ a_{\mathbf{p}}^-$. The interpretation of the first term is straightforward: $(\bar{n}_{\mathbf{p}})\, n_{\mathbf{p}}$ *counts* (anti-)particles with definite momentum $\mathbf{p}$ and there is a sum over the corresponding energies. The second term in Eq.(13) looks disturbing. It is associated with *infinite* vacuum (no particles) energy:

$$E_0 = \langle 0|\hat{\mathcal{H}}|0\rangle = \int d\mathbf{p}\, \omega_p \delta(\mathbf{0}).$$

Actually, there are two kinds of infinities in $E_0$:

  – InfraRed    (large distances, $L \to \infty$) due to $L^3 \to (2\pi)^3 \delta(\mathbf{0})$;
  – UltraViolet (small distances, $\mathbf{p}, \omega_p \to \infty$).

One usually "solves" this problem by introducing *normal-ordered* Hamiltonians, e.g.,

$$:\!\hat{\mathcal{H}}_{osc}\!:= \frac{\omega_j}{2}\left( :\!a_j^+ a_j^- + a_j^- a_j^+\!: \right) = \omega_j :\!a_j^+ a_j^-\!:= \omega_j a_j^+ a_j^-.$$

With $:\!\hat{\mathcal{H}}\!:$ we measure all energies with respect to the vacuum $\hat{\mathcal{H}} \to :\!\hat{\mathcal{H}}\!:= \hat{\mathcal{H}} - \langle 0|\hat{\mathcal{H}}|0\rangle$ and ignore (non-trivial) dynamics of the latter. In what follows we assume that operators are normal-ordered by default.

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

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

are eigenvectors of the Hamiltonian with *positive* energies and we avoid introduction of negative energies in our formalism from the very beginning. One can generalize Eq. (12) and "construct" the momentum $\hat{\mathbf{P}}$ and charge $\hat{Q}$ operators[4]:

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

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

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

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

and consider the following amplitudes:

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

$$\text{Particle (charge } +1) \qquad\qquad \text{Antiparticle (charge } -1)$$
$$\text{propagates from } x_1 \text{ to } x_2 \qquad\qquad \text{propagates from } x_2 \text{ to } x_1$$

---

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

Both possibilities can be taken into account in one function:

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

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

Equation (17) is nothing else but the famous Feynman propagator, which has the following momentum representation:

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

The $i\epsilon$-prescription ($\epsilon \to 0$) picks up certain poles in the complex $p_0$ plane (see Fig. 2) and gives rise to the time-ordered expression (17). The propagator plays a key role in the construction of perturbation theory for interacting fields (see Sec. 4.1).

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

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

This gives us an alternative way to find the expression (18). One can also see that $D_c(x - y)$ is a Lorentz and translational invariant function.

The propagator of particles can be connected to the force between two classical static sources $J_i(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_i)$ located at $\mathbf{x}_i = (\mathbf{x}_1, \mathbf{x}_2)$. The sources disturb the vacuum $|0\rangle \to$



**Fig. 2:** Integration contours in $p_0$ plane.

$|\Omega\rangle$, since the Hamiltonian of the system is modified $\mathcal{H} \to \mathcal{H}_0 + J \cdot \phi$. Assuming for simplicity that $\phi = \phi^\dagger$, we can find the energy of the disturbed vacuum from

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

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

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

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

## 3 Symmetries and fields

Let us switch to the discussion of symmetries and their role in QFT. A convenient way to deal with quantum fields and the symmetries of the corresponding physical systems is to consider the following
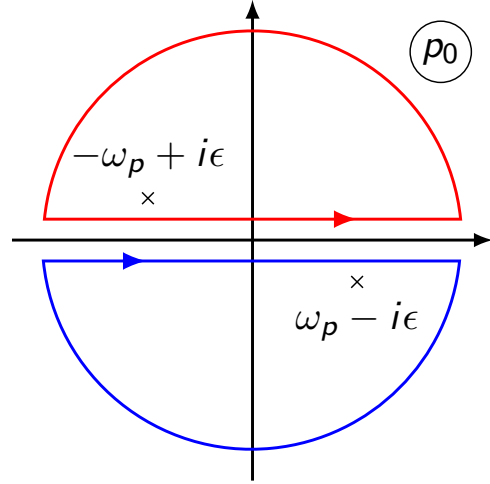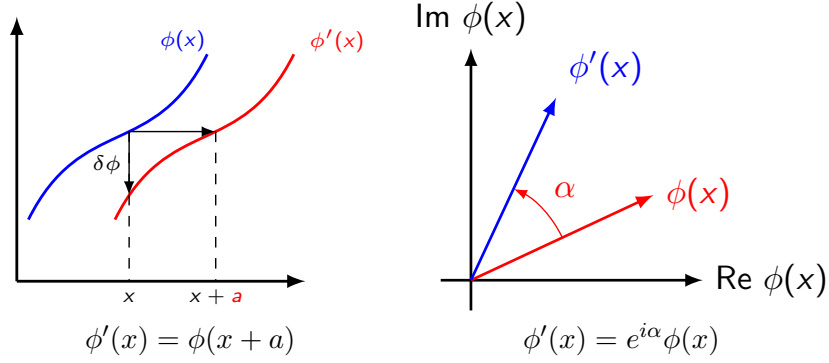
$$\phi'(x) = \phi(x+a) \qquad\qquad \phi'(x) = e^{i\alpha}\phi(x)$$

**Fig. 3:** Translations (left) and phase transformations (right).

*Action* functional[5]

$$\mathcal{A}[\phi(x)] = \int d^4x \;\underbrace{\mathcal{L}(\phi(x),\partial_\mu\phi)}_{\text{Lagrangian (density)}} = \int d^4x \;\Big(\underbrace{\partial_\mu\phi^\dagger\partial_\mu\phi - m^2\phi^\dagger\phi}_{\phi^\dagger\cdot K\cdot\phi}\Big). \tag{20}$$

To have an analogy with a mechanical system, one can rewrite $\mathcal{A}[\phi]$ as

$$\mathcal{A}[\phi(x)] = \int dt\, L(t), \qquad L = T - U, \quad H = T + U$$

$$T = \int d\mathbf{x}|\partial_t\phi|^2, \quad U = \int d\mathbf{x}(|\partial_\mathbf{x}\phi|^2 + m^2|\phi|^2)$$

with $T$ and $U$ being kinetic and potential energy of a system of *coupled* oscillators (a "mattress").

Given a Lagrangian $\mathcal{L}$, one can derive the *equations of motions* (EOM) via the *Action Principle*. For this we consider variation of the action

$$\underbrace{\mathcal{A}[\phi'(x)] - \mathcal{A}[\phi(x)]}_{\delta\mathcal{A}[\phi(x)]=0} = \int d^4x\left[\underbrace{\left(\partial_\mu\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi} - \frac{\partial\mathcal{L}}{\partial\phi}\right)}_{(\partial_\mu^2+m^2)\phi=0}\delta\phi + \underbrace{\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi}\delta\phi\right)}_{\text{surface term}=0}\right]. \tag{21}$$

due to tiny (infinitesimal) shifts in the field $\phi'(x) = \phi(x) + \delta\phi(x)$. If we require that $\delta\mathcal{A}[\phi(x)] = 0$ for *any* variation $\delta\phi(x)$ of *some* $\phi(x)$, we will immediately deduce that this can be achieved only for *specific* $\phi(x)$ that satisfy EOM. These *particular* fields are usually called "on-mass-shell". From the Lagrangian for our free scalar field (20) we derive the KG equation. It is related in a straightforward way to the quadratic form $K$ in Eq.(20). Having in mind Eq.(19), one can see that the (Feynman) propagator can also be obtained by inverting $K$. This statement is easily generalized to the case of other fields.

The Action functional for a physical system allows one to study *Symmetries*. The latter are intimately connected with *transformations*, which leave something *invariant*. The transformations can be *discrete*, such as

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

---

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

or depend on *continuous* parameters. One distinguishes *space-time* from *internal* transformations. Lorentz boosts, rotations, and translations are typical examples of the former, while phase transformations belong to the latter (see Fig. 3). At the moment, we only consider *global* symmetries with parameters independent of space-time coordinates and postpone the discussion of $x$-dependent or *local* (*gauge*) transformations to Sec. 5.

Given $\mathcal{A}[\phi]$, one can *find* its *symmetries*, which can be defined as *particular* infinitesimal variations $\delta\phi(x)$ that for *any* $\phi(x)$ leave $\mathcal{A}[\phi]$ invariant up to a surface term (*cf.* the Action Principle)

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

If we compare this with the general expression

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

and require in addition that $\phi$ *satisfy* EOM[6], we get a *local conservation law*

$$\partial_\mu J_\mu = 0, \quad J_\mu \equiv \mathcal{K}_\mu - \frac{\partial \mathcal{L}}{\partial \partial_\mu \phi} \delta\phi. \tag{22}$$

The integration of Eq. (22) over *space* leads to *conserved* charge:

$$\frac{d}{dt} Q = 0, \qquad Q = \int d\mathbf{x} J_0. \tag{23}$$

If $\delta\phi = \rho_i \delta_i \phi$ depends on parameters $\rho_i$, we have a conservation law for every $\rho_i$. This is the essence of the *first Noether theorem* [17].

A careful reader might notice that we somehow forgot about the quantum nature of our fields and in our discussion of symmetries treat them as classical objects. Let us comment on this fact. In Classical Physics *symmetries* allow one to find

- new solutions to EOM from the given one, keeping some features of the solutions (*invariants*) intact;
- how a solution in one coordinate system (as seen by one observer) looks in other coordinates (as seen by another observer).

In a quantum world a symmetry $\mathcal{S}$ guarantees that transition *probabilities* $\mathcal{P}$ between *states* do not change upon transformation:

$$|A_i\rangle \xrightarrow{\mathcal{S}} |A_i'\rangle, \qquad \mathcal{P}(A_i \to A_j) = \mathcal{P}(A_i' \to A_j'), \qquad |\langle A_i | A_j \rangle|^2 = |\langle A_i' | A_j' \rangle|^2. \tag{24}$$

One can see that symmetries can be *represented* by *unitary*[7] operators $U$:

$$|A_i'\rangle = U|A_j\rangle, \quad \langle A_i' | A_j' \rangle = \langle A_i | \underbrace{U^\dagger U}_{1} |A_j\rangle. \tag{25}$$

In QFT one usually reformulates a symmetry transformation of *states* as a change of *operators* $\mathcal{O}_k$ via

$$\langle A_i | \mathcal{O}_k(x) | A_j \rangle \xrightarrow{\mathcal{S}} \langle A_i' | \mathcal{O}_k(x) | A_j' \rangle = \langle A_i | \mathcal{O}_k'(x) | A_j \rangle, \quad \mathcal{O}_k'(x) \equiv U^\dagger \mathcal{O}_k(x) U. \tag{26}$$

---

[6]This requirement is crucial.

[7]or anti-unitary (e.g., in the case of time reversal).

For example, translational invariance leads to a relation between matrix elements of *quantum* fields, e.g.,

$$\langle A_i|\phi(x)|A_j\rangle = \langle A_i|\phi'(x+a)|A_j\rangle = \langle A_i|U^\dagger(a)\phi(x+a)U(a)|A_j\rangle \tag{27}$$

for any states. As a consequence, any *quantum* field in a translational invariant theory should satisfy

$$\phi(x+a) = U(a)\phi(x)U^\dagger(a). \tag{28}$$

One can also find similar *constraints* on quantum operators due to other symmetries.

By means of the Noether theorem we can get almost at no cost the expressions for energy-momentum $P_\mu = (\mathcal{H}, \mathbf{P})$ and charge $Q$, which we used in Sec. 2. For example, $P_\mu$ is nothing else but the conserved "charges", which correspond to space-time translations. Indeed, the Noether current in this case is just the energy-momentum tensor $T_{\mu\nu}$

$$\phi'(x+a) = \phi(x), \qquad \text{expand in } a \Rightarrow \delta\phi(x) = -a_\nu\partial_\nu\phi(x), \tag{29}$$

$$\delta\mathcal{L}(\phi(x), \partial_\mu\phi(x)) = \partial_\nu(-a_\nu\mathcal{L}) \Rightarrow J_\mu = -a_\mu\mathcal{L} + a_\nu\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi}\partial_\nu\phi = a_\nu T_{\mu\nu}. \tag{30}$$

According to Eq. (23), for every $a_\mu$ we have $P_\nu = \int d\mathbf{x}T_{0\nu}$, i.e., conserved total energy-momentum. In the same way, we can apply the Noether theorem to phase transformations of our *complex* field and get

$$\phi'(x) = e^{i\alpha}\phi(x), \quad \delta\phi(x) = i\alpha\phi(x), \quad J_\mu = i(\phi^\dagger\partial_\mu\phi - \phi\partial_\mu\phi^\dagger), \quad Q = \int d\mathbf{x}J_0. \tag{31}$$

The corresponding quantum operators, i.e., $\hat{\mathcal{H}}$ (12) or $\hat{Q}$ (16), are obtained[8] (modulo ordering issues) from these (classical) expressions by plugging in quantum field $\hat{\phi}$ from Eq.(8). It turns out that the charges act as *generators* of symmetries, e.g., for space-time translations the unitary operator from Eq.(28) is given by

$$U(a) = \exp\left(i\hat{P}_\mu a_\mu\right), \qquad \hat{\phi}(x+a) = U(a)\hat{\phi}(x)U^\dagger(a). \tag{32}$$

In addition, *conserved* quantities can be used to define a convenient *basis* of states, e.g., we characterize our particle states by eigenvalues of $P_\mu$, and $Q$:

$$|\mathbf{p}\rangle \equiv |\mathbf{p}, +1\rangle, |\bar{\mathbf{p}}\rangle \equiv |\mathbf{p}, -1\rangle \Rightarrow \hat{Q}|\mathbf{p}, q\rangle = q|\mathbf{p}, q\rangle, \hat{\mathbf{P}}|\mathbf{p}, q\rangle = \mathbf{p}|\mathbf{p}, q\rangle. \tag{33}$$

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

$$\phi_i'(x') = S_{ij}(a)\phi_j(x) \Rightarrow \phi_i(x') = S_{ij}(a)U(a)\phi_j(x)U^\dagger(a), \quad x' = x'(x, a). \tag{34}$$

Typical examples are fields with non-zero spin: they have several components, which also change under coordinate rotations (more generally, under Lorentz transformations). Moreover, it is the Lorentz symmetry that allows us to classify fields as different *representations* of the corresponding *group*.

Let us discuss this in more detail. We can describe fields involving several degrees of freedom (per space point) by adding more (and more) *indices* $\phi(x) \to \Phi_\alpha^i(x)$. One can split the indices into two groups: *space-time* ($\alpha$) and *internal* ($i$). The former are associated with space-time transformations, while the latter with transformations in the "internal" space:

$$\text{Lorentz transform } \Lambda: \Phi_\alpha'^i(\Lambda x) = S_{\alpha\beta}(\Lambda)\Phi_\beta^i(x), \tag{35}$$

$$\text{Internal transform } a: \quad \Phi_\alpha'^i(x) = U^{ij}(a)\Phi_\alpha^j(x). \tag{36}$$

---

[8]We leave this as an exercise.

A quantum field in this case can be represented as

$$\Phi_\alpha^i(x) = \frac{1}{(2\pi)^{3/2}} \sum_s \int \frac{d\mathbf{p}}{\sqrt{2\omega_p}} \left[ u_\alpha^s(\mathbf{p})\, (a_\mathbf{p}^-)_s^i \, e^{-ipx} + v_\alpha^{*s}(\mathbf{p})(b_\mathbf{p}^+)_s^i \, e^{+ipx} \right]. \tag{37}$$

Here the factors $e^{\pm ipx}$ with $p_0 = \omega_\mathbf{p}$ (plane waves) guarantee that every component of $\Phi_\alpha^i$ satisfies the KG equation. The sum in Eq.(37) is over all polarization states, which are characterized by polarization "vectors" for particles $u_\alpha^s(\mathbf{p})$ annihilated by $(a_\mathbf{p}^-)_s^i$, and anti-particles $v_\alpha^{*s}(\mathbf{p})$ created by $(b_\mathbf{p}^+)_s^i$ . The conjugated field $(\Phi_\alpha^i)^\dagger$ involves (conjugated) polarization vectors for (anti) particles that are (annihilated) created. Let us give a couple of examples:

- Quarks are *coloured fermions* $\Psi_\alpha^i$ and, e.g., $(a_\mathbf{p}^-)_s^b$ annihilates the "blue" quark in a spin state $s$. The latter is characterized by a spinor $u_\alpha^s(\mathbf{p})$;
- There are *eight vector* gluons $G_\mu^a$. So $(a_\mathbf{p}^-)_s^a$ annihilates a gluon $a$ in spin state $s$ having polarization $u_\alpha^s(\mathbf{p}) \to \epsilon_\mu^s(\mathbf{p})$.

Since Lorentz symmetry plays a key role in QFT, we elaborate on some of its non-trivial representations and consider vector and fermion fields in more detail.

## 3.1 Massive vector fields

A charged Vector Field (e.g., a $W$-boson) can be written as

$$W_\mu(x) = \frac{1}{(2\pi)^{3/2}} \sum_{\lambda=1}^3 \int \frac{d\mathbf{p}}{\sqrt{2\omega_p}} \left[ \left( \epsilon_\mu^\lambda(\mathbf{p})a_\lambda^-(\mathbf{p})e^{-ipx} + \epsilon_\mu^{*\lambda}(\mathbf{p})b_\lambda^+(\mathbf{p})\, e^{+ipx} \right) \right]. \tag{38}$$

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

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

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

$$\langle 0|T(W_\mu(x)W_\nu^\dagger(y))|0\rangle = \frac{1}{(2\pi)^4} \int d^4p\, e^{-ip(x-y)} \left[ \frac{-i\left( g_{\mu\nu} - \frac{p_\mu p_\nu}{m^2} \right)}{p^2 - m^2 + i\epsilon} \right] \quad [p_0 - \text{arbitrary}] \tag{39}$$

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

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

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

## 3.2 Massless vector fields

Massless (say photon) vectors are usually represented by

$$A_\mu(x) = \frac{1}{(2\pi)^{3/2}} \sum_{\lambda=0}^3 \int \frac{d\mathbf{p}}{\sqrt{2\omega_p}} \left[ \epsilon_\mu^\lambda(\mathbf{p})a_\lambda^-(\mathbf{p})e^{-ipx} + \text{h.c.} \right]. \tag{40}$$

with

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

The corresponding Feynman propagator can be given by

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

In spite of the fact that we sum over four polarizations in Eq.(40) only *two* of them are *physical*! This reflects the fact that the vector-field Lagrangian in the massless case $m = 0$

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

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

## 3.3 Fermion fields

Spin-1/2 fermion fields (e.g., leptons) are given by[9]

$$\psi^\alpha(x) = \frac{1}{(2\pi)^{3/2}}\int\frac{d\mathbf{p}}{\sqrt{2\omega_p}}\sum_{s=1,2}\left[u_s^\alpha(\mathbf{p})a_s^-(\mathbf{p})e^{-ipx} + v_s^\alpha(\mathbf{p})b_s^+(\mathbf{p})e^{+ipx}\right],$$

where we explicitly write the *spinor* (Dirac) index $\alpha$ for $u_s$, $v_s$ and the quantum operator $\psi$. The former satisfy the $4 \times 4$ matrix (Dirac) equations

$$(\hat{p} - m)u_s(\mathbf{p}) = 0, \quad (\hat{p} + m)v_s(\mathbf{p}) = 0, \quad \hat{p} \equiv \gamma_\mu p_\mu, \quad p_0 \equiv \omega_\mathbf{p} \tag{41}$$

and correspond to particles ($u_s$) or antiparticles ($v_s$). In Eq.(41) we use gamma-matrices

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

to account for *two* spin states ($s = 1, 2$) of particles and antiparticles. Fermion fields transform under the Lorentz group $x' = \Lambda x$ as (*cf.* Eq.(35))

$$\psi'(x') = \mathcal{S}_\Lambda\psi(x), \qquad \psi'(x')^\dagger = \psi(x)\mathcal{S}_\Lambda^\dagger. \tag{42}$$

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

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

This allows us to convince ourselves that the Dirac Lagrangian

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

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

$$\bar{u}_s(\mathbf{p})u_r(\mathbf{p}) = 2m\delta_{rs}, \qquad \bar{v}_s(\mathbf{p})v_r(\mathbf{p}) = -2m\delta_{rs},$$

---

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

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

$$\left[a^{-}_{r,\mathbf{p}}, a^{+}_{s,\mathbf{p}'}\right]_{+} = \left[b^{-}_{r,\mathbf{p}}, b^{+}_{s,\mathbf{p}'}\right]_{+} = \delta_{sr}\delta(\mathbf{p} - \mathbf{p}')$$

$$\left[a^{\pm}_{r,\mathbf{p}}, a^{\pm}_{s,\mathbf{p}'}\right]_{+} = \left[b^{\pm}_{r,\mathbf{p}}, b^{\pm}_{s,\mathbf{p}'}\right]_{+} = \left[a^{\pm}_{r,\mathbf{p}}, b^{\pm}_{s,\mathbf{p}'}\right]_{+} = 0.$$

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

Let us continue our discussion of free fermions by emphasizing the difference between the notions of *Chirality* and *Helicity*. Two independent solutions for *massive* fermions ($u_{1,2}$) can be chosen to correspond to two different *helicities* — projections of spin vector **s** onto direction of **p**:

$$\mathcal{H} = \mathbf{s}\cdot\mathbf{n}, \quad \mathbf{n} = \mathbf{p}/|\mathbf{p}|. \tag{43}$$



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

*By definition* Left ($\psi_L$) and Right ($\psi_R$) *chiral* spinors are eigenvectors of

$$\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3 \Rightarrow [\gamma_\mu, \gamma_5]_+ = 0, \quad \gamma_5^2 = 1, \quad \gamma_5^\dagger = \gamma_5, \tag{44}$$

where

$$\gamma_5\psi_L = -\psi_L, \qquad \gamma_5\psi_R = +\psi_R. \tag{45}$$

*Any* spinor $\psi$ can be decomposed as

$$\psi = \psi_L + \psi_R, \qquad \psi_{L/R} = P_{L/R}\psi, \quad P_{L/R} = \frac{1\mp\gamma_5}{2}. \tag{46}$$

Rewriting the Dirac Lagrangian it terms of chiral components

$$\mathcal{L} = i(\underbrace{\bar{\psi}_L\hat{\partial}\psi_L + \bar{\psi}_R\hat{\partial}\psi_R}_{\text{conserve chirality}}) - m(\underbrace{\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L}_{\text{break chirality}}), \tag{47}$$

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

$$\psi \to e^{i\gamma_5\alpha}\psi. \tag{48}$$

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

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

$$\mathcal{L} = \frac{1}{2}(i\bar{\psi}_L\hat{\partial}\psi_L - m\bar{\psi}_L\psi_L^c). \tag{49}$$

As a consequence, to describe Majorana particles, we need only two components instead of four since antiparticles coincide with particles in this case. At the moment, the nature of neutrinos is unclear and we refer to [18] for more details.

## 4  From free to interacting fields

Let us summarize what we have learned so far. If we have a Lagrangian $\mathcal{L}$ at hand, we can

- Derive *EOM* (via the *Action* Principle);
- Find the *Symmetries* of the Action $\mathcal{A} = \int d^4x \mathcal{L}$;
- Find *Conserved* quantities (via the Noether Theorem).

However, we usually *start* building our models by *postulating* symmetries. Indeed, we assume that a general QFT Lagrangian $\mathcal{L}$ is

- a Lorentz (Poincare) invariant (i.e., a sum of Lorentz scalars),
- Local (involves a finite number of partial derivatives),
- Real (hermitian) (respects unitarity=conservation of probability)

In addition, one can impose other symmetries and get further restrictions on the model. Having all this in mind, we can proceed further and discuss particle *interactions*.

In HEP, a typical collision/scattering experiment deals with *"free"* initial and final states and considers *transitions* between these states. To account for this in a quantum theory, one introduces the *S-matrix* with matrix elements

$$\mathcal{M} = \langle \beta | S | \alpha \rangle, \qquad \mathcal{M} = \delta_{\alpha\beta} + (2\pi)^4 \delta^4(p_\alpha - p_\beta) i M_{\alpha\beta} \tag{50}$$

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

$$|\alpha\rangle = \tilde{a}^+_{\mathbf{p}_1} ... \tilde{a}^+_{\mathbf{p}_r} |0\rangle, \quad |\beta\rangle = \tilde{a}^+_{\mathbf{k}_1} ... \tilde{a}^+_{\mathbf{k}_s} |0\rangle, \quad \tilde{a}^+_{\mathbf{p}} = (2\pi)^{3/2} \sqrt{2\omega_{\mathbf{p}}} a^+_{\mathbf{p}}, \tag{51}$$

where for convenience[10] (see also Eq.(62)) we rescale our creation/annihilation operators. Given the matrix element $M_{\alpha\beta}$, one can calculate the differential probability (per unit volume per unit time) to evolve from $|\alpha\rangle$ to $|\beta\rangle$:

$$dw = \frac{n_1 ... n_r}{(2\omega_{p_1}) ... (2\omega_{p_r})} |M_{\alpha\beta}|^2 d\Phi_s, \tag{52}$$

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

$$d\Phi_s = (2\pi)^4 \delta^4(p_{in} - k_{out}) \frac{d\mathbf{k}_1}{(2\pi)^3 (2\omega_{k_1})} ... \frac{d\mathbf{k}_i}{(2\pi)^3 (2\omega_{k_i})} \tag{53}$$

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

$$d\Gamma = \Phi_\Gamma |M_{1 \to s}|^2 d\Phi_s, \qquad \Phi_\Gamma = \frac{1}{2m}, \tag{54}$$

$$d\sigma = \Phi_\sigma |M|^2 d\Phi_s, \qquad \Phi_\sigma = \frac{1}{4\sqrt{(p_1 p_2)^2 - p_1^2 p_2^2}}. \tag{55}$$

In Eq.(55) the factor $\Phi_\sigma$ is *Lorentz-invariant* and is expressed in terms of four-momenta of initial particles $p_1$ and $p_2$. The total width $\Gamma$ and total cross-section $\sigma$ can be obtained by integration over the momenta of final particles restricted by energy-momentum conservation due to the four-dimensional $\delta$-function in Eq.(53).

In QFT, the S-matrix is given by the time-ordered exponent

$$S = Te^{-i \int d^4x \mathcal{H}_I(x)} = Te^{i \int d^4x \mathcal{L}_I(x)}. \tag{56}$$

---

[10] The states created by $\tilde{a}^+$ are normalized in the *relativistic-invariant* way.
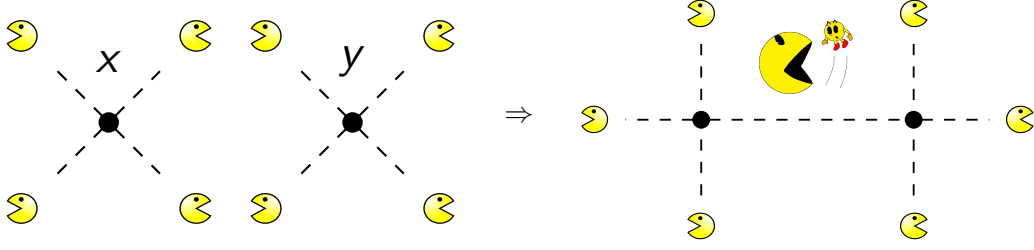
**Fig. 4:** The Wick theorem at work: one of the contributions.

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

The interaction Lagrangian $\mathcal{L}_I = \mathcal{L}_{full} - \mathcal{L}_0$ is a sum of *Lorentz-invariant* terms having more than *two* fields and more $\partial_\mu$ than in the quadratic part $\mathcal{L}_0$, which corresponds to free particles. It is worth noting that in Eq.(56) we treat $\mathcal{L}_I$ ($\mathcal{H}_I$) as an operator built from *free*[11] quantum fields (i.e., certain combinations of $a^\pm$ and $b^\pm$).

The *time-ordering* operation, which was used to define particle propagators, is generalized in Eq.(56) to account for more than two fields originating from $\mathcal{L}_I$

$$T\Phi_1(x_1)...\Phi_n(x_n) = (-1)^k \Phi_{i_1}(x_{i_1})...\Phi_{i_n}(x_{i_n}), \qquad x_{i_1}^0 > ... > x_{i_n}^0. \tag{57}$$

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

As it was mentioned earlier, (interaction) Lagrangians should be hermitian. Any scalar combination of quantum fields can, in principle, be included in $\mathcal{L}_I$, e.g.,

$$\begin{aligned} \mathcal{L}_I: \quad & g\phi^3(x), \qquad \lambda\phi^4(x), \qquad y\bar{\psi}(x)\psi(x)\phi(x) \\ & e\bar{\psi}(x)\gamma_\mu\psi(x)A_\mu(x), \qquad G\left[(\bar{\psi}_1\gamma_\mu\psi_2)(\bar{\psi}_3\gamma_\mu\psi_4) + \text{h.c.}\right] \end{aligned}$$

The parameters (couplings) $g$, $\lambda$, $e$, $y$, and $G$ set the strength of the interactions. An important characteristic of any coupling in the QFT model is its *dimension*, which can be deduced from the fact that Lagrangian has dimension $[\mathcal{L}] = 4$. One can notice that all the couplings (hidden) in the T-shirt Lagrangian are *dimensionless*. This fact has crucial consequences for the self-consistency of the SM model.

## 4.1 Perturbation theory

In an interacting theory it is very hard, if not impossible, to calculate the S-matrix (56) exactly. Usually, we assume that the couplings in $\mathcal{L}_I$ are small allowing us to treat the terms in $\mathcal{L}_I$ as *perturbations* to $\mathcal{L}_0$. As a consequence, we expand the T-exponent and restrict ourselves to a finite number of terms. In the simplest case of $\mathcal{L}_I = -\lambda\phi^4/4!$ we have at the $n$th order

$$\frac{i^n}{n!}\left[\frac{\lambda}{4!}\right]^n \langle 0|\tilde{a}_{\mathbf{k}_1}^-...\tilde{a}_{\mathbf{k}_s}^- \int dx_1...dx_n T\left[\phi(x_1)^4...\phi(x_n)^4\right] \tilde{a}_{\mathbf{p}_1}^+...\tilde{a}_{\mathbf{p}_r}^+|0\rangle. \tag{58}$$

To proceed, one uses the *Wick* theorem:

$$T\Phi_1...\Phi_n = \sum(-1)^\sigma \langle 0|T(\Phi_{i_1}\Phi_{i_2})|0\rangle...\langle 0|T(\Phi_{i_{m-1}}\Phi_{i_m})|0\rangle :\Phi_{i_{m+1}}...\Phi_{i_n}:, \tag{59}$$

where the sum goes over all possible ways to pair the fields. The Wick theorem (59) expresses *time-ordered* products of fields in terms of *normal-ordered* ones and propagators. The normal-ordered operation puts *all* annihilation operators originating from different $\Phi$s to the right. It also cares about fermions, e.g.,

$$:a_1^- a_2^+ a_3^- a_4^- a_5^+ a_6^-: = (-1)^\sigma a_2^+ a_5^+ a_1^- a_3^- a_4^- a_6^-, \tag{60}$$

---

[11]More precisely, operators in the *interaction* picture.

**Table 1:** Feynman rules for external states.

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

where $\sigma$ correspond to the number of fermion permutations (*cf.* Eq.(57)). In Fig. 4 a cartoon, which illustrates Eq.(59) for one of the contributions to $T[\mathcal{L}_I(x)\mathcal{L}_I(y)]$, is provided.

After application of the Wick theorem we have to calculate

$$\langle 0|\tilde{a}_{\mathbf{k}_1}^- ...\tilde{a}_{\mathbf{k}_s}^- :\Phi_{i_{m+1}}...\Phi_{i_n}: \tilde{a}_{\mathbf{p}_1}^+ ...\tilde{a}_{\mathbf{p}_r}^+|0\rangle. \tag{61}$$

To get a *non-zero* matrix element, all $a^-(a^+)$ in the normal product of fields from the Lagrangian have to be "killed" by (commuted with) $a^+(a^-)$ from the initial (final) states.

For our *generalized* field (37) we have

$$\left[\Phi_\alpha^i(x), (a_{\mathbf{p}}^+)_s^i\right] = \underbrace{\frac{e^{-ipx}}{(2\pi)^{3/2}\sqrt{2\omega_p}}}_{\text{common to all fields}} u_\alpha^s(\mathbf{p}), \qquad \text{initial state polarization (particle)};$$

$$\left[(b_{\mathbf{p}}^-)_s^i, \Phi_\alpha^i(x)\right] = \frac{e^{+ipx}}{(2\pi)^{3/2}\sqrt{2\omega_p}} v_\alpha^{*s}(\mathbf{p}), \qquad \text{final state polarization (antiparticle)}. \tag{62}$$

and one clearly sees that the factors in the denominators Eq.(62) are avoided when the rescaled $\tilde{a}^\pm$ (or $\tilde{b}^\pm$) operators (51) are used.

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

$$\left.\begin{array}{l} \langle 0|T(\phi(x)\phi^\dagger(y))|0\rangle \\[2mm] \langle 0|T(\psi(x)\bar{\psi}(y))|0\rangle \\[2mm] \langle 0|T(W_\mu(x)W_\nu^\dagger(y))|0\rangle \end{array}\right\} = \int \frac{d^4p}{(2\pi)^4}\frac{ie^{-ip(x-y)}}{p^2-m^2+i\epsilon} \left\{\begin{array}{ll} 1 & \phi; \\[2mm] \hat{p}+m & \psi; \\[2mm] -g_{\mu\nu}+p_\mu p_\nu/m^2 & W_\mu. \end{array}\right. \tag{63}$$

One can notice that all the dependence on $x_i$ of the integrand in Eq.(58) comes from either Eq.(62) or Eq.(63). As a consequence, it is possible to carry out the integration for *every* $x_i$

$$\int d^4x_i e^{-ix_i(p_1+...+p_n)} = (2\pi)^4\delta^4(p_1+...+p_n) \tag{64}$$

and obtain a $\delta$-function reflecting energy-momentum conservation at the corresponding vertex.

Depending on the direction of momenta, the external lines represent incoming or outgoing particles (see Table 1). Again, the corresponding factors (=polarization vectors) are derived from $\mathcal{L}_0$. Notice that we explicitly write the Lorentz indices for vector particles and suppress the Dirac indices for

15

fermions. To keep track of the index contractions in the latter case, one uses *arrows* on the fermion lines.[12]

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

$$i \frac{\delta^4 \mathcal{A}_I[\phi]}{\delta\phi(p_1)\delta\phi(p_2)\delta\phi(p_3)\delta\phi(p_4)}\bigg|_{\phi=0} \Rightarrow \underbrace{(2\pi)^4 \delta^4(p_1 + p_2 + p_3 + p_4)}_{\text{conservation of energy-momentum}} \times [-i\lambda]. \tag{65}$$

In a typical diagram all $(2\pi)^4\delta(...)$ factors (but *one*[13]) reflecting the energy-momentum conservation at each vertex, are removed by the momentum integration originating from propagators, Eq.(63). Due to this, we also omit these factors (see, Table 2 for examples).

Given Feynman rules, we can draw all possible diagrams that contribute to a process and evaluate the amplitude. We do not provide the precise prescription here (see textbooks [12–16] for details) but just mention the fact that one should keep in mind various *symmetry* factors and relative *signs* that can appear in real calculations.

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

$$|M|^2 = MM^\dagger \Rightarrow \qquad \tag{66}$$



Sometimes we do not care about polarization states of initial or final particles, so we have to *sum* over *final* polarization and *average* over *initial* ones. That is where *spin-sum* formulas,e.g.,

$$\sum_s u_s(\mathbf{p_1})\bar{u}_s(\mathbf{p_1}) = \hat{p}_1 + m, \qquad \sum_s v_s(\mathbf{p_2})\bar{v}_s(\mathbf{p_2}) = \hat{p}_2 - m \tag{67}$$

become handy

$$MM^\dagger \rightarrow \sum_{s,r}(\bar{u}_s A v_r)(\bar{v}_r A^\dagger u_s) = \mathrm{Tr}\left[(\hat{p}_1 + m)A(\hat{p}_2 - m)A^\dagger\right]. \tag{68}$$

As a consequence, one can utilize the well-known machinery for gamma-matrix traces to evaluate probabilities in an efficient way.

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

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

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

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

---

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

[13]We factor it out in the definition of $M_{\alpha\beta}$, see Eq.(50).

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

$$\mathcal{L}_I = -y\bar{\psi}\psi\phi \qquad\qquad \mathcal{L}_I = e\bar{\psi}\gamma_\mu\psi A_\mu \qquad\qquad \mathcal{L}_I = ieA_\mu\left(\phi^\dagger\partial_\mu\phi - \phi\partial_\mu\phi^\dagger\right)$$



$$-iy \qquad\qquad\qquad ie\gamma_\mu \qquad\qquad\qquad ie(p_1+p_2)_\mu$$
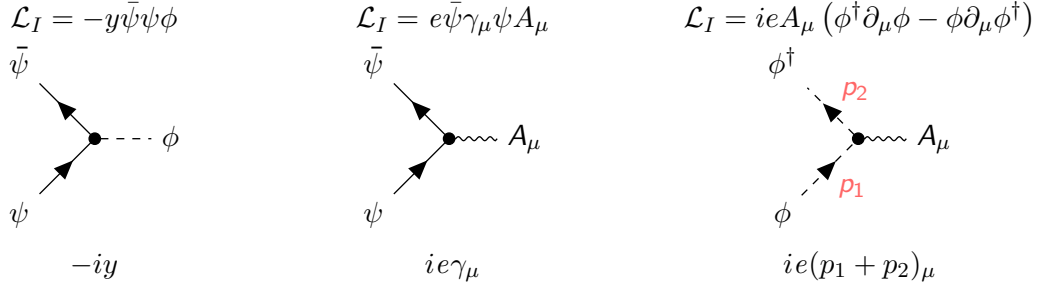
which can lead to *divergent* (meaningless?) results. This is again a manifestation of *UV* divergences due to *large* momenta ("small distances").

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

$$I_2^\Lambda(k) = i\pi^2\left[\ln\frac{\Lambda^2}{k^2} + 1\right] + \mathcal{O}\left(\frac{k^2}{\Lambda^2}\right) = i\pi^2\left[\ln\frac{\Lambda^2}{\mu^2} - \ln\frac{k^2}{\mu^2} + 1\right] + \mathcal{O}\left(\frac{k^2}{\Lambda^2}\right) \qquad (69)$$

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

$$I_2^{4-2\varepsilon}(k) = \mu^{2\varepsilon}\int\frac{d^{4-2\varepsilon}q}{q^2(k-q)^2} = i\pi^2\left(\frac{1}{\varepsilon} - \ln\frac{k^2}{\mu^2} + 2\right) + \mathcal{O}(\varepsilon). \qquad (70)$$

Both the regularized integrals are now convergent[14] and share the same logarithmic dependence on external momentum $k$. One can also notice a (one-to-one) correspondence between a *logarithmically* divergent contribution $\log\Lambda^2/\mu^2$ in Eq.(69) and the pole term $1/\varepsilon$ in Eq.(70). However, the constant terms are *different*. How do we make sense of this ambiguity?

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

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

It is obvious that working with *bare* quantities is not very convenient. One usually makes the dependence on $\Lambda$ (or $\varepsilon$) explicit by introduction of divergent $Z$-factors for *bare* fields ($\phi_B$), masses ($m_B^2$), and couplings ($\lambda_B$), e.g.,

$$\mathcal{L}_{full} = \frac{1}{2}(\partial\phi_B)^2 - \frac{m_B^2}{2}\phi_B^2 + \frac{\lambda_B\phi_B^4}{4!} = \frac{Z_2}{2}(\partial\phi)^2 - \frac{Z_m m^2}{2}Z_2\phi^2 + \frac{Z_\lambda\lambda}{4!}(Z_2\phi^2)^2 \qquad (71)$$

---

[14]We do not discuss the issue of possible IR divergences here.
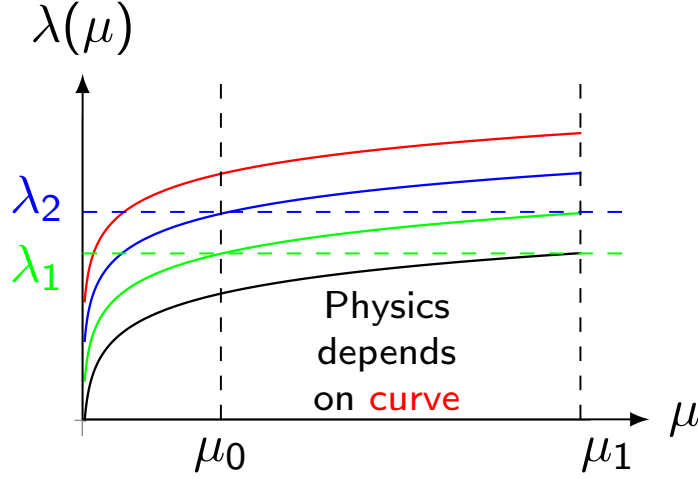
**Fig. 5:** Solutions of RGE for different boundary conditions.

$$= \frac{(\partial\phi)^2}{2} - \frac{m^2\phi^2}{2} + \frac{\lambda\phi^4}{4!} + \underbrace{\frac{(Z_2-1)}{2}(\partial\phi)^2 - \frac{(Z_m Z_2 - 1)m^2}{2}\phi^2 + (Z_4 Z_2^2 - 1)\frac{\lambda\phi^4}{4!}}_{\text{counterterms}}. \quad (72)$$

Here $\phi$, $m$ and $\lambda$ denote *renormalized* (finite) quantities. Since we can always subtract something finite from infinity, there is a certain freedom[15] in this procedure. So we have to impose additional *conditions* on $Z$s, i.e., define a *renormalization* scheme. For example, in the minimal (MS) schemes we subtract only the divergent terms, e.g., only poles in $\varepsilon$, while in the so-called momentum-subtraction (MOM) schemes we require amplitudes (more generally *Green functions*) to have a certain value at some fixed kinematics.

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

$$\text{(diagram)} = \text{(diagram)} + \text{(diagram)} + \text{permutations} + \text{more loops} \quad (73)$$

$$= \lambda_B(\Lambda) - \frac{\lambda_B(\Lambda)^2}{2(16\pi^2)}\left(\ln\frac{\Lambda^2}{\mu^2} - \ln\frac{k^2}{\mu^2} + \dots\right) + \dots \quad (74)$$

$$= \left[\lambda(\mu) + \frac{3}{2}\frac{\lambda^2(\mu)}{16\pi^2}\ln\frac{\Lambda^2}{\mu^2}\right] - \frac{\lambda(\mu)^2}{2(16\pi^2)}\left(\ln\frac{\Lambda^2}{\mu^2} - \ln\frac{k^2}{\mu^2} + \dots\right) + \dots \quad (75)$$

$$= \lambda(\mu) + \frac{\lambda(\mu)^2}{2(16\pi^2)}\left(\ln\frac{k^2}{\mu^2} + \dots\right) + \dots. \quad (76)$$

In Eq.(73) the tree-level and one-loop diagrams contributing to the matrix element are shown. The corresponding expression in terms of the bare coupling $\lambda_B(\Lambda)$ that implicitly depends on the regularization parameter $\Lambda$ is given in Eq.(74). We introduce a renormalized[16] coupling $\lambda(\mu)$ in Eq.(75) to make the dependence explicit:

$$\lambda_B(\Lambda) = \lambda(\mu)Z_\lambda = \lambda(\mu)\left(1 + \frac{3}{2}\frac{\lambda(\mu)}{16\pi^2}\ln\frac{\Lambda^2}{\mu^2} + \dots\right). \quad (77)$$

---

[15]Different constant terms in Eq.(69) and Eq.(70) are one manifestation of this fact.

[16]We use minimal subtractions here and the factor of three comes from the fact that all three one-loop graphs ($s$, $t$ and $u$) give rise to the same *divergent* term.
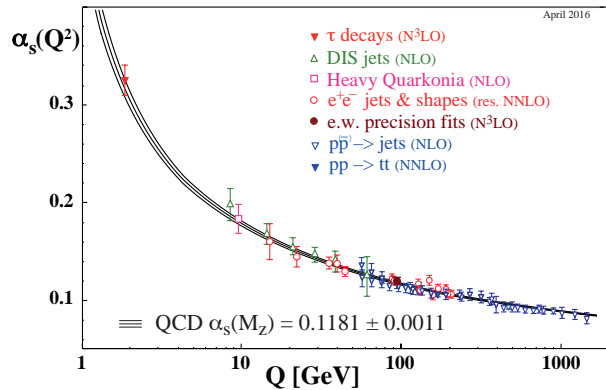
**Fig. 6:** The scale dependence of the strong coupling $\alpha_s$.

The final result (76) is finite (when $\Lambda \to \infty$) and can be confronted with experiment. It seems to depend on an auxiliary scale $\mu$, which inevitably appears in any renormalization scheme. The crucial point here is that *observables* (if all orders of PT are taken into account) actually do *not* depend on $\mu$. Changing $\mu$ corresponds to a certain reshuffling of the PT series: some terms from loop corrections are absorbed into the rescaled (*running*) couplings. This allows one to improve the "convergence"[17] of the series.

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

$$\lambda(\mu_0) \to \lambda(\mu), \quad \frac{d}{d\ln\mu}\lambda = \beta_\lambda(\lambda), \quad \beta_\lambda = \frac{3}{2}\frac{\lambda^2}{16\pi^2} + ... \tag{78}$$

The *beta-function* $\beta_\lambda$ can be calculated order-by-order in PT. However, the (initial) value $\lambda(\mu_0)$ needed to solve Eq.(78) is *not predicted* and has to be extracted from experiment.

It is worth pointing out here that two different numerical values of the *renormalized* self-coupling, $\lambda_1$ and $\lambda_2$, do not necessarily correspond to different Physics. Indeed, if they are fitted from measurements at different scales, e.g., $\mu_0$ and $\mu$, and are related by means of RGE, they represent the *same* Physics (see Fig. 5). A prominent example is the running of the strong coupling in Quantum chromodynamics (QCD) described by (see [19])

$$\beta_{\alpha_s} = -\frac{\alpha_s^2}{4\pi}\left(11 - \frac{2}{3}n_f\right) + ... + \mathcal{O}(\alpha_s^7), \quad n_f - \text{number of flavours.} \tag{79}$$

In Fig. 6 one can see a remarkable consistency between different measurements of $\alpha_s(\mu)$ and the scale dependence predicted by perturbative QCD.

### 4.2 Renormalizable or non-renormalizable?

Let us stress again that the model is called *renormalizable* if *all* the divergences that appear in loop integrals can be canceled by local counterterms due to renormalization of bare parameters and couplings from $\mathcal{L}_{full}$. But what happens if there is a divergent amplitude but the structure of the required subtraction does not have a counter-part in our initial Lagrangian, i.e., we do not have a coupling to absorb the infinity? Obviously, we can modify $\mathcal{L}_{full}$ and *add* the required term (and the coupling).

---

[17]Actually, the PT series are *asymptotic* (divergent) and we speak about the behavior of a limited number of first terms here.
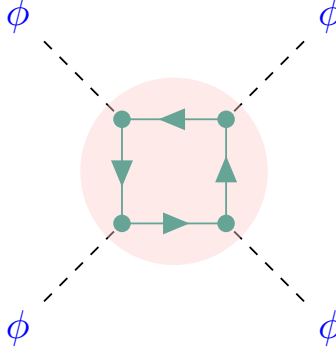
**Fig. 7:** One-loop correction to higgs self-interaction.

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

$$\mathcal{L}_I \ni \delta\mathcal{L}_Y = -y \cdot \bar{\psi}\psi\phi. \tag{80}$$

Let us assume for the moment that we set the self-coupling to zero $\lambda = 0$ and want to calculate the Higgs-scattering amplitude due to top quarks (see, Fig. 7). We immediately realize that the contribution is divergent and without $\delta\mathcal{L}_4 = -\lambda\phi^4/4!$ we are not able to cancel it. Due to this, we are forced to consider the $\phi^4$ term in a consistent theory.

Since we modified $\mathcal{L}_{full}$, we have to re-calculate all the amplitudes. In principle, new terms in $\mathcal{L}_I$ will generate new diagrams, which can require new interactions to be added to $\mathcal{L}_I$. Will this process terminate? In the case of *renormalizable* models the answer is positive. We just need to make sure that $\mathcal{L}_I$ include *all* possible terms with *dimensionless* couplings[18], or, *equivalently*, local dimension-4 *operators* built from quantum fields and their derivatives.

On the contrary, if one has to add more and more terms to $\mathcal{L}_I$, this is a signal of a *non-renormalizable* model. It looks like that we have to abandon such models since we need to measure an infinite number of couplings to predict something in this situation! However, it should be stressed that non-renormalizable models, contrary to renormalizable ones, involve couplings $G_i$ with *negative* mass dimension $[G_i] < 0$! Due to this, not all of them are important at *low* energies, such as

$$G_i E^{-[G_i]} \ll 1. \tag{81}$$

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

$$-\mathcal{L}_I = G\bar{\Psi}_p\gamma_\rho\Psi_n \cdot \bar{\Psi}_e\gamma_\rho\Psi_\nu + \text{h.c.} \tag{82}$$

in the description of the $\beta$-decay $n \to p + e^- + \bar{\nu}_e$. Since the model turns out to be a harbinger of the modern electroweak theory, let us consider it in more detail and discuss its features, which eventually lead to the construction of the SM.

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

$$-\mathcal{L}_{\text{Fermi}} = \frac{G_F}{2\sqrt{2}}(J_\mu^+ J_\mu^- + \text{h.c.}). \tag{83}$$

Here the current

$$J_\rho^- = (V - A)_\rho^{\text{nucleons}} + \overline{\Psi}_e\gamma_\rho\left(1 - \gamma_5\right)\Psi_{\nu_e} + \overline{\Psi}_\mu\gamma_\rho\left(1 - \gamma_5\right)\Psi_{\nu_\mu} + ... \tag{84}$$

---

[18]Remember the T-shirt Lagrangian?
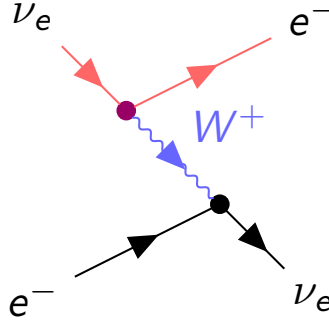
**Fig. 8:** A contribution to $\nu e$-scattering due to charged $W$-boson

is the difference between Vector ($V$) and Axial ($A$) parts. This kind of *current-current* interactions can describe not only the proton beta-decay but also the muon decay $\mu \to e\nu_\mu\bar\nu_e$ or the process of $\nu_e e$ - scattering. Since the *Fermi* constant $G_F \simeq 10^{-5}$ GeV$^{-1}$, from simple *dimensional* grounds we have

$$\sigma(\nu_e e \to \nu_e e) \propto G_F^2 s, \qquad s = (p_e + p_\nu)^2. \tag{85}$$

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

However, a modern view on the Fermi model treats it as an *effective* field theory [20] with certain *limits of applicability*. It perfectly describes low-energy experiments and one can fit the value of $G_F$ very precisely (see [21]). The *magnitude* of $G_F$ tells us something about a *more fundamental* theory (the SM in our case): around $G_F^{-1/2} \sim 10^2 - 10^3$ GeV there should be some "New Physics" (NP) to cure the above-mentioned shortcomings. Indeed, by analogy with (renormalizable) QED we can introduce *mediators* of the weak interactions – electrically charged *vector* fields $W_\mu^\pm$ (see, e.g., Fig. 8):

$$\mathcal{L}_{\text{Fermi}} = -\frac{G_F}{2\sqrt{2}}(J_\mu^+ J_\mu^- + \text{h.c.}) \to \mathcal{L}_I = \frac{g}{2\sqrt{2}}(W_\mu^+ J_\mu^- + \text{h.c.}) \tag{86}$$

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

$$T = i(2\pi)^4 \frac{g^2}{8} J_\alpha^+ \left[ \frac{g_{\alpha\beta} - p_\alpha p_\beta/M_W^2}{p^2 - M_W^2} \right] J_\beta^-. \tag{87}$$

In the limit $|p| \ll M_W$, Eq.(87) reproduces the prediction of the effective theory (Fermi model) if we identify ("match")

$$(\text{effective theory}) \quad \frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2} \quad (\text{more fundamental theory}). \tag{88}$$

However, one can see that in the UV region ($|p| \gg M_W$) the amplitude (87) still has bad behavior, leading to all the above-mentioned problems. To deal with the issue, we utilize *gauge* symmetry, which will be discussed in the next section.

## 5 Gauge symmetries

We are seeking for a model of weak interactions that has good UV-properties. Let us revise how the *gauge principle* is implemented in QED. First of all, consider

$$\mathcal{L}_0 = \bar\psi \left( i\hat\partial - m \right) \psi \tag{89}$$
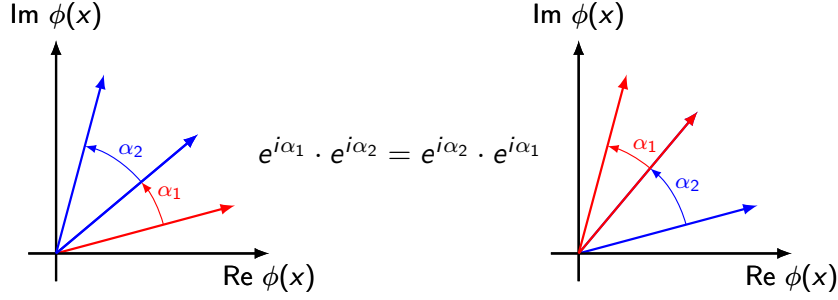
**Fig. 9:** $U(1)$ transformations commute with each other.

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

$$\psi \to \psi' = e^{ie\omega}\psi \tag{90}$$

*local*, i.e., $\omega \to \omega(x)$. In this case, the Lagrangian ceases to be invariant[19]:

$$\delta\mathcal{L}_0 = \partial_\mu\omega \cdot J_\mu, \qquad J_\mu = -e\bar{\psi}\gamma_\mu\psi, \tag{91}$$

To compensate this term, we add the interaction of the current $J_\mu$ with the photon field $A_\mu$:

$$\mathcal{L}_0 \to \mathcal{L} = \mathcal{L}_0 + A_\mu J_\mu = \bar{\psi}\left[i(\hat{\partial} + ie\hat{A}) - m\right]\psi, \qquad A_\mu \to A'_\mu = A_\mu - \partial_\mu\omega. \tag{92}$$

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

$$\mathcal{L}_{QED} = \bar{\psi}\left(i\hat{D} - m\right)\psi - \frac{1}{4}F_{\mu\nu}^2 \tag{93}$$

$$D_\mu = \partial_\mu + ieA_\mu, \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{94}$$

Here we introduce a *covariant* derivative $D_\mu$ and a *field-strength* tensor $F_{\mu\nu}$. One can check that Eq.(93) is invariant under

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

The *second* Noether theorem [17] states that theories possessing *gauge* symmetries are *redundant*, i.e., some degrees of freedom are not physical. This makes quantization non-trivial. To deal with this problem in QED, one adds a *gauge-fixing term* to the free vector-field Lagrangian:

$$\mathcal{L}_0(A) = -\frac{1}{4}F_{\mu\nu}^2 - \frac{1}{2\xi}\left(\partial_\mu A_\mu\right)^2 \equiv -\frac{1}{2}A_\mu K_{\mu\nu}A_\nu. \tag{95}$$

This term allows one to obtain the photon propagator by inverting[20] $K_{\mu\nu}$:

$$\langle 0|TA_\mu(x)A_\nu(y)|0\rangle = \int \frac{d^4p}{(2\pi)^4}\frac{-i\left[g_{\mu\nu} - (1-\xi)p_\mu p_\nu/p^2\right]}{p^2 + i\epsilon}e^{-ip(x-y)} \tag{96}$$

---

[19]Note that one can use this fact to get an expression for the Noether current $J_\mu$.

[20]If we omit the gauge-fixing term, we will not be able to invert the quadratic form.

The propagator now involves an auxiliary parameter $\xi$. It controls the propagation of *unphysical* longitudinal polarization $\epsilon_\mu^L \propto p_\mu$. The polarization turns out to be harmless in QED since the corresponding terms *drop out* of physical quantities, e.g., due to current conservation

$$e_\mu^L J_\mu \propto p_\mu J_\mu = 0 \quad \text{[we have no source for unphysical } \gamma\text{]}. \tag{97}$$

One can see that the propagator has good UV behaviour and falls down as $1/p^2$ for large $p$. The gauge symmetry of QED is $U(1)$. It is *Abelian* since the order of two transformations is irrelevant (see Fig. 9). However, if we want to apply the *gauge principle* to the case of EW interactions, we have to generalize $U(1)$ to the *Non-Abelian* case. Let us consider the $SU(n)$ group, i.e., unitary $n \times n$ matrices $U_{ij}$ depending on $n^2 - 1$ parameters $\omega^a$ and having $\det U = 1$:

$$\psi_i \to \psi_i' = U_{ij}(\omega)\psi_j, \quad U(\omega) = e^{igt^a\omega^a}. \tag{98}$$

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

$$[t^a, t^b] = if^{abc}t^c, \quad f^{abc} - \text{ structure constants }. \tag{99}$$

For *constant* $\omega^a$ the transformation (98) is a symmetry of the Lagrangian

$$\mathcal{L}_0 = \bar{\psi}_i \left( i\hat{\partial} - m \right) \psi_i, \quad i = 1, ..., n \tag{100}$$

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

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

$$(D_\mu)_{ij} = \partial_\mu \delta_{ij} - igt_{ij}^a W_\mu^a. \tag{101}$$

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

$$D_\mu'\psi' = U(\omega)(D_\mu\psi), \quad U(\omega) = e^{igt^a\omega^a}. \tag{102}$$

One can find that

$$W_\mu^a \to W_\mu'^a = W_\mu^a + \partial_\mu\omega^a + gf^{abc}W_\mu^b\omega^c \tag{103}$$

$$= W_\mu^a + (D_\mu)^{ab}\omega_b, \quad (D_\mu)^{ab} \equiv \partial_\mu\delta^{ab} - ig(-if^{abc})W_\mu^c, \tag{104}$$

where we introduce the covariant derivative (101) $D_\mu^{ab}$ with generators $(t^c)^{ab} = -if^{cab}$ in the *adjoint* representation. The field-strength tensor for each component of $W_\mu^a$ is given by the commutator

$$[D_\mu, D_\nu] = -igt^a\mathcal{F}_{\mu\nu}^a, \quad \mathcal{F}_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\mu W_\nu^a + gf^{abc}W_\mu^a W_\nu^b. \tag{105}$$

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

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

$$\mathcal{L} = \bar{\psi} \left( i\hat{D} - m \right) \psi - \frac{1}{4}\mathcal{F}_{\mu\nu}^a \mathcal{F}_{\mu\nu}^a = \mathcal{L}_0 + \mathcal{L}_I, \tag{106}$$

$$\mathcal{L}_0 = \bar{\psi} \left( i\hat{\partial} - m \right) - \frac{1}{4}F_{\mu\nu}^a F_{\mu\nu}^a, \quad F_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a, \tag{107}$$
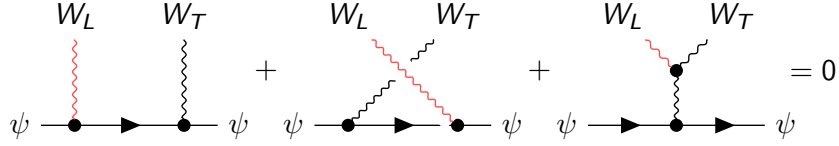
**Fig. 10:** Gauge symmetry at work: tree-level amplitudes with unphysical polarization (L) vanish.
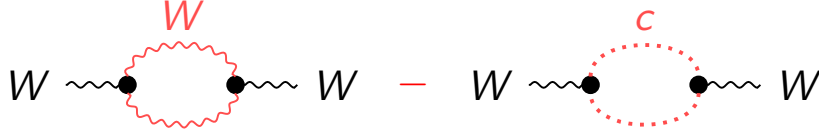


**Fig. 11:** Ghosts cancel contributions due to virtual unphysical states.

$$\mathcal{L}_I = g\bar{\psi}^i_\alpha \gamma^\mu_{\alpha\beta} t^a_{ij}\psi^j_\beta W^a_\mu - \frac{g}{2}f^{abc}W^b_\mu W^c_\nu F^a_{\mu\nu} - \frac{g^2}{4}f^{abc}f^{ade}W^a_\mu W^b_\nu W^d_\mu W^e_\nu. \tag{108}$$

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

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

$$\mathcal{L}_{gf} = -\frac{1}{2\xi}\left(F^a\right)^2, \qquad F^a = \partial_\mu W^a_\mu \tag{109}$$

with $F^a$ being a gauge-fixing function. This again introduces unphysical states in the $W^a_\mu$ propagator. However, contrary to the case of QED, the *fermionic* current $J^a_\mu = g\bar{\psi}t^a\gamma_\mu\psi$ is not conserved and can produce longitudinal $W^a_\mu$. Nevertheless, the *structure* of vector-boson self-interactions guarantees that at *tree* level amplitudes involving unphysical polarizations for external $W^a_\mu$ *vanish* (see, e.g., Fig. 10).

Unfortunately, this is not sufficient to get rid of unphysical states in loops. To deal with the problem in a *covariant* way, one introduces the so-called *Fadeev-Popov ghosts* $\bar{c}_a$ and $c_a$. They are *anticommuting* "scalars" and precisely cancel the annoying contribution. The Lagrangian for the fictitious particles is related to the gauge-fixing function $F_a(W_\mu) = \partial_\mu W^a_\mu$ via

$$\mathcal{L}_{ghosts} = -\bar{c}^a \frac{\partial F_a(W^\omega)}{\partial \omega_b} c^b = -\bar{c}^a \partial_\mu D^{ab}_\mu c^b$$
$$= -\bar{c}^a \partial^2 c^a - gf^{abc}(\partial_\mu\bar{c}^a)c^b A^b_\mu. \tag{110}$$

The ghosts are charged under $SU(n)$ and interact with gauge fields in the same way as the unphysical modes. However, there is an additional minus sign for the loops involving anticommuting ghosts (see, e.g., Fig. 11) that leads to the above-mentioned cancellations.

# 6 Gauge theory of electroweak interactions

## 6.1 Fermion couplings to gauge bosons

In the SM we use the gauge principle to introduce EW interactions. Indeed, we utilize

$$SU(2)_L \otimes U(1)_Y \tag{111}$$

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

SM fermions are charged under the group (111). To account for the $(V - A)$ pattern only *left* fermions interact with $W_\mu$ and form $SU(2)_L$ doublets:

$$L = \begin{pmatrix} \nu_l \\ l^- \end{pmatrix}_L, Q = \begin{pmatrix} q_u \\ q_d \end{pmatrix}_L, \qquad q_u = u, c, t; \; q_d = d, s, b; \; l = e, \mu, \tau. \tag{112}$$

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

$$D_\mu^L = \begin{pmatrix} \partial_\mu - \frac{i}{2} \left( g W_\mu^3 + g' Y_L^f B_\mu \right) & -i \frac{g}{\sqrt{2}} W_\mu^+ \\ -i \frac{g}{\sqrt{2}} W_\mu^- & \partial_\mu + \frac{i}{2} \left( g W_\mu^3 - g' Y_L^f B_\mu \right) \end{pmatrix}. \tag{113}$$

The *right* fermions[21] are $SU(2)_L$ singlets and do not couple to $W_\mu$:

$$D_\mu^R = \partial_\mu - i g' \frac{Y_R^f}{2} B_\mu. \tag{114}$$

The covariant derivatives involve two gauge couplings $g$, $g'$ corresponding to $SU(2)_L$ and $U(1)_Y$, respectively. Different $Y_{L/R}^f$ denote weak hypercharges of the fermions and up to now the values are not fixed. Let us put some constraints on $Y_{L/R}^f$. The first restriction comes from the $SU(2)_L$ symmetry, i.e., $Y_L^u = Y_L^d \equiv Y_L^Q$, and $Y_L^\nu = Y_L^e \equiv Y_l^L$.

One can see that the EW interaction Lagrangian

$$\mathcal{L}_W = \mathcal{L}_{NC} + \mathcal{L}_{CC}, \tag{115}$$

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

$$\mathcal{L}_{CC}^l = \frac{g}{\sqrt{2}} \bar{\nu}_L^e \gamma_\mu W_\mu^+ e_L + \text{h.c.} = \frac{g}{2\sqrt{2}} \bar{\nu}_e \gamma_\mu W_\mu^+ (1 - \gamma_5) e + \text{h.c.} \tag{116}$$

also involves *neutral-current* interactions

$$\mathcal{L}_{NC}^l = \bar{\nu}_L^e \gamma_\mu \left( \frac{1}{2} g W_\mu^3 + \frac{Y_L^l}{2} g' B_\mu \right) \nu_L^e + \bar{e}_L \gamma_\mu \left( -\frac{1}{2} g W_\mu^3 + \frac{Y_L^l}{2} g' B_\mu \right) e_L + g' \bar{e}_R \gamma_\mu \frac{Y_R^e}{2} B_\mu e_R. \tag{117}$$

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

$$\begin{aligned} W_\mu^3 &= Z_\mu \cos\theta_W + A_\mu \sin\theta_W \\ B_\mu &= -Z_\mu \sin\theta_W + A_\mu \cos\theta_W. \end{aligned} \tag{118}$$

Here we introduce the *Weinberg* angle $\theta_W$. One can try to fix $\sin\theta_W$ and various $Y_{L/R}^f$ from the requirement that, e.g., $A_\mu$ has the same interactions as the photon in QED. Indeed, given fermion *electric* charges $Q_f$ (see Fig. 1) in the units of the elementary charge $e$, one can derive the following relations:

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

---

[21]In what follows we do not consider right-handed neutrino and refer again to Ref. [18].

$$g' Y_R^f \cos\theta_W = 2eQ_f, \qquad f = e,\, u,\, d. \tag{119}$$

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

$$Y_L^l = -3Y_L^Q, \quad Y_R^e = -6Y_L^Q, \quad Y_R^u = 4Y_L^Q, \quad Y_R^d = -2Y_L^Q \tag{120}$$

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

$$\mathcal{L}_{NC} \ni \bar{f} \left[ \left( g T_3^f \sin\theta_W + g' \frac{Y_f^L}{2} \cos\theta_W \right) P_L + \left( g' \frac{Y_f^R}{2} \cos\theta_W \right) P_R \right] \gamma_\mu f A_\mu \tag{121}$$

$$= e\bar{f} \left( T_3 + \frac{Y}{2} \right) \gamma_\mu f A_\mu = e Q_f \bar{f} \gamma_\mu f A_\mu, \tag{122}$$

where in Eq.(122) we assume that $T_3$ and $Y$ are operators, which give $T_3^f$ and $Y_L^f$, when acting on left components, and $T_3 = 0$ and $Y_R^f = 2Q_f$ for right fermions.

The relations (120) allow one to rewrite the neutral-current Lagrangian as

$$\mathcal{L}_{NC} = e J_\mu^A A^\mu + \frac{g}{\cos\theta_W} J_\mu^Z Z_\mu, \tag{123}$$

where the photon $A_\mu$ and a new $Z$-boson couple to the currents of the form

$$J_\mu^A = \sum_f Q_f \bar{f} \gamma_\mu f, \qquad J_\mu^Z = \frac{1}{4} \sum_f \bar{f} \gamma_\mu \left( v_f - a_f \gamma_5 \right) f, \tag{124}$$

$$v_f = 2T_3^f - 4Q_f \sin^2\theta_W, \qquad a_f = 2T_3^f, \tag{125}$$

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

$$v_u = 1 - \frac{8}{3} \sin^2\theta_W, \qquad a_u = 1. \tag{126}$$

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

$$\mathcal{L}_{CC} = \frac{g}{\sqrt{2}} \left( J_\mu^+ W^{+\mu} + J_\mu^- W^{-\mu} \right), \quad J_\mu^+ = \frac{1}{2} \sum_f \bar{f}_u \gamma_\mu \left( 1 - \gamma_5 \right) f_d, \tag{127}$$

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

It turns out that it was *a prediction* of the electroweak SM that there should be an additional neutral gauge boson $Z_\mu$. Contrary to the photon, the $Z$-boson also interacts with neutrinos. This crucial property was used in the experiment called *Gargamelle* at CERN, where in 1973 the discovery was presented (Fig. 8). About ten years later both $W$ and $Z$ were directly produced at Super Proton Synchrotron (SPS) at CERN. Finally, in the early 90s a comprehensive analysis of the $e^+ e^- \to f\bar{f}$ process, which was carried out at the Large Electron Proton (LEP) Collier (CERN) and at the Standford Linear Collider (SLAC) confirmed the SM predictions for the $Z$ couplings to fermions (125).

It is also worth mentioning the fact that the (hyper)-charge assignment (120) satisfies very nontrivial constraints related to cancellation of *gauge anomalies*. Anomalies correspond to situations when
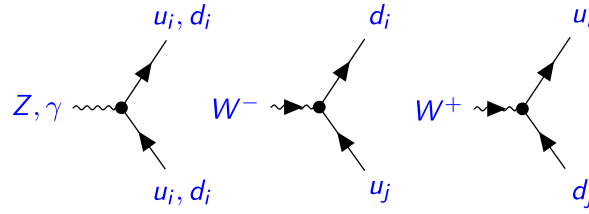
**Fig. 12:** Gauge-boson–quark vertices. Leptons interact with the EW bosons in the same way.
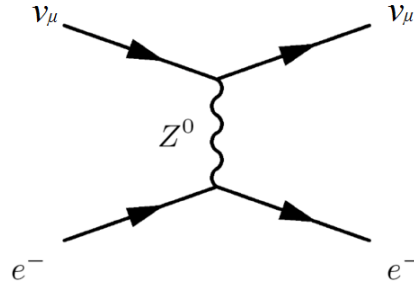


**Fig. 13:** The chamber of Gargamelle at CERN (left), $\nu_\mu$ scattering due to $Z$-boson (right). From Wikipedia.

a symmetry of the classical Lagrangian is violated at the quantum level. A well-known example is *Axial or Chiral or Adler–Bell–Jackiw(ABJ)* anomaly when the classical conservation law for the axial current $J_\mu^A$ is modified due to quantum effects:

$$J_\mu^A = \bar{\Psi}\gamma_\mu\gamma_5\Psi, \qquad \partial_\mu J_\mu^A = 2im\Psi\gamma_5\Psi + \underbrace{\frac{\alpha}{2\pi}F_{\mu\nu}\tilde{F}_{\mu\nu}}_{\text{anomaly}}, \qquad \tilde{F}_{\mu\nu} = 1/2\epsilon_{\mu\nu\rho\sigma}F_{\rho\sigma}. \qquad (128)$$

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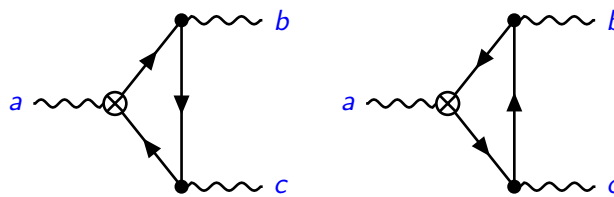


**Fig. 14:** Diagrams contributing to the anomaly of an axial current (crossed vertex).

There is nothing wrong when the anomalous current $J_\mu^A$ corresponds to a global symmetry and does not enter into $\mathcal{L}$. It just implies that a classically forbidden processes may actually occur in the quantum theory. For example, it is the anomaly in the *global* axial *flavour* symmetry that is responsible for the decay $\pi \to \gamma\gamma$. On the contrary, if an axial current couples to a gauge field, anomalies break gauge invariance, thus rendering the corresponding QFT inconsistent. In the SM left and right fermions (eigenvectors of $\gamma_5$) have different $SU(2)_L \times U(1)_Y$ quantum numbers, leaving space for potential anomalies. However, since we have to take into account all fermions which couple to a gauge field, there is a possibility that contributions from different species cancel each other due to a special assignment of
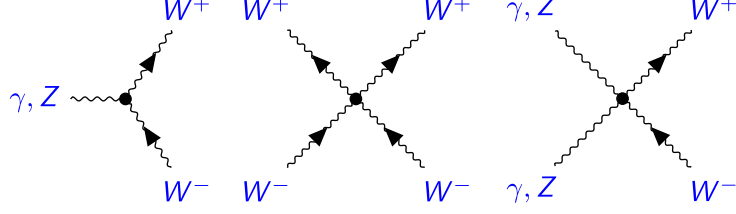
**Fig. 15:** Gauge-boson self-interaction vertices.

fermion charges. Indeed, in the case of chiral theories[22], anomalies are proportional to ($\gamma_5 = P_R - P_L$)

$$\text{Anom} \propto \text{Tr}[t^a, \{t^b, t^c\}]_L - \text{Tr}[t^a, \{t^b, t^c\}]_R, \tag{129}$$

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

$$0 = 2Y_L^Q - Y_R^u - Y_R^d, \qquad\qquad U(1)_Y - SU(3)_c - SU(3)_c, \tag{130a}$$

$$0 = N_c Y_L^Q + Y_L^l, \qquad\qquad U(1)_Y - SU(2)_L - SU(2)_L, \tag{130b}$$

$$0 = N_c \left[ 2(Y_L^Q)^3 - (Y_R^u)^3 - (Y_R^d)^3 \right] + \left[ 2(Y_L^l)^3 - (Y_R^e)^3 \right], \qquad U(1)_Y - U(1)_Y - U(1)_Y, \tag{130c}$$

$$0 = N_c \left[ 2Y_L^Q - Y_R^u - Y_R^d \right] + \left[ 2Y_L^l - Y_R^e \right], \qquad\qquad U(1)_Y - grav. - grav., \tag{130d}$$

where, in addition to the EW gauge group, we also consider strong interactions of quarks that have $N_c = 3$ colours[23]. While the first three conditions come from the SM interactions, the last one (130d) is due to the coupling to gravity. Other anomalies are trivially zero. One can see that the hypercharges introduced in Eq.(120) do satisfy the equations. It is interesting to note that contributions due to colour quarks miraculously cancel those of leptons and the cancellation works within a single generation. This put a rather strong restriction on possible new fermions that can couple to the SM gauge bosons: new particles should appear in a complete generation (quarks + leptons) in order not to spoil anomaly cancellation within the SM. Moreover, the anomaly cancellation condition can select viable models that go beyond the SM (BSM).

## 6.2 Properties of the EW gauge bosons

Due to the non-Abelian nature of the $SU(2)_L$ group, the gauge fields $W_i$ have triple and quartic self-interactions (see Eq.(108)). Since $W_3$ is a linear combination of the $Z$-boson and photon, the same is true for $Z$ and $\gamma$. In Fig. 15, self-interaction vertices for the EW gauge bosons are depicted.

The triple vertices $WW\gamma$ and $WWZ$ predicted by the SM were tested at LEP2 in the $e^+e^- \to W^+W^-$ process (Fig. 16) and agreement with the SM predictions was found. Subsequent studies at hadron colliders (Tevatron and LHC) aimed at both quartic and triple gauge couplings (QGC and TGC, respectively) also show consistency with the SM and put limits on possible deviations (so-called anomalous TGC and QGC).

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

---

[22] Chiral theories distinguish left and right fermions.

[23] In the SM coloured quarks belong to the fundamental representation of the corresponding gauge group $SU(3)_c$.
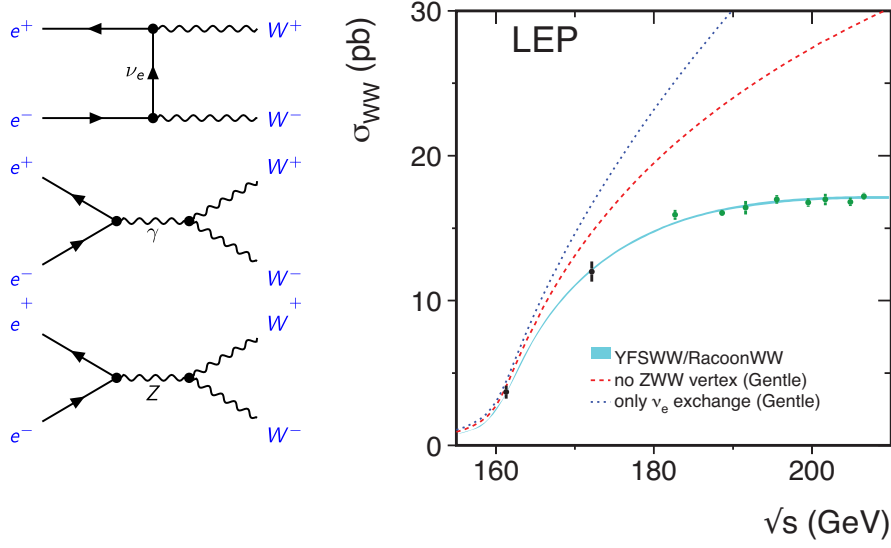
**Fig. 16:** $e^+e^- \to W^+W^-$.

measured by the parameter $\rho$:

$$\rho \equiv \frac{M_W^2}{M_Z^2 \cos^2 \theta_W}. \tag{131}$$

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

The fact that both $W$ and $Z$ should be massive poses a serious problem for theoretical description of the EW interactions. First of all, the naive introduction of the corresponding mass terms breaks the *gauge* symmetry (111). For example, $m_W^2 W_\mu^+ W_\mu^-$ is forbidden due to $W_\mu \to W_\mu + \partial_\mu \omega + ....$. One can also mention an issue with unitarity, which arises in the scattering of longitudinal EW bosons due to gauge self-interactions in Fig. 15.

In addition, the symmetry also forbids *explicit* mass terms for fermions, since e.g., $m_\mu(\bar{\mu}_L \mu_R +$ h.c.), which accounts for muon mass, mixes left and right fields that transform differently under the electroweak group (111). In the next section, we discuss how these problems can be solved by coupling the SM fermions and gauge bosons to the scalar (Higgs) sector (see also [22]).

### 6.3 Spontaneous symmetry breaking and gauge-boson masses

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

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial_\mu \phi - V(\phi^\dagger \phi) - \frac{1}{4} F_{\mu\nu}^2 + ie\left(\phi^\dagger \partial_\mu \phi - \phi \partial_\mu \phi^\dagger\right) A_\mu + e^2 A_\mu A_\mu \phi^\dagger \phi \equiv \mathcal{L}_1, \tag{132}$$

which is invariant under $U(1)$

$$\phi \to e^{ie\omega(x)}\phi, \quad A_\mu \to A_\mu + \partial_\mu \omega. \tag{133}$$

In Eq.(132) a *complex* scalar $\phi$ interacts with the photon $A_\mu$. We can use *polar* coordinates to rewrite the Lagrangian in terms of new variables

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \rho)^2 + \frac{e^2 \rho^2}{2}\left(A_\mu - \frac{1}{e}\partial_\mu \theta\right)\left(A_\mu - \frac{1}{e}\partial_\mu \theta\right) - V(\rho^2/2) - \frac{1}{4}F_{\mu\nu}^2, \tag{134}$$
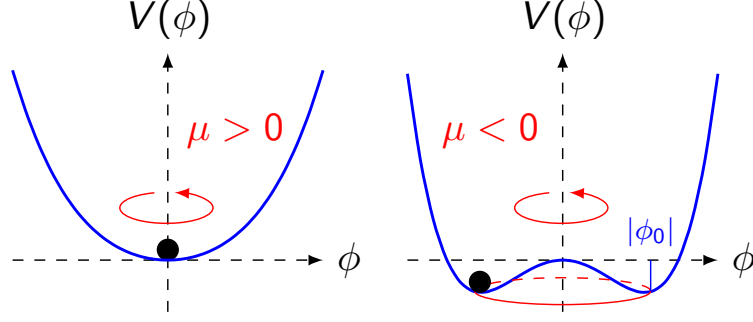
**Fig. 17:** A symmetric vacuum (left) and degenerate vacua (right).

$$= \frac{1}{2}(\partial_\mu \rho)^2 + \frac{e^2 \rho^2}{2} B_\mu B_\mu - V(\rho^2/2) - \frac{1}{4} F_{\mu\nu}^2(B), \tag{135}$$

where $\rho$ is gauge invariant, while the $U(1)$ transformation (133) gives rise to a *shift* in $\theta$:

$$\phi = \frac{1}{\sqrt{2}} \rho(x) e^{i\theta(x)}, \quad \rho \to \rho, \quad \theta \to \theta + e\omega. \tag{136}$$

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

If in Eq.(132) we replace our *dynamical* field $\rho(x)$ by a constant $\rho \to v = \text{const}$, we get the mass term for $B_\mu$. This can be achieved by considering the potential $V(\phi)$ of the form (written in terms of initial variables)

$$V = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{137}$$

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

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

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

$$\left. \frac{\partial V}{\partial \phi^\dagger} \right|_{\phi=\phi_0} = 0 \Rightarrow \phi_0^\dagger \phi_0 = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2} > 0 \Rightarrow \phi_0 = \frac{v}{\sqrt{2}} e^{i\beta}, \tag{138}$$

which are *related* by *global* $U(1)$ transformations (133) that change $\beta \to \beta + e\omega$. So, in spite of the fact that we do not break the symmetry *explicitly*, it is *spontaneously broken* (SSB) due to a particular choice of our solution ($\beta$).

In QFT we interpret $\phi_0$ as a characteristic of our *vacuum* state, i.e., as a *vacuum expectation value* (vev) or *condensate* of the quantum field:

$$\phi_0 = \langle 0|\phi(x)|0\rangle \stackrel{\beta=0}{=} \frac{v}{\sqrt{2}}. \tag{139}$$

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

$$\phi(x) = \frac{v + h(x)}{\sqrt{2}} e^{i\zeta(x)/v}, \quad \langle 0|h(x)|0\rangle = 0, \quad \langle 0|\zeta(x)|0\rangle = 0. \tag{140}$$

As a consequence, Eq.(135) can be rewritten as

$$\mathcal{L} = \frac{1}{2}(\partial_\mu h)^2 + \frac{e^2 v^2}{2}\left(1 + \frac{h}{v}\right)^2 B_\mu B_\mu - V(h) - \frac{1}{4}F_{\mu\nu}^2(B) \equiv \mathcal{L}_2, \tag{141}$$

$$V(h) = -\frac{|\mu|^2}{2}(v+h)^2 + \frac{\lambda}{4}(v+h)^4 = \frac{2\lambda v^2}{2}h^2 + \lambda v h^3 + \frac{\lambda}{4}h^4 - \frac{\lambda}{4}v^4. \tag{142}$$

One can see that the Lagrangian (142) describes a massive vector field $B_\mu$ with $m_B^2 = e^2 v^2$ and a massive scalar $h$ with $m_h^2 = 2\lambda v^2$. We do not break the symmetry explicitly. It is again *hidden* in the relations between couplings and masses. This is the essence of the *Brout-Englert-Higgs-Hagen-Guralnik-Kibble* mechanism [23–25].

The Lagrangians $\mathcal{L}_1$ (132) and $\mathcal{L}_2$ (142) describe the same Physics but written in terms of different quantities (variables). Expression (132) involves a *complex* scalar $\phi$ with 2 (real) degrees of freedom (DOFs) and a *massless* gauge field ($A_\mu$) also having 2 DOFs. It is manifestly gauge invariant but not suitable for perturbative expansion ($\phi$ has imaginary mass).

On the contrary, in $\mathcal{L}_2$ the gauge symmetry is hidden[24] and it is written in terms of *physical* DOFs, i.e., a *real* scalar $h$ (1 DOF) and a *massive* vector $B_\mu$ (3 DOFs). In a sense, one *scalar* DOF ($\zeta$) is "eaten" by the gauge field to become massive. It is important to note that the postulated *gauge* symmetry allows us to avoid the consequences of the *Goldstone* theorem, which states that if the vacuum breaks a *global* continuous symmetry there is a *massless* boson (Nambu-Goldstone) in the spectrum[25]. This boson is associated with 'oscillations" along the valley, i.e., in the *broken* direction (see Fig. 17). However, due to the local character of symmetry, $\chi$ is not physical anymore, its disappearance (or appearance, see below) reflects the *redundancy*, which was mentioned above.

In Sec. 4.2, we demonstrated that the massive-vector propagator has rather bad UV behavior and is not very convenient for doing calculations in PT. It looks like we gain nothing from the gauge principle. But it is not true. We can write the model Lagrangian in the *Cartesian* coordinates $\phi = \frac{1}{\sqrt{2}}(v + \eta + i\chi)$:

$$\mathcal{L}_3 = -\frac{1}{4}F_{\mu\nu}F_{\mu\nu} + \frac{e^2 v^2}{2}A_\mu A_\mu + \frac{1}{2}\partial_\mu\chi\,\partial_\mu\chi - \underline{ev A_\mu \partial_\mu\chi} + \frac{1}{2}\partial_\mu\eta\,\partial_\mu\eta - \frac{2v^2\lambda}{2}\eta^2 + \frac{v^4\lambda}{4} \tag{143}$$

$$+ eA_\mu\chi\partial_\mu\eta - eA_\mu\eta\partial_\mu\chi - v\lambda\eta(\eta^2 + \chi^2) - \frac{\lambda}{4}(\eta^2 + \chi^2)^2 + \frac{e^2}{2}A_\mu A_\mu(2v\eta + \eta^2 + \chi^2). \tag{144}$$

The "free" part (143) of $\mathcal{L}_3$ seems to describe 5 real DOFs: a massive scalar $\eta$, a *massless* (would-be *Nambu-Goldstone*) boson $\chi$ and a massive $A_\mu$. However, there is a mixing between the *longitudinal* component of $A_\mu$ and $\chi$ that spoils this naive counting (unphysical $\chi$ is "partially eaten" by $A_\mu$) .

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

$$\delta\mathcal{L}_{g.f.} = -\frac{1}{2\xi}(\partial_\mu A_\mu + ev\xi\chi)^2 = -\frac{1}{2\xi}(\partial_\mu A_\mu)^2 - \underline{ev\chi\partial_\mu A_\mu} - \frac{e^2 v^2\xi}{2}\chi^2. \tag{145}$$

It removes the mixing from Eq.(143) and introduces a mass for $\chi$, $m_\chi^2 = (e^2 v^2)\xi$. In addition, the vector-boson propagator in this case looks like

$$\langle 0|T A_\mu(x) A_\nu(y)|0\rangle = \int \frac{d^4p}{(2\pi)^4} \frac{-i\left[g_{\mu\nu} - (1-\xi)\frac{p_\mu p_\nu}{p^2 - \xi m_A^2}\right]}{p^2 - m_A^2 + i\epsilon} e^{-ip(x-y)}, \quad m_A = ev. \tag{146}$$

One can see that for $\xi \to \infty$ we reproduce Eq.(39), while for finite $\xi$ the propagator behaves like $1/p^2$ as $p \to \infty$, thus making it convenient for PT calculations.

---

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

[25]Any non-derivative interactions violate the shift symmetry $\zeta \to \zeta + ev\omega$ for $\omega = $ const.

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

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

$$SU(2)_L \times U(1)_Y \to U(1)_{em}. \tag{147}$$

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

$$\Phi = \frac{1}{\sqrt{2}} \exp\left( i\frac{\zeta_j(x)\sigma^j}{2v} \right) \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \qquad \Phi_0 \equiv \langle 0|\Phi|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{148}$$

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

$$\Phi \to \exp\left( ig\frac{\sigma^i}{2}\omega_a + ig'\frac{Y_H}{2}\omega' \right) \Phi. \tag{149}$$

We do not want to break $U(1)_{em}$ spontaneously so the vacuum characterized by the vev $\Phi_0$ should be invariant under $U(1)_{em}$, i.e., has no electric charge $Q$

$$e^{ieQ\theta}\Phi_0 = \Phi_0 \to Q\Phi_0 = 0. \tag{150}$$

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

$$Q\Phi_0 = \left( T_3 + \frac{Y}{2} \right)\Phi_0 = \frac{1}{2}\begin{pmatrix} 1+Y_H & 0 \\ 0 & -1+Y_H \end{pmatrix}\begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \overset{?}{=} 0. \tag{151}$$

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

$$\mathcal{L}_\Phi = (D_\mu\Phi)^\dagger(D_\mu\Phi) - V(\Phi), \quad \text{with} \quad V(\Phi) = m_\Phi^2 \Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2. \tag{152}$$

For $m_\Phi^2 < 0$ the symmetry is spontaneously broken. In the *unitary* gauge (Goldstone bosons are gauged away: in Eq.(148) we put $\zeta_j = 0$) the first term in Eq.(152) can be cast into

$$|D_\mu\Phi|^2 = \frac{1}{2}(\partial_\mu h)^2 + \frac{g^2}{8}(v+h)^2|W_\mu^1 + iW_\mu^2|^2 + \frac{1}{8}(v+h)^2(gW_\mu^3 - g'Y_H B_\mu)^2 \tag{153}$$

$$= \frac{1}{2}(\partial_\mu h)^2 + \frac{g^2}{4}(v+h)^2 W^+ W^- \quad \left[ \sqrt{2}W_\mu^\pm = W_\mu^1 \mp iW_\mu^2 \right]$$

$$+ \frac{1}{8}(v+h)^2 \left[ Z_\mu(g\cos\theta_W + g'\sin\theta_W) + A_\mu(g\sin\theta_W - g'\cos\theta_W) \right]^2 \tag{154}$$

$$= \frac{1}{2}(\partial_\mu h)^2 + M_W^2\left(1 + \frac{h}{v}\right)^2 W^+ W^- + \frac{M_Z^2}{2}\left(1 + \frac{h}{v}\right)^2 Z_\mu Z_\mu, \tag{155}$$

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

$$g\sin\theta_W - g'\cos\theta_W = 0 \quad \Rightarrow \quad \sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}}, \cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}} \tag{156}$$
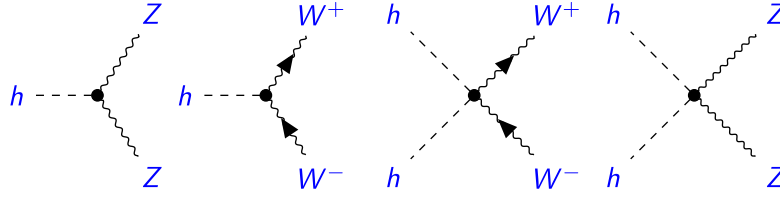
**Fig. 18:** Gauge-boson–Higgs interactions.

and, consequently,

$$g\cos\theta_W + g'\sin\theta_W = \sqrt{g^2 + g'^2}, \qquad e = g\sin\theta_W = g'\cos\theta_W = \frac{gg'}{\sqrt{g^2 + g'^2}}. \tag{157}$$

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

$$M_W^2 = \frac{g^2 v^2}{4}, \quad M_Z^2 = \frac{(g^2 + g'^2)v^2}{4}. \tag{158}$$

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

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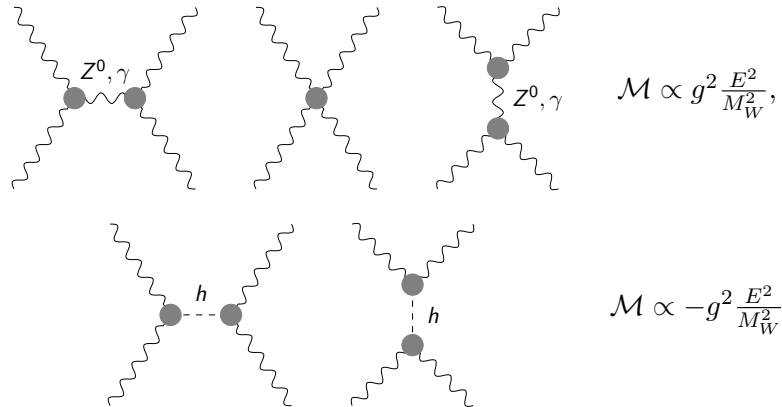


**Fig. 19:** WW-scattering and Unitarity.

Having in mind Eq.(88), one can derive the relation

$$G_F = \frac{1}{\sqrt{2}v^2} \Rightarrow v \simeq 246\,\text{GeV}, \tag{159}$$

which gives a numerical estimate of $v$. One can also see that due to (158) we have (at the tree level)

$$\rho = \frac{M_W^2}{M_Z^2 \cos^2 \theta_W} = 1. \tag{160}$$

Let us emphasize that it is a consequence of the fact that the SM Higgs is a weak *doublet* with *unit* hypercharge. Due to this, $\rho \simeq 1$ imposes important constraints on possible extensions of the SM Higgs sector. For example, we can generalize expression (160) to account for $n$ scalar $(2I_i + 1)$-plets ($i = 1, ..., n$) that transform under $SU(2)_L$ and have hypercharges $Y_i$. In case they acquire vevs $v_i$, which break the EW group, we have

$$\rho = \frac{\sum_i (I_i(I_i + 1) - Y_i^2) v_i^2}{\sum_i 2Y_i^2 v_i^2}. \tag{161}$$

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

### 6.4 Fermion-higgs interactions and masses of quarks and leptons

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

$$\mathcal{L}_Y = -y_e \underset{+1}{(\bar{L}} \underset{+1}{\Phi}) e_R - y_d \underset{-\frac{1}{3}}{(\bar{Q}} \underset{+1}{\Phi}) d_R - y_u \underset{-\frac{1}{3}}{(\bar{Q}} \underset{-1}{\Phi^c}) u_R + \text{h.c.}, \tag{162}$$

which involves *dimensionless* Yukawa couplings $y_f$. It describes interactions between the Higgs field $\Phi$, left fermion doublets (112) and right singlets. In Eq.(162) we also indicate weak hypercharges of the corresponding fields. One can see that combinations of two doublets, $(\bar{Q}\Phi)$ etc., are invariant under $SU_L(2)$ but have a non-zero charge under $U(1)_Y$. The latter is compensated by hypercharges of right fermions. In addition, $U(1)_Y$ symmetry forces us to use a charge-conjugated Higgs doublet $\Phi^c = i\sigma_2\Phi^*$ with $Y = -1$ to account for Yukawa interactions involving $u_R$.

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

$$-\mathcal{L}_Y = \sum_f \frac{y_f v}{\sqrt{2}} \left(1 + \frac{h}{v}\right) \bar{f}f = \sum_f m_f \left(1 + \frac{h}{v}\right) \bar{f}f, \quad f = u, \, d, \, e, \tag{163}$$

where unitary gauge is utilized. One can see that SSB generates fermion masses $m_f$ and, similarly to Eq.(155), *relates* them to the corresponding couplings of the Higgs boson $h$ (see Fig.20a).
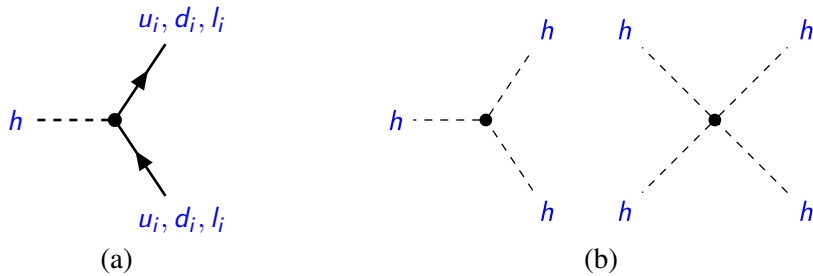


**Fig. 20:** Higgs–fermion couplings (a) and self-interactions of the Higgs boson (b).

It is worth noting that Eq.(162) is not the most general renormalizable Lagrangian involving the SM scalars and fermions. One can introduce *flavour* indices and non-diagonal *complex* Yukawa matrices $y_f^{ij}$ to account for a possible mixing between the SM fermions, i.e.,

$$\mathcal{L}_{\text{Yukawa}} = -y_l^{ij}(\bar{L}_i\Phi)l_{jR} - y_d^{ij}(\bar{Q}_i\Phi)d_{jR} - y_u^{ij}(\bar{Q}_i\Phi^c)u_{jR} + \text{h.c.}. \tag{164}$$

Substituting $\Phi \to \Phi_0$ we derive the expression for fermion mass matrices $m_f^{ij} = y_f^{ij} \frac{v}{\sqrt{2}}$, which can be diagonalized by suitable unitary rotations of left and right fields. In the SM the Yukawa matrices (164) are also diagonalized by the *same* transformations. This leads again (in the unitary gauge) to Eq.(163) but with the fields corresponding to the *mass* eigenstates. The latter *do not* coincide with *weak* states, which enter into $\mathcal{L}_W$ (115). However, one can rewrite $\mathcal{L}_W$ in terms of mass eigenstates. Due to large *flavour symmetry* of weak interactions[26], this introduces a single mixing matrix (the Cabibbo–Kobayashi–Maskawa matrix, or CKM), which manifests itself in the charged-current interactions $\mathcal{L}_{CC}$. A remarkable fact is that three generations are *required* to have $\mathcal{CP}$ violation in the quark sector. Moreover, a single CKM with only four physical parameters (angles and one phase) proves to be very successful in accounting for plethora of phenomena involving transitions between different flavours. We will not discuss further details but refer to the dedicated lectures on Flavor Physics [26].

## 7 The SM: theory vs. experiment

Let us summarize and write down the full SM Lagrangian as

$$\mathcal{L}_{SM} = \mathcal{L}_{Gauge}(g_s, g, g') + \mathcal{L}_{Yukawa}(y_u, y_d, y_l) + \mathcal{L}_{Higgs}(\lambda, m_\Phi^2) + \mathcal{L}_{Gauge\text{-}fixing} + \mathcal{L}_{Ghosts}. \quad (165)$$

The Yukawa part $\mathcal{L}_{Yukawa}$ is given in Eq.(164), while $\mathcal{L}_{Higgs} = -V(\Phi)$ is the Higgs potential from Eq.(152). After SSB the corresponding terms give rise to the Higgs couplings to the SM fermions (Fig.20a) and Higgs self-interactions (Fig.20b). The former are diagonal in the *mass* basis. The kinetic term for the Higgs field is included in

$$\mathcal{L}_{Gauge} = -\frac{1}{4} \underbrace{G_{\mu\nu}^a G_{\mu\nu}^a}_{SU(3)_c} - \frac{1}{4} \underbrace{W_{\mu\nu}^i W_{\mu\nu}^i}_{SU(2)_L} - \frac{1}{4} \underbrace{B_{\mu\nu} B_{\mu\nu}}_{U(1)_Y} + (D_\mu \Phi)^\dagger (D_\mu \Phi) \quad (166)$$

$$+ \underbrace{\bar{L}_i \, i\hat{D} \, L_i + \bar{Q}_i \, i\hat{D} \, Q_i}_{SU(2)_L \text{ doublets}} + \underbrace{\bar{l}_{Ri} \, i\hat{D} \, l_{R_i} + \bar{u}_{Ri} \, i\hat{D} u_{R_i} + \bar{d}_{Ri} \, i\hat{D} \, d_{R_i}}_{SU(2)_L \text{ singlets}}, \quad (167)$$

where for completeness we also add the colour group $SU(3)_c$ responsible for the strong force. The first three terms in Eq.(166) introduce gauge bosons for the SM gauge groups and in the non-Abelian case account for self-interactions of the latter (Fig. 15). The fourth term in (166) written in the form (155) accounts for gauge interactions of the Higgs field (Fig. 18). Finally, Eq.(167) gives rise to interactions between gauge bosons and the SM fermions (see, e.g., Fig. 12).

The SM Lagrangian Eq.(165) depends on 18 physical[27] parameters — 17 dimensionless couplings (gauge, Yukawa, and scalar self-interactions) and only 1 mass parameter $m_\Phi^2$ (see Table. 3). It is worth emphasizing here that there is certain freedom in the definition of *input* parameters. In principle, one can write down the SM predictions for a set of 18 observables (e.g., physical particle masses or cross-sections at fixed kinematics) that can be measured in experiments. With the account of loop corrections the predictions become non-trivial functions of *all* the Lagrangian parameters. By means of PT it is possible to invert these relations and express these primary parameters in terms of the chosen measured quantities. This allows us to *predict* other *observables in terms of* a finite set of measured *observables*[28].

However, it is not always practical to strictly follow this procedure. For example, due to confinement we are not able to directly probe the strong coupling $g_s$ and usually treat it as a scale-dependent parameter $(4\pi)\alpha_s = g_s^2$ defined in the modified minimal-subtraction ($\overline{\text{MS}}$) scheme. It is customary to use the value of $\alpha_s^{(5)}(M_Z) = 0.1181 \pm 0.011$ at the $Z$-mass scale as an input for theoretical predictions. A convenient choice of other input parameters is presented in Table.3. It is mostly dictated by the fact

---

[26]In the SM the symmetry is $U(3)^5$ and corresponds to flavour rotations of left doublets, $Q$ and $L$, and right singlets, $u_R$, $d_R$ and $l_R$. Neutrinos are assumed to be massless.

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

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

**Table 3:** Parameters of the SM.

| 18= | 1 | 1 | 1 | 1 | 1 | 9 | 4 |
|---|---|---|---|---|---|---|---|
| primary: | $g_s$ | $g$ | $g'$ | $\lambda$ | $m_\Phi^2$ | $y_f$ | $y_{ij}$ |
| practical: | $\alpha_s$ | $M_Z^2$ | $\alpha$ | $M_H^2$ | $G_F$ | $m_f$ | $V_{CKM}$ |

that the parameters from the "practical" set are measured with better precision than the others.

At the *tree* level one can write

$$\alpha_s = \frac{g_s^2}{4\pi}, \quad (4\pi)\alpha = g^2 g'^2/(g^2 + g'^2), \quad M_Z^2 = \frac{(g^2+g'^2)v^2}{4},$$

$$G_F = \frac{1}{\sqrt{2}v^2}, \quad M_h^2 = 2\lambda v^2 = 2|m_\Phi|^2, \quad m_f = y_f v/\sqrt{2}. \tag{168}$$

The relations are modified at higher orders in PT and perturbative corrections turn out to be mandatory if one wants to confront theory predictions [27–29] with high-precision experiments. A simple example to demonstrate this fact comes from the tree-level "prediction" for the $W$-mass $M_W$. From Eq.(158) and Eq.(168) we can derive

$$\frac{G_F}{\sqrt{2}} = \frac{\pi\alpha}{2M_W^2(1 - M_W^2/M_Z^2)}. \tag{169}$$

Plugging recent PDG [21] values

$$\alpha^{-1} = 137.035999139(31), \ M_Z = 91.1876(21) \ \text{GeV}, \ G_F = 1.1663787(6) \times 10^{-5} \ \text{GeV}^{-2}, \tag{170}$$

in Eq.(169), one can predict

$$M_W^{tree} = 80.9387(25) \ \text{GeV}, \tag{171}$$

where only uncertainties due to the input parameters (170) are taken into account. Comparing $M_W^{tree}$
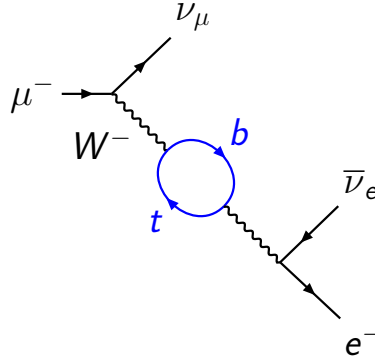


**Fig. 21:** An example of loop corrections to the muon decay, which give rise to the modification of the tree-level relation in Eq.(169).

with the measured value $M_W^{exp} = 80.379(12)$ GeV, one can see that our naive prediction is off by about $47\sigma$! Of course, this is not the reason to abandon the SM. We just need to take radiative corrections into account (see, e.g., Fig.21). Among other things the latter allows one to *connect* phenomena at different scales in the context of a single model.

A modern way to obtain the values of the SM parameters is to perform a global fit to confront state-of-the-art SM predictions with high-precision experimental data. Due to quantum effects, we can even probe New Physics that can contribute to the SM processes at low energies via virtual states. Indeed, LEP precision measurements interpreted in the context of the SM were used in a multidimensional
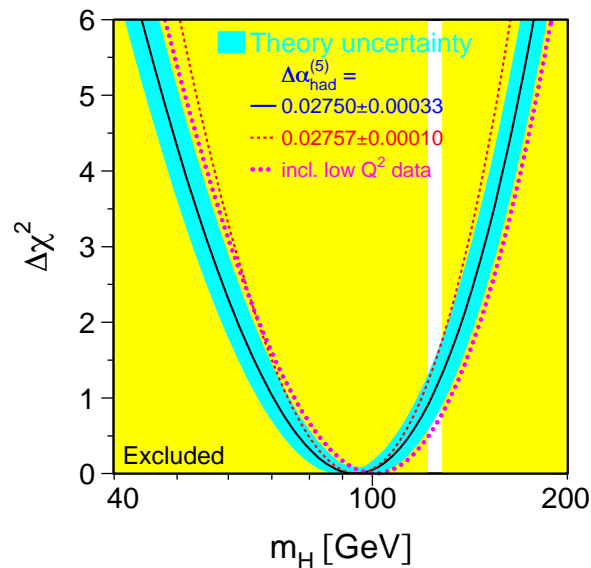
**Fig. 22:** The dependence of $\Delta\chi^2_{\min}(M_H^2) = \chi^2_{\min}(M_H^2) - \chi^2_{\min}$ on the value of $M_H$. The width of the shaded band around the curve shows the theoretical uncertainty. Exclusion regions due to LEP and LHC are also presented.

parameter fits to predict the mass of the top quark $M_t$ ("New Physics"), prior to its discovery at the Tevatron. After $M_t$ was measured it was included in the fit as an additional constraint, and the same approach led to the prediction of a *light* Higgs boson. In Fig.22, the famous *blue-band* plot by the LEP Electroweak Working Group (LEPEWWG [30]) is presented. It was prepared a couple of months before the official announcement of the Higgs-boson discovery. One can see that the best-fit value corresponding to $\Delta\chi^2_{min} = 0$ lies just about $1\sigma$ below the region *not* excluded by LEP and LHC.

Obviously, at the moment the global EW fit is *overconstrained* and can be used to test overall consistency of the SM. In Fig. 23 we present the comparison between measurements of different (pseudo)observables $O^{\text{meas}}$ and the SM predictions $O^{\text{fit}}$ corresponding to the best-fit values of fitted parameters. Although there are several quantities where *pulls*, i.e., deviations between the theory and experiment, reach more than two standard deviations, the average situation should be considered as extremely good. A similar conclusion can be drawn from the recent Figs. 24 and 25, in which experimental results for various cross-sections measured by ATLAS and CMS are compared with the SM predictions. In case one is interested in the behavior of the SM at ultra-high energies, it is more convenient to get back to the primary parameters and use the renormalization group to estimate how they change with scale. In Fig. 26, the scale dependence of the SM parameters is presented. One can see that the gauge couplings tend to converge to a single value at about $10^{13-15}$ GeV, thus providing a hint for Grand Unification. Another important consequence of this kind of studies is related to the EW vacuum (meta)stability (see, e.g., [33]). In Fig. 26, it manifests itself at the scale $\mu \simeq 10^{10}$ GeV, at which the self-coupling $\lambda$ becomes negative, making the tree-level potential unbounded from below.

## 8    Conclusions

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

–   it is based on Symmetry principles: Lorentz + $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry;

–   it is renormalizable and unitary;
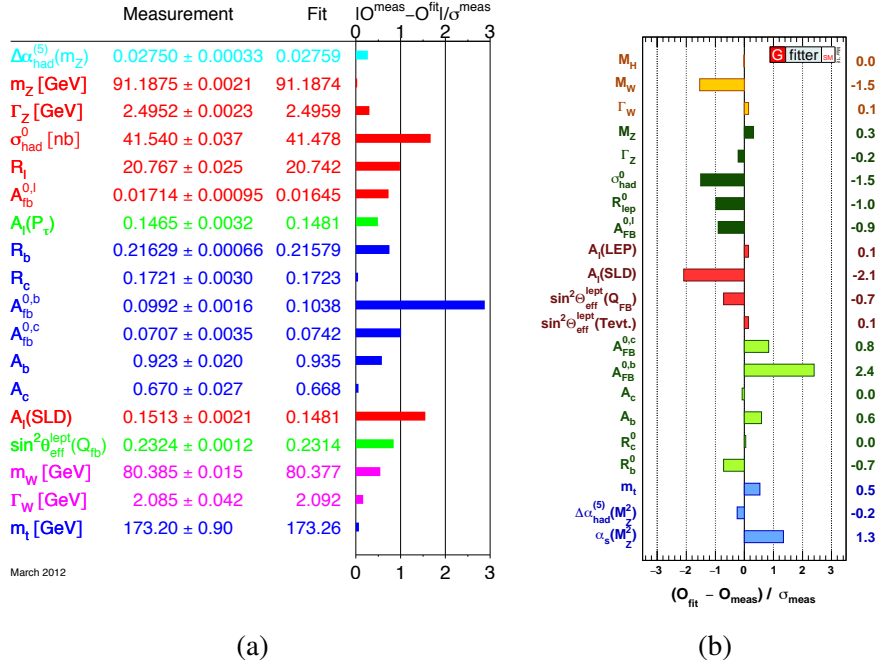
**Fig. 23:** Pulls of various (pseudo)observables due to (a) LEPEWWG [30] and (b) Gfitter [31].

– the structure of all interactions is fixed (but not all couplings are tested experimentally);
– it is an anomaly-free theory;
– it can account for rich Flavour Physics (see [26]);
– three generations allow $\mathcal{CP}$-violation (see [26]);
– it can be extended to incorporate neutrino masses and mixing (see [18]);
– it allows making systematic predictions for a wide range of phenomena at different scales;
– all predicted particles have been discovered experimentally;
– it survives stringent experimental tests.

Due to this, the SM is enormously successful (*Absolutely Amazing Theory of Almost Everything*). Since it works so well, *any* New Physics should reproduce it in the low-energy limit. Unfortunately, contrary to the Fermi-like non-renormalizable theories, the values of the SM parameters do not give us obvious *hints* for a New Physics scale. But why do we need New Physics if the model is so perfect? It turns out that we do not *understand*, why the SM works so well. For example, one needs to clarify the following:

– What explains the pattern behind Flavour Physics (hierarchy in masses and mixing, 3 generations)?
– Is there a symmetry behind the SM (electric) charge assignment?
– What is the origin of the Higgs potential?
– What is the origin of accidental Baryon and Lepton number symmetries?
– Why is there no CP-violation in the strong interactions (strong CP problem)[29]?
– Why is the Higgs-boson mass so low? (Hierarchy/Naturalness problem, see [22])
– Is it possible to unify all the interactions, including gravity?

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

---

[29]The SM Gauge group allows such a term in the SM Lagrangian, $\mathcal{L} \ni \theta_{CP} \frac{1}{16\pi^2} F^a_{\mu\nu} \tilde{F}^a_{\mu\nu}$. But it turns out that $\theta_{CP} = 0$.
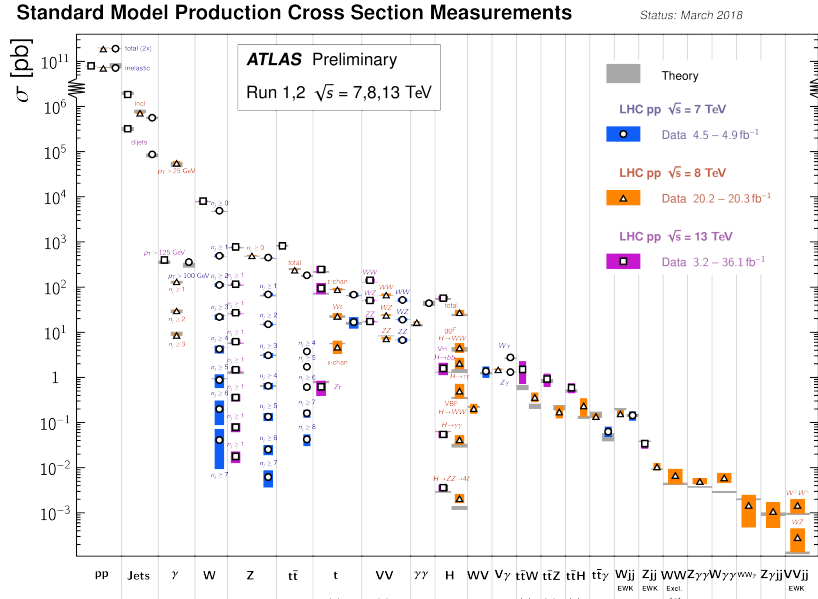
**Fig. 24:** ATLAS results of the SM cross-section measurements.
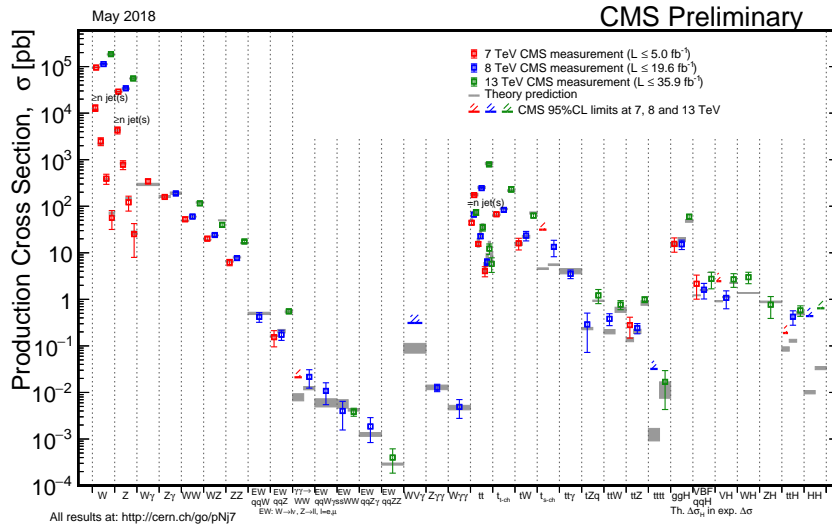


**Fig. 25:** SM processes at CMS.

– Origin of neutrino masses (see [18]);

– Baryon asymmetry (see [34]);

– Dark matter, Dark energy, Inflation (see [34]);

– Tension in $(g - 2)_\mu$, $b \to s\mu\mu$, $b \to cl\nu$;

– Possible problems with Lepton Universality of EW interactions (see [26, 35]).

In view of the above-mentioned issues we believe that the SM is not an ultimate theory (see [35]) and enormous work is ongoing to prove the existence of some New Physics. In the absence of a direct signal a key role is played by *precision* measurements, which can reveal tiny, yet significant, deviations from the SM predictions. The latter should be accurate enough (see, e.g., Ref. [36]) to compete with modern and future experimental precision [37].
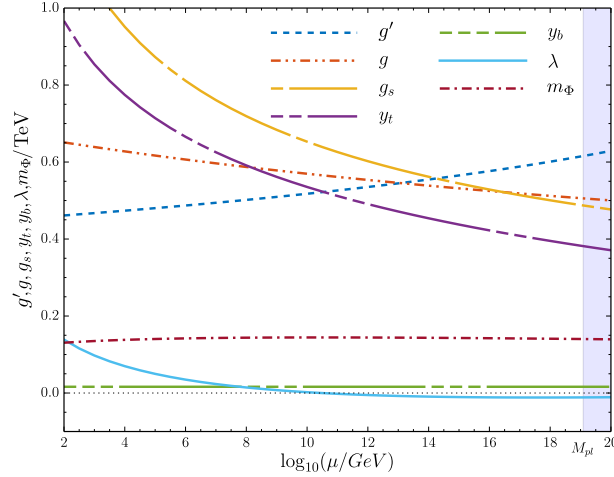
**Fig. 26:** Scale dependence of the SM parameters obtained by means of `mr` package [32].

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

## Acknowledgements

## References

[1] S. L. Glashow, Nucl. Phys. **22** (1961) 579

[2] S. Weinberg, Phys. Rev. Lett. **19** (1967) 1264

[3] A. Salam, 8th Nobel Symposium 19–25 May 1968, Lerum, Sweden, (1968) 367

[4] G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **716** (2012) 1

[5] S. Chatrchyan *et al.* [CMS Collaboration], Phys. Lett. B **716** (2012) 30

[6] R. Oerter, The theory of almost everything: The standard model, the unsung triumph of modern physics, New York, USA: Pi Press, 2006

[7] R. Kleiss, Quantum field theory for the electroweak standard model,
European School of High-Energy Physics, Trest, Czech Republic, 2007,
Eds. N. Ellis and R. Fleischer, CERN-2008-007, 1–137

[8] E. Boos, Quantum field theory and the electroweak standard model,
European School of High-Energy Physics, Paradfurdo, Hungary, 2013,
Eds. M. Mulders and G. Perez, CERN-2015-004, 1–64

[9] A. B. Arbuzov, Quantum field theory and the electroweak standard model,
European School of High-Energy Physics, Bansko, Bulgaria, 2015,
Eds. M. Mulders and G. Zanderighi, CERN-2017-004, 1–34

[10] J. Iliopoulos, Introduction to the standard model of the electro-weak interactions,
European School of High-Energy Physics, La Pommeraye, Anjou, 2012,
Eds. C. Grojean and M. Mulders, CERN-2014-008, 1–29

[11] L. B. Okun, Leptons and Quarks, World Scientific, Singapore, 2014

[12] M. E. Peskin, D. V. Schroeder, An Introduction to Quantum Field Theory, Westview Press, 1995

[13] N. N. Bogoliubov, D. V. Shirkov, Introduction to the theory of quantized fields, Interscience Publishers, 1959

[14] L. H. Ryder, Quantum Field Theory, 2nd Edition, Cambridge University Press, 1996

[15] S. Weinberg, *The Quantum Theory of Fields,* Cambridge University Press, Vol.1, 1995, Vol.2, 1996

[16] A. Zee, Quantum field theory in a nutshell, Princeton, UK: Princeton Univ. Press., 2010

[17] E. Noether, Gott. Nachr. **1918** (1918) 235, English translation: Transp.Theory Statist.Phys. 1 (1971) 186, arXiv:physics/0503066

[18] Silvia Pascoli, Neutrino Physics, these proceedings

[19] Francesco Tramontan, QCD, these proceedings

[20] H. Georgi, Ann. Rev. Nucl. Part. Sci. **43** (1993) 209

[21] M. Tanabashi *et al.* [Particle Data Group], Phys. Rev. D **98** (2018) no.3, 030001

[22] Fabio Maltoni, Higgs Physics, these proceedings

[23] F. Englert and R. Brout, Phys. Rev. Lett. **13** (1964) 321

[24] P. W. Higgs, Phys. Rev. Lett. **13** (1964) 508

[25] Scientific Background on the Nobel Prize in Physics 2013: The BEH-mechanism, interactions with short range forces and scalar particles, AAPPS Bull. **23**, no. 6 (2013), 03–19

[26] Jure Zupan, Flavour Physics and CP Violation, these proceedings

[27] D. Y. Bardin and G. Passarino, The standard model in the making: Precision study of the electroweak interactions, Oxford University Press, International series of monographs on physics, 104, Oxford, 1999

[28] A. B. Arbuzov, M. Awramik, M. Czakon, A. Freitas, M. W. Grunewald, K. Monig, S. Riemann and T. Riemann, Comput. Phys. Commun. **174** (2006) 728, arXiv:hep-ph/0507146

[29] G. Montagna, O. Nicrosini, F. Piccinini and G. Passarino, Comput. Phys. Commun. **117** (1999) 278, arXiv:hep-ph/9804211

[30] The LEP Electroweak Working Group, archived 23 September 2019

[31] M. Baak *et al.* [Gfitter Group], Eur. Phys. J. C **74** (2014) 3046

[32] B. A. Kniehl, A. F. Pikelner and O. L. Veretin, Comput. Phys. Commun. **206** (2016) 84, arXiv:1601.08143

[33] A. V. Bednyakov, B. A. Kniehl, A. F. Pikelner and O. L. Veretin, Phys. Rev. Lett. **115** (2015), 201802, arXiv:1507.08833

[34] Andrea De Simone, Cosmology and Dark Matter, these proceedings

[35] Ben Allanach, Physics Beyond the Standard Model, these proceedings

[36] A. Blondel *et al.*, Standard Model Theory for the FCC-ee Tera-Z stage, CERN-2019-003

[37] M. Dam, arXiv:1601.03849 [hep-ex]

# Introduction to perturbative QCD

*F. Tramontano*
Università di Napoli Federico II e INFN sezione di Napoli, Napoli, Italia

**Abstract**
Following a phenomenological guiding principle, in these notes I present a selection of old and recent developments in perturbative QCD and collider physics.

**Keywords**
Quantum Chromodynamics; lectures; gauge theory; hadrons; parton shower.

## 1 Basics and $e^+e^-$ annihilation into hadrons

Protons, neutrons and hadrons mainly interact via the so called strong force. The strong force has a higher level of symmetry than weak and electromagnetic interactions, since it conserves parity and isospin (almost). Unlikely the case of QED, there is no evidence of a low energy small parameter which controls the interaction strength. On the contrary, the strong interaction is characterized by a dimensional scale of about $200 \sim 300$ MeV, corresponding to the typical lifetime of its excitations that is about $10^{-24}$ s. As can be seen in Fig.(1), taken from the Review of Particle Physics [1], the typical hadron-hadron cross section is of the order of $10$ mb, that again corresponds to about $(300\,\mathrm{MeV})^2$, so it's very much harder
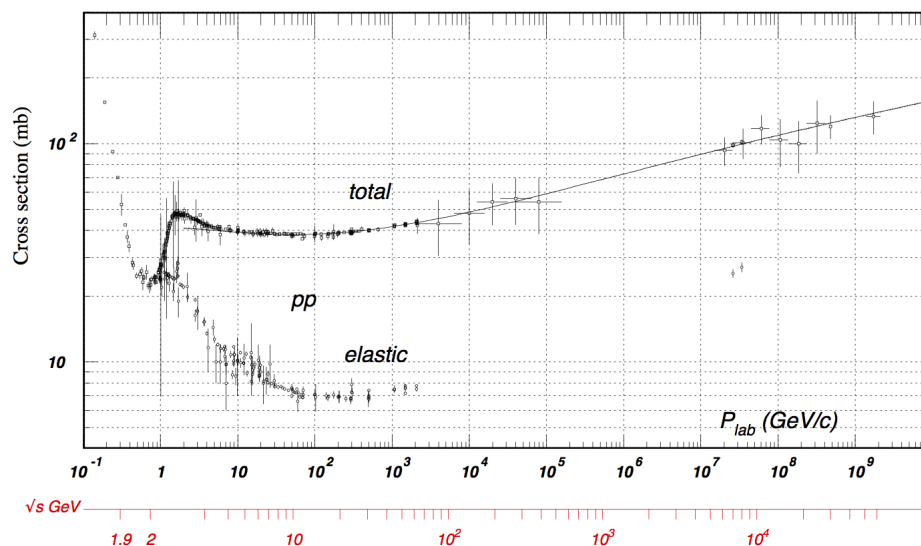


**Fig. 1:** Measurements of the total proton-proton cross sections as function of the beam energy.

then EW interactions and furthermore we don't really see the vertices.

## 1.1 Motivations for QCD

First of all, the whole hadron spectrum can be classified by assuming that the hadrons are made up of quarks carring different flavours (see Fig.(2) [1]), that each quark $q$ with flavour $f$ comes in three

| quark flavour | spin | charge (e) |
|---|---|---|
| d=down | 1/2 | -1/3 |
| u=up | 1/2 | 2/3 |
| s=strange | 1/2 | -1/3 |
| c=charm | 1/2 | 2/3 |
| b=beauty | 1/2 | -1/3 |
| t=top | 1/2 | 2/3 |

**Fig. 2:** Quark flavours and their properties.

different colours ($q_f^i$ with index $i = 1, 2, 3$) and that observable hadrons are colour singlets under the $SU(3)_C$ colour group. Singlets under the $SU(3)_C$ are easily built for a quark-antiquark (meson) or a three quark system (baryon). In Fig.(3) [1], for example, there are the lowest mass multiplets classified as representation of favour $SU(4)_F$. There is a nice symmetry in each group and mass differences are well
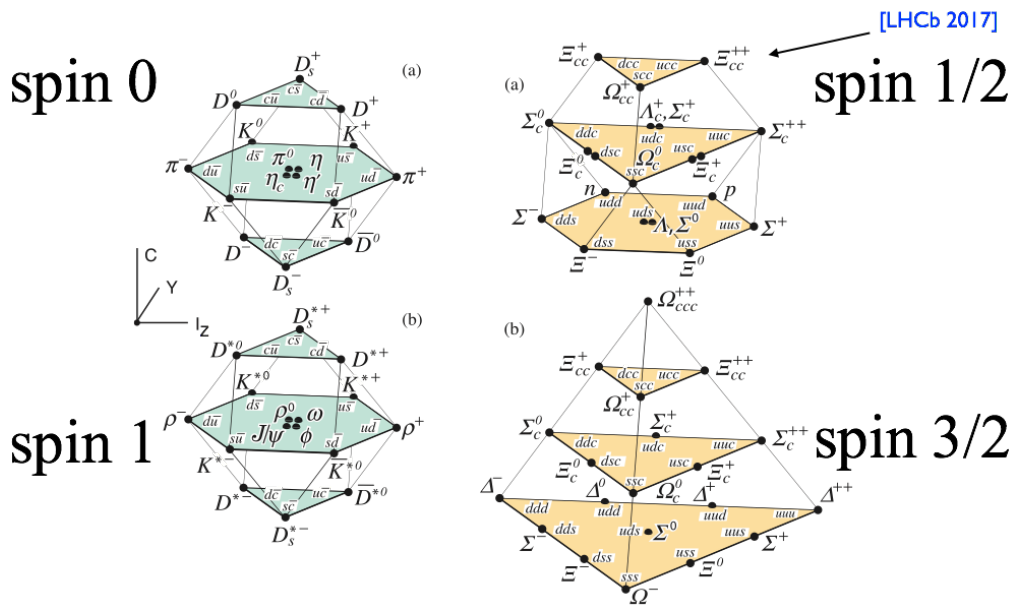


**Fig. 3:** Classification of the light mesons and baryons spectra according to the multiplets of the $SU(4)_F$ flavour group.

explained by quark masses. The relatively low number of multiplets and their composition is difficult to explain without colour constraint. For example, consider the $\Delta^{++}$ state, that is composed of three up quarks with their spin aligned. For the Pauli principle this state would not be possible without assuming total antisymmetry of its colour wave function. One could in principle assume an antisymmetric space wave function, but this would naturally require to this state to have a mass larger than the others in the same multiplet that have a different flavour content so that for them a symmetric wave function is natural; on the contrary they are all close by. It's remarkable that the $\Xi_{cc}^{++}$, the last missing state in the proton

multiplet to be discovered, has been found only in 2017 by the LHCb experiment [2]. However, the simple vision of just $q\bar{q}$ and $qqq$ states does not represent the whole story. There is now evidence of states that do not fit the spectrum dictated by these two combinations only, both for the meson and the baryon sector. The new states found involve the presence of both heavy ($c$ or $b$) or light ($u$, $d$ or $s$) quarks, so that there are light-light, heavy-light and heavy-heavy new resonances. For example, in the light-light sector, there is observation of $\pi\pi$ resonances with isospin 2 while quarks have only isospin $1/2$, and for the hidden charm and beauty sectors (resonances containing a $c\bar{c}$ or a $b\bar{b}$ pair), there is evidence of many states that for their masses, decay modes and decay widths do not fit at all the spectrum of heavy $q\bar{q}$ resonances (quarkonium). All such states have to be necessarily multi-quark (more than three) systems! However, the only necessary condition to build a colour singlet is that:

$$n_q - n_{\bar{q}} = \text{ multiple of 3} \tag{1}$$

which also implies the absence of hadrons with fractional charge although quarks have fractional charges. At the same time Eq.(1) predicts the existence of multi-quark colourless states. So far, indeed, QCD has more then one solution to the problem! See the Fig.(4). The new states could resemble nuclei, being
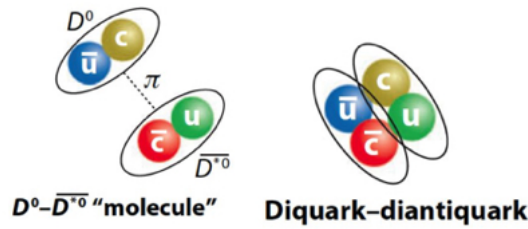


**Fig. 4:** Two QCD allowed solutions for a multi-quark state: a "molecule" made of regular hadrons binded together by the nuclear force (left) and new exotic tetra-quark (right).

molecules of color singlet states kept together by kind of nuclear forces: that would explain the fact that these states have masses typically close to the respective thresholds. Or they could also be made of colored parts confined by long-range color forces, similarly to what happens in normal mesons, making tetraquarks (or pentaquarks, hexaquarks...) as indicated by the spectrum of the $X(3872)$ (a $1^{++}$ meson decaying into $J/\psi\,\pi^+\,\pi^-$) measured by CMS, that is much harder than the (rescaled) deuteron and tritium spectrum measured in Pb-Pb collisions, see Fig.(5) [3]. We refrain here to discuss this aspects further, nevertheless we stress that such studies might indeed provide new clues to the understanding of QCD in the fully non-perturbative regime.

The second pillar in favour of QCD is the phenomenon of scaling in lepton-hadron Deep Inelastic Scattering (DIS). By this, we mean the fact that, for inclusive inelastic large angle scattering, the differential cross section, as expressed in terms of dimensionless variables, "scales" with energy:

$$\frac{d\sigma}{dxdy} \sim \frac{1}{s} \qquad (s = E_{CM}^2) \tag{2}$$

As we will discuss further, this is a prediction of the parton model of QCD. So, the strong interaction at high energy resembles a weakly interacting theory with dimensionless coupling!

Last but not the least, there has been the theoretical discovery of asymptotically free theories. Non abelian gauge theories are weakly coupled at high energies (short distances) while they can become strongly approaching a characteristic low energy scale. In the latter highly non perturbative regime, the strong interacting fields may lead to a spectrum of asymptotic binding states, "confining" the high energy degrees of freedom. So far colour has been used to justify the spectrum that is a static property of the theory, nevertheless assuming that the outcome of every strong interaction are only colourless particles
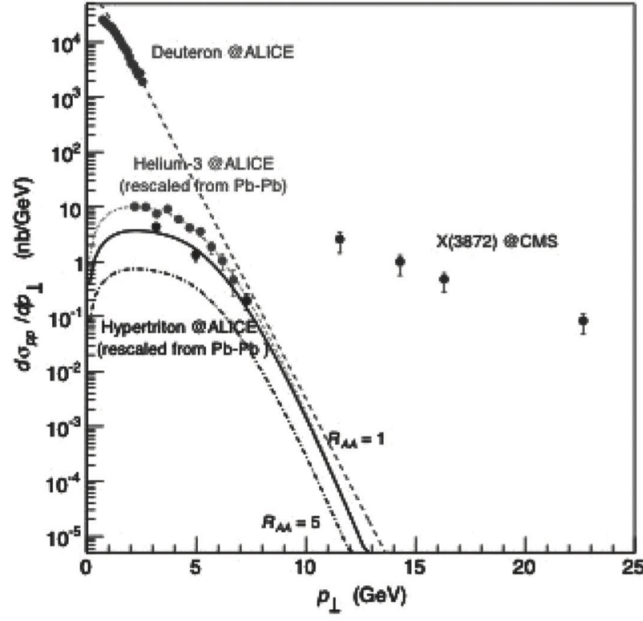
**Fig. 5:** Comparison of the measureed transverse momentum spectrum of $X(3872)$ in p-p collision (CMS) with the one of deuteron and tritium in Pb-Pb collisions (ALICE). The huge discrepancy disfavors the interpretation of the new state $X(3872)$ as a molecule.

we are implicitly also giving to colour a dynamical role. For all these reasons, non abelian gauge theories $(SU(3)_C$ in our case) are good candidates to be the theory of strong interactions. One has to assume that the strong interaction group is completely independent from the weak interaction group (commute) to guarantee that parity violating terms remain of order $\alpha_e/M_W^2$, but still one has same type of theory for weak, strong and electromagnetic interactions! The basic aspects of QCD are discussed in excellent books like for eample [4] and [5].

## 1.2 QCD Lagrangian and Feynman rules

The QCD Lagrangian is given by:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^A F_{\mu\nu A} + \sum_{f=1}^{N_f} \bar{q}_f^i(i\gamma^\mu\partial_\mu\delta_{ij} - g_s\gamma^\mu t_{ij}^A A_\mu^A - m_f\delta_{ij})q_f^j. \tag{3}$$

For $N_c = 3$ colours we have $i, j = 1, 2, 3$ and $A, B, C = 1, 2, ..., 8$. The field strength is

$$F_{\mu\nu}^A = \partial_\mu A_\nu^A - \partial_\nu A_\mu^A - g_s f_{ABC} A_\mu^B A_\mu^C. \tag{4}$$

The eight SU(3) generators fulfill the commutation rules:

$$[t^A, t^B] = if_{ABC}t^C \tag{5}$$

where $f_{ABC}$ are numbers specific of the SU(3) group named the structure constants of the group. The terms in the gluon field strength generate 3-gluon and 4-gluon couplings. Then, unlike the neutral single photon in the case of QED, the eight gluons carry colour charge. Furthermore, gluon radiation from a quark or a gluon changes its colour charge generating a colour flow. The explicit colour matrices are not

important for most practical purposes. It is customary to normalise them according to

$$Tr(t^A t^B) = T_R \delta^{AB} \qquad T_R = \frac{1}{2} \tag{6}$$

and use $t^A = \frac{1}{2}\lambda^A$, where $\lambda^A$ are the Gell-Mann matrices:

$$\lambda^1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \lambda^3 = ... \tag{7}$$

From this Lagrangian one finds the Feynman rules given in Fig.(6). Before using them for the computa-



**Fig. 6:** List of the Feynman rules corresponding to the QCD Lagrangian in Eq. 3.

tion of amplitudes in perturbation theory, one has to specify the gluon polarization tensor $d_{\mu\nu}(k)$ which appears in the Feynman rule for the gluon propagator. Several choices are possible, for example:

$$d_{\mu\nu}(k) = \sum_\lambda \varepsilon_\lambda^\mu(k) \varepsilon_\lambda^{\nu*}(k) = \begin{cases} -g^{\mu\nu} + (1-\rho)\frac{k^\mu k^\nu}{k^2 + i\epsilon} & \text{covariant} \\[2ex] -g^{\mu\nu} + \frac{n_\mu k_\nu + k_\mu n_\nu}{n\cdot k} - \frac{(n^2 - \rho k^2)k^\mu k^\nu}{(n\cdot k)^2} & \text{axial} \end{cases} \tag{8}$$

With $\rho$ and $n$ a free parameter and fixed four vector respectively, and the property that physical results have to be independent from them. The easiest and most natural choice seems to be the covariant gauge

with $\rho = 1$. However, it propagates also unphysical (non transverse) degrees of freedom, as a consequence of the non-abelian nature of the theory. This spurious degrees of freedom have to be ruled out in the calculation of physical quantities. The simplest way to implement this cancellation is via the on purpose introduction of other non physical degrees of freedom named ghosts. So, if one adopts the covariant gauge, the Feynman rules have to be extended including the complementary ones shown in Fig.(7). To meet the purpose, ghosts are coloured scalars with a fermionic property: a (-1) factor for each
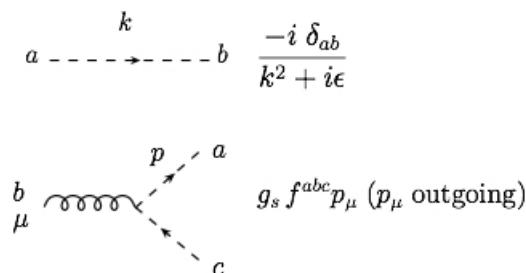


**Fig. 7:** Feynman rules involving ghost particles in covariant gauge.

ghost loop has to be included as, indeed, for the fermion loop. The axial gauges, instead, propagates only the physical degrees of freedom, and so ghosts are not required, to the price of a more complicated expression in the propagator. Let us focus on a specific example. Suppose you want to compute the scattering amplitude for the process $q\bar{q} \to gg$ that is one of the subprocesses at work to create two jets at an hadron collider. The corresponding Feynman diagrams, which involve only the physical degrees of freedom, are the ones in Fig.(8). The cross section is proportional to the squared amplitude, and one can
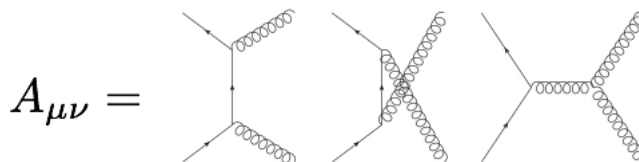


**Fig. 8:** Lowest order Feynman diagrams contribution to the scattering amplitude for $q\bar{q} \to gg$.

compute it according to the formula:

$$\sigma = A_{\mu\nu} A^*_{\mu'\nu'} d^{\mu\mu'} d^{\nu\nu'} \tag{9}$$

generically written in terms of the polarization tensors for the two gluons. If you decide to use the gluon polarization tensor as given in the axial guage there is nothing more to add. The point is that if one uses naively the simple covariant expression, i.e. $d^{\alpha\beta} = -g^{\alpha\beta}$, to compute the squared amplitude in Eq.(9), will end up with a wrong result. The correct one is recovered after adding also the ghost contribution shown in Fig.(9). What is done in several automated programs designed to calculate scattering
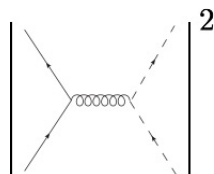


**Fig. 9:** Ghost contribution to the matrix element for the $q\bar{q} \to gg$ scattering process in a covariant gauge

amplitudes is to compute the helicity amplitudes, that is the collection of amplitudes each with a specific

choice for the helicities of the external particles. In this case one considers ab initio only physically meaningful helicity states. Then the square is obtained using the formula:

$$\sigma = \sum_{\substack{helicity \\ configurations}} |A_{h.c.}|^2 \tag{10}$$

Numerically, this procedure is very efficient. Beyond Dirac algebra we also need to compute the colour factors associated to the colour structure of the QCD vertices. The most relevant ones are related to the computations showed in Fig.(10). As we see from the second and third line in Fig.(10), the gluon
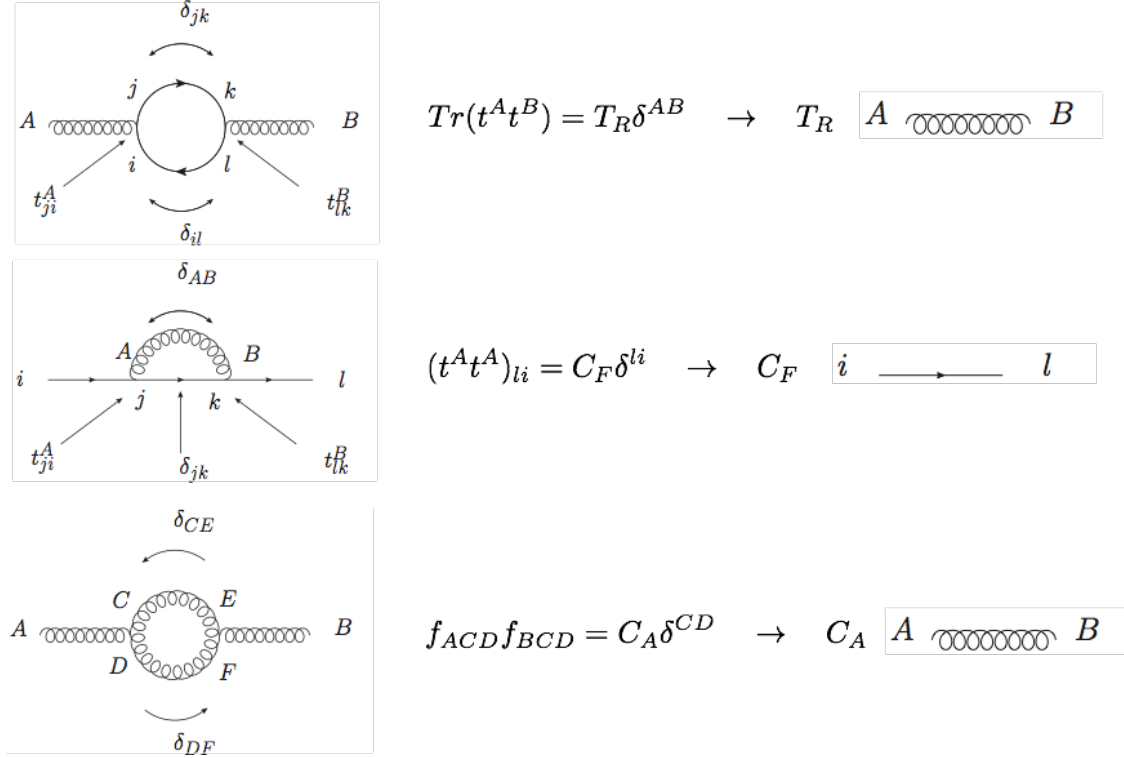


**Fig. 10:** Examples of fundamental colour structures and associated colour factors.

radiative correction to the fermion propagator receives a factor $C_F = 4/3$ while the correction to the gluon propagator receives a factor $C_A = 3$. From this difference, we argue that gluons radiate more then fermions, their coupling is effectively almost twice as much as the one of the fermions. Another colour identity that is very useful in practice is the Fierz transformation:

$$t_{ij}^A t_{kl}^A = \frac{1}{2}\delta_{kj}\delta_{il} - \frac{1}{2N_c}\delta_{ij}\delta_{kl} \tag{11}$$

which allows one to deal with colour flowing from one line to another as in Fig.(11).

In principle, starting from the Feynman diagrams, one has that a helicity amplitude can be written as a combination of the colour tensor structures ($C_i$), that are just strings of colour matrices, and fully contracted Lorentz structures (complex numbers $K_{ij}$) for each colour structure ($i$) and external helicity configurations ($j$)

$$A_j = \sum_i C_i \cdot K_{i,j}. \tag{12}$$

The same numerical construction can be generalized also for diagrams containing one closed loop. Indeed, we can factorize one loop diagrams into kinematical factors and integrals over the unconstrained
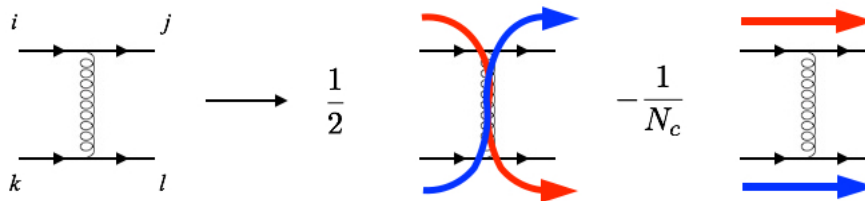
**Fig. 11:** Diagrammatic colour flow in a quark-quark scattering.

loop momentum. Note that possible one loop integrals (there is finite number of them) are all known analytically so that, beyond Dirac algebra, what is needed to build the "interference" among tree level and one loop diagrams, that is part of the Next-to-Leading Order (NLO) correction, is just the colour matrix built by interfering (contracting) all the colour structures present in both. Full automation of tree level and one loop amplitude computation has been reached in recent years. We talk of the "NLO revolution" to express the fact that, based on a large number of optimizations in the computation of the kinematical factors ($K_{i,j}$) and the availability of faster computers and larger memories with respect to the past, nowadays it is possible to generate full codes for very complicated processes with up to hundreds of thousands of Feynman diagrams just pushing a button and waiting (only once) an amount of time that is much lower then the running time needed to make simulations with that code with sufficient numerical precision. Although, as we will see in another section, the NLO revolution has a prequel.

## 1.3 QCD Phenomenology in $e^+e^-$ collisions

We now start the phenomenological exploration of QCD. It will be clearer in the following that we have to distinguish among different regimes: at low energy QCD studies address hadron spectrum as a consequence of the symmetries, the properties of finite temperature hadronic matter and lattice calculations. We will instead focus on the high energy phenomenology: in particular on perturbation theory with quarks and gluons, although only hadrons appear in the initial and final states of the scattering processes. The assumption here is that the processes of quark extraction from an initial hadron and that of hadron formation do not spoil the predictions that one can perform working with quarks and gluons. This has implications on the definition of observables that can be predicted in perturbation theory. Other fundamental applications of QCD are connected with the effective field theories. They cover several aspects and topics which play a very important role both to validate QCD and to establish QCD effects on processes that are beyond the Standard Model. This is a field that is relevant at all energies, nevertheless we will not treat it here as it would require more dedicated lectures.

Let's start from the easiest situation that is represented by the hadron production in electron-positron collisions, namely the process $e^+e^- \to \gamma^* \to$ hadrons. For energies below the Z pole, the prediction of QCD at the lowest order in the strong coupling constant, i.e. $\alpha_s^0 = (g_s^2/(4\pi))^0$, for the ratio among the cross sections for hadrons and muon pair creation is:

$$R = \frac{\sigma(\gamma^* \to hadrons)}{\sigma(\gamma^* \to \mu^+\mu^-)} = N_c \sum_i q_i^2 = 3 \cdot \left( \frac{4}{9} + \frac{1}{9} + \frac{1}{9} + \frac{4}{9} + \frac{1}{9} + ... \right) \quad (13)$$

that is obtained squaring the diagram in Fig.(12) and the rational numbers in parenthesis are the squared charges of $u$, $d$, $s$, $c$, $b$... quarks ($q_i$). At the next order in QCD, i.e. at order $\alpha_s$, one has to compute all the diagrams shown in Fig.(13), corresponding to the emission of a virutal or real gluon off the external quarks. They are indeed the same as if the corrections would have been induced by a QED photon, but for a colour factor of $C_F = 4/3$ (in QED the correction is $3\alpha_e/4$). At order $\alpha_s^2$, instead, we encounter a
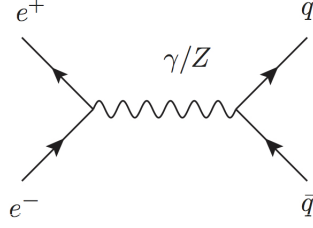
**Fig. 12:** Lowest order diagram to the process $e^+e^- \rightarrow$ hadrons. An electron-positron pair annihilates into a photon or a Z boson which converts into quark anti-quark pair.



**Fig. 13:** At Next-to-Leading Order in the QCD interaction, the scattering amplitude receives contributions from both virtual and real emission of an extra gluon off the external quark lines. After squaring the amplitude and integrating it, the correction amount to the leading order cross section $\sigma_0$ times $\frac{\alpha_s}{\pi}$.

problem. Once we add all the diagrams we get:

$$\sigma = \sigma_0 \left\{ 1 + \frac{\alpha_s}{\pi} + \left[ c + \pi b_0 \log \frac{M^2}{Q^2} \right] \frac{\alpha_s^2}{\pi^2} \right\} \tag{14}$$

with

$$c = 1.986 - 0.115\, n_f \qquad b_0 = \frac{33 - 2\, n_f}{12\pi} \tag{15}$$

In this formula for the total cross section we have indeed a logarithmic ultraviolet divergence that we have regulated by introducing the ultraviolet cutoff energy scale $M$, while $Q$ is the invariant mass of the system, i.e. the "hard" scale of the process, and $n_f$ is the total number of quark flavours that can enter in the loops. More in general, one can show that for any physical quantity $G$ computed as a power series in the coupling constant:

$$G = G_0 \alpha_s^n + (...)\alpha_s^{n+1} + ... \tag{16}$$

the expansion has the form:

$$G = G_0 \alpha_s^n + \left( G_1 + n\, G_0\, b_0 \log \frac{M^2}{Q^2} \right) \alpha_s^{n+1} + ... \tag{17}$$

Indeed, for the $e^+e^-$ cross section into hadrons up to second order Eq.(14) gives:

$$\frac{\sigma}{\sigma_0} - 1 = \frac{1}{\pi}\alpha_s + \left( \frac{c}{\pi^2} + \frac{1}{\pi} b_0 \log\left( \frac{M^2}{Q^2} \right) \right) \alpha_s^2 \tag{18}$$

where we can easily make the identifications: $n = 1$, $G_0 = 1/\pi$ and $G_1 = c/\pi^2$. Now we define:

$$\widetilde{\alpha}_s(\mu) = \alpha_s + b_0 \log\left(\frac{M^2}{\mu^2}\right) \alpha_s^2 \tag{19}$$

introducing a new fictitious scale $\mu$. So that in terms of $\widetilde{\alpha}_s(\mu)$, neglecting terms of order $\alpha_S^{n+2}$ and higher, we find that any physical quantity gets the form:

$$G = G_0 \widetilde{\alpha}_s^n(\mu) + \left(G_1 + n b_0 \log\left(\frac{\mu^2}{Q^2}\right)\right) \widetilde{\alpha}_s^{n+1}(\mu) + ... \tag{20}$$

where all the dependence on the cutoff regulator $M$ has been adsorbed, and it is now contained, in the redefinition of the coupling constant. This is what is called renormalization.

More formally, we have that for any physical quantity $G(\alpha_s, M, ...)$ depending on the strong coupling, an ultraviolet cutoff, and physical variables like masses and momenta, we can always define a charge containing the whole dependence on the UV cutoff $M$:

$$\widetilde{\alpha}_s(\mu, M, \alpha_s) = \alpha_s + c_1(\mu, M)\alpha_s^2 + ... \tag{21}$$

in such a way that the physical quantity has a finite expression in terms of $\widetilde{\alpha}$, $\mu$ and the physical variables:

$$G(\alpha_s, M, ...) = G'(\widetilde{\alpha}_s(\mu, M, \alpha_s), \mu, ...). \tag{22}$$

Even if the definition of the coupling contains divergences, they are of ultraviolet origin and we might think that other dynamical effects may solve the problem at high energies while for the moment, at the energies we are probing the theory, we can assume the coupling to be finite, measure it in one process and use it to make predictions for another process! Before doing this, let us look more carefully to Eq.(22). Taking derivative on both sides wrt $\log(\mu^2)$ we have (note that $G$ in the left hand side of Eq.(22) has no $\mu$ dependence):

$$0 = \frac{\partial G'(\widetilde{\alpha}_s, \mu, ...)}{\partial \widetilde{\alpha}_s} \frac{\partial \widetilde{\alpha}_s}{\partial \log(\mu^2)} + \frac{\partial G'(\widetilde{\alpha}_s, \mu, ...)}{\partial \log(\mu^2)} \tag{23}$$

or, equivalently:

$$\frac{\partial \widetilde{\alpha}_s}{\partial \log(\mu^2)} = -\frac{\frac{\partial G'(\widetilde{\alpha}_s, \mu, ...)}{\partial \log(\mu^2)}}{\frac{\partial G'(\widetilde{\alpha}_s, \mu, ...)}{\partial \widetilde{\alpha}_s}} = \beta(\widetilde{\alpha}_s, \mu) \tag{24}$$

But $\beta$ is dimensionless, so if it does not depend explicitly on $M$ it cannot depend explicitly on $\mu$, so that:

$$\beta \equiv \beta(\widetilde{\alpha}_s) = \frac{\partial \widetilde{\alpha}_s}{\partial \log(\mu^2)} \tag{25}$$

Then, from $e^+ e^-$ cross section into hadrons we have:

$$\widetilde{\alpha}_s(\mu) = \alpha_s + b_0 \log\left(\frac{M^2}{\mu^2}\right) \alpha_s^2 \quad \Rightarrow \quad \frac{\partial \widetilde{\alpha}_s}{\partial \log \mu^2} = -b_0 \widetilde{\alpha}_s^2 + \mathcal{O}\left(\widetilde{\alpha}_s^3\right) \tag{26}$$

In summary, any physical quantity can be given as an expansion in $\widetilde{\alpha}_s$:

$$G'(\widetilde{\alpha}_s, \mu, ...) = \sum_i G_i(\mu, ...)\widetilde{\alpha}_s^i \tag{27}$$

and if we change $\mu$ and $\widetilde{\alpha}_s$ in such a way that:

$$\delta \widetilde{\alpha}_s = \beta(\widetilde{\alpha}_s)\delta \log(\mu^2) \tag{28}$$

$$\text{QCD: } b_0 = \frac{33 - 2n_f}{12\pi} > 0$$

$$\text{QED: } b_0 = -\frac{4n_f}{12\pi} < 0$$

**Fig. 14:** One loop contributions to the beta function in QCD and QED for the running of the coupling constant. While the fermionic loops gives raise to the same contribution in both cases (apart for the constant factor $C_F = 4/3$), the gluon bubble insertion in the QCD case turns the sign of the $b_0$ coefficient from negative to positive.

the prediction of physical quantities does not change. The last property goes under the name of Renormalization Group Invariance. After renormalization, the $e^+e^-$ cross section into hadrons has become (dropping the tilde from now on):

$$\sigma = \sigma_0 \left\{ 1 + \frac{\alpha_s(\mu)}{\pi} + \left[ c + \pi b_0 \log\left(\frac{\mu^2}{Q^2}\right) \right] \frac{\alpha_s^2(\mu)}{\pi^2} \right\} \tag{29}$$

It will be better to choose $\mu \sim Q$ otherwise the second order may become larger then the first (and subsequent orders even larger!). Remarkably, choosing $\mu = Q$ (or taking the two in a fixed ratio), the energy dependence of the cross section is entirely described by the scale dependence (running) of $\alpha_s$ and so it is of fundamental importance to solve the equation on the right hand side of Eq.(26). This equation can be recast in the following form:

$$\frac{\partial}{\partial \log \mu^2} \frac{1}{\alpha_s} = b_0 \tag{30}$$

with the obvious solution:

$$\frac{1}{\alpha_s} = b_0 \log \mu^2 + constant \tag{31}$$

and the integration constant can be used to get the expression:

$$\alpha_s(\mu) = \frac{1}{b_0 \log \frac{\mu^2}{\Lambda^2}} \tag{32}$$

The parameter $\Lambda$, or $\Lambda_{QCD}$, has nothing to do with a scale value where the coupling diverges. It is just a smart way to rewrite the integration constant. Furthermore, for scales where the coupling becomes large, one must include higher order corrections to the $\beta$ function in Eq.(26) leading to a different structure of the solution. The logarithmic ultraviolet divergences come mainly from the bubble graphs in Fig.(14) and the main difference with the QED case is due to the gluon bubble that brings an opposite sign with respect to the contribution of the fermionic bubbles. In QCD, $\Lambda$ has to be in the range of the typical hadronic scale we have mentioned above, the scale at which the hadronic systems become strongly coupled. Assuming $\Lambda$ in the range $100 \div 500\,\text{MeV}$, Eq.(32) gives $\alpha_s(M_Z) = 0.1 \div 0.13\,(\pm 13\%)$ and $\alpha_s(10^7\text{GeV}) = 0.040 \div 0.044\,(\pm 5\%)$, i.e. going up with energy, the logarithmic dependence nicely narrows the relative uncertainty. The current status of the computation of $e^+e^-$ annihilation into hadron is given by the formula:

$$\frac{\sigma}{\sigma_0} = 1 + \frac{\alpha_s}{\pi}\left(1 + 0.448\alpha_s - 1.30\alpha_s^2 - 2.59\alpha_s^3\right) + ... \tag{33}$$

where the last term has been computed in [6]. The numerical coefficients are obtained for the case $n_f = 5$ (the general expression is very complicated) and $\alpha_s \equiv \alpha_s^{\overline{MS}}(Q)$, with $\overline{MS}$ reflecting the details of the renormalization prescription used to define the coupling constant. Corrections are well behaved and amount to about $5\%$ at the second order, $2\%$ at the third order and $4\%_{oo}$ at the fourth order. In principle

one could think to use this formula to determine $\alpha_s$ with an accuracy better then 1%, although the $e^+e^-$ cross section is not an observable very sensitive to QCD effects since the perturbative expansion only starts at order $\alpha_s^0$. As for the $\beta$ function, the situation is as follows:

$$\frac{d\alpha_s}{d\log\mu^2} = -b_0\alpha_s^2 - b_1\alpha_s^3 - b_2\alpha_s^4 - b_3\alpha_s^5 - b_4\alpha_s^6 \tag{34}$$

where the last coefficient has been computed by three groups [7–9]. The first coefficients are:

$$b_0 = \frac{33 - 2n_f}{12\pi}, \quad b_1 = \frac{153 - 19n_f}{24\pi^2}, \quad b_2 = \frac{2857 - \frac{5033n_f}{18} + \frac{325n_f^2}{54}}{(4\pi)^4} \tag{35}$$

In an actual computation at a given order in perturbation theory, the usual prescription is to use the running of the coupling constant at one higher order. Therefore, since there are not so many very higher order computations, what is used in general is the second order running:

$$\alpha_s(\mu) = \frac{1}{b_0 \log \frac{\mu^2}{\Lambda^2}} \left( 1 - \frac{b_1 \log \log \frac{\mu^2}{\Lambda^2}}{b_0^2 \log \frac{\mu^2}{\Lambda^2}} \right) \tag{36}$$

Measurements of $\alpha_s$ have been done combining large data sets as the one reported in Fig.(15). The curvi-
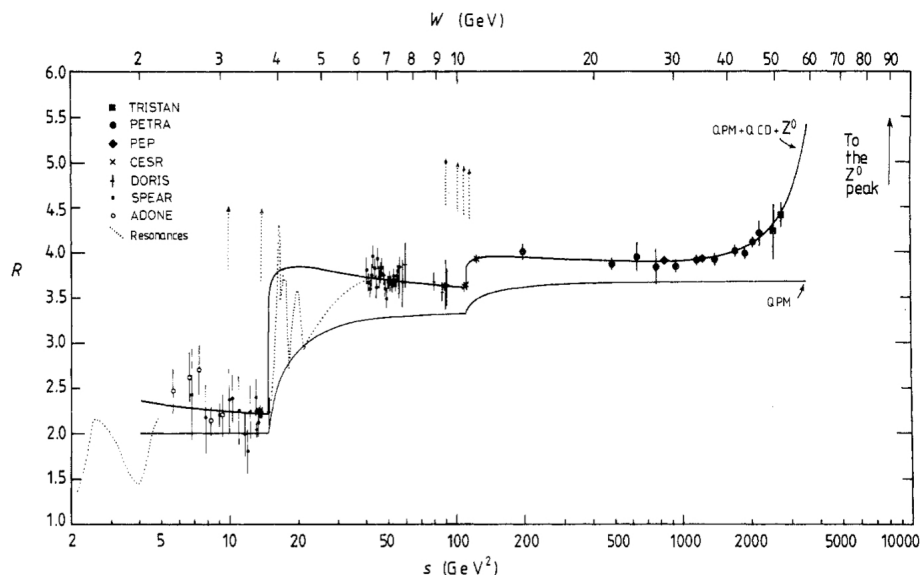


**Fig. 15:** A compilation of $R$ ratio measurements at different energies used for the extraction of $\alpha_s$.

linear behaviour in the prediction is given by a threshold velocity term to take into account that quarks are not massless. NLO corrections are clearly demanded for the fit. Combining results of experiments from 20 up to 65GeV, the value $\alpha_s(35\text{GeV}) = 0.146 \pm 0.030$ is found, that after evolution implies: $\alpha_s(M_Z) = 0.124 \pm 0.021$. Other ways to determine $\alpha_s$ is through the measurement of the total hadronic decay width at the Z pole at LEP, which gives $\alpha_s(M_Z) = 0.122 \pm 0.009$, or comparing hadronic and leptonic tau decays. In this case one has that QCD corrections displaces the value of the ratio from 3, that is a factor entirely due to colour. It is found $\alpha_s(M_\tau) = 0.36 \pm 0.05$ that after evolution implies: $\alpha_s(M_Z) = 0.122 \pm 0.005$.

## 1.4 How to compute more features of the final state

The nice agreement of the coupling constants extracted from total cross sections and decay rates mentioned at the end of the previous section is certainly an indication of the goodness of the theory, but how

to go further and prove the validity of the theory also at differential level? Indeed, to compute $e^+e^-$ annihilation into hadrons at the first order in the strong coupling, we have combined diagrams for producing two quarks and two quarks and a gluon, all particles that btw have never been observed. Furthermore, the final state partons bring colour quantum number that has never been observed neither. It is clear that we have to find the conditions that allow for the description of the final state in terms of an evolution from the elementary processes (among quark and gluons) to the hadrons that are observed. In doing that we will also gain the possibility to make differential predictions with QCD. We will again use the radiative corrections for $e^+e^-$ to hadrons to give an idea of such a procedure. As we have seen in Fig.(13), the first perturbative correction to $e^+e^-$ annihilation into hadrons (NLO QCD) is computed adding real and virtual emission diagrams. The technical difficulty is that the contributions live in different phase spaces. We start with the computation of the diagrams for the real emission in Fig.(16).
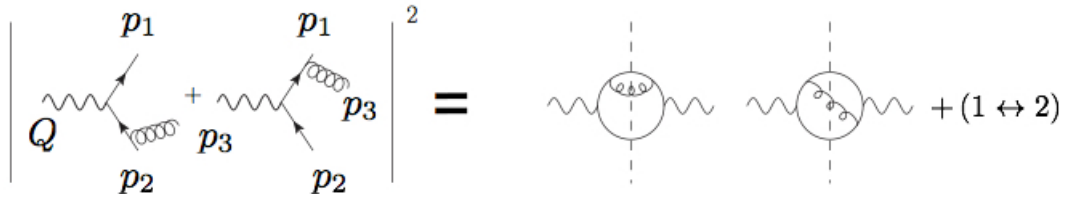


**Fig. 16:** Diagrammatic representation of the square amplitude for the real gluon emission process $e^+e^- \to q\bar{q}g$.

To present the result of the computation we introduce the energy fractions:

$$x_i = \frac{2\, p_i \cdot Q}{Q^2} = \frac{2\, E_i}{Q} \text{ (c.m. frame)} \quad \to \quad x_i > 0. \tag{37}$$

Energy conservation implies:

$$x_1 + x_2 + x_3 = \frac{2 \sum_i p_i \cdot Q}{Q^2} = 2. \tag{38}$$

Furthermore, neglecting the mass of the quarks, we have that:

$$2\, p_1 \cdot p_3 = (p_1 + p_3)^2 = (Q - p_2)^2 = Q^2 - 2\, p_2 \cdot Q \tag{39}$$

that can also be written as:

$$2\, E_1 E_3 (1 - \cos\theta_{13}) = Q^2(1 - x_2). \tag{40}$$

This tells that $x_i < 1$ and that $x_2 \to 1$ in the limit $\theta_{13} \to 0$ with $\theta_{13}$ the angle between the three momenta of the outgoing quark with momentum $p_1$ and the gluon with momentum $p_3$. Including the phase space factors, in terms of the energy fractions, the real emission contribution to the cross section is given by:

$$\sigma^R = \int_0^1 dx_1 dx_2 dx_3 \delta(2 - x_1 - x_2 - x_3)|M_R(x_1, x_2, x_3)|^2 \tag{41}$$

and from the diagrams in Fig.(16) we have:

$$|M_R(x_1, x_2, x_3)|^2 = \sigma_0 C_F \frac{\alpha_S}{2\,\pi} \frac{x_1^2 + x_2^2}{(1 - x_1)(1 - x_2)} \tag{42}$$

So we see that, from the mathematical point of view, there are non-integrable divergences. This of course implies that the diagrams with the virtual emission will diverge in exactly the same way so to cancel the divergence of the real in the inclusive result. The collection of the divergences in Eqs.(41,42) is shown pictorially in Fig.(17).
Actually, from the running of the strong coupling and the experimental evidence, we already know that

$$x_1 \to 1 \iff \theta_{23} \to 0 \quad \textbf{collinear} \; \textbf{(a)}$$
$$x_2 \to 1 \iff \theta_{13} \to 0 \quad \textbf{collinear} \; \textbf{(b)}$$
$$\frac{1-x_1}{1-x_2} \sim const. \; x_3 \to 0 \iff E_3 \to 0 \quad \textbf{soft} \qquad \textbf{(c)}$$
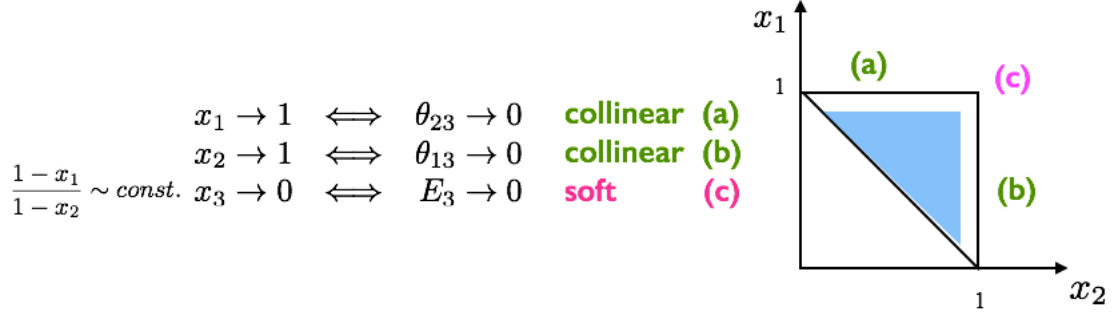
**Fig. 17:** Singular limits of the real matrix element squared and their representation in the 3-body Dalitz plot $x_1 x_2$. For massless partons, the available phase space is given by the upper right triangle. The singular limits corresponds to: the up ($x_1 = 1$) and right ($x_2 = 1$) edges, respectively corresponding to the configuration with the gluon becoming collinear to the quark and the anti-quark, and to the vertex (1,1) corresponding to a zero energy gluon (soft gluon limit).

there is a physical cut-off for the energy fraction, let's call it $\varepsilon$, given by the ratio of the typical hadron mass $M_{had}$ and the hard scale of the process $Q$. Physically, here is indeed a change of regime for:

$$\varepsilon \sim \frac{M_{had}}{Q} \sim \frac{\Lambda_{QCD}}{Q} \tag{43}$$

Note that, however, even if the infinite might not be there, expressing the logarithmic dependence in term of the strong coupling, mathematically one has that:

$$\infty = \alpha_s(Q) \int_0^1 \frac{dx}{1-x} \quad \Rightarrow \quad \alpha_s(Q) \int_0^{1-\varepsilon} \frac{dx}{1-x} \sim \alpha_s(Q) \log \frac{1}{\varepsilon} \sim \alpha_s(Q) \frac{1}{\alpha_s(Q)}$$

from which one has (including more and more orders does not improve the situation):

$$\sigma \sim \sigma_0 \left( 1 + \alpha_s(Q) \frac{1}{\alpha_s(Q)} + ... \right) \sim \sigma_0 (1 + 1 + 1 + ...). \tag{44}$$

Ultimately, then, the divergences that show up in the computation are intimately connected with the non perturbative regime of the theory and they are a characteristic feature of long distance (small $\epsilon$ limit) phenomena. It is useful to rewrite the real matrix element in Eq.(42) as:

$$\frac{x_1^2 + x_2^2}{(1-x_1)(1-x_2)} = \frac{1 + (1-x_3)^2}{(1-x_1)(1-x_2)} - 2 \tag{45}$$

and then use the relation:

$$\frac{1}{(1-x_1)(1-x_2)} = \frac{1}{x_3} \left( \frac{1}{1-x_1} + \frac{1}{1-x_2} \right) \tag{46}$$

so that the real matrix element, and in particular its most divergent part, gets the form:

$$d\sigma_R = \sigma_0 dW_{13} + \sigma_0 dW_{23} + \; finite \; terms \tag{47}$$

(and similarly for $dW_{13}$) with:

$$dW_{23} = \frac{dx_1}{1-x_1} dx_3 P_{qg}(x_3) = \frac{d\cos\theta_{23}}{1-\cos\theta_{23}} dx_3 P_{qg}(x_3) \tag{48}$$

and

$$P_{qg}(x_3) = C_F \frac{\alpha_s}{2\pi} \frac{1 + (1-x_3)^2}{x_3}. \tag{49}$$

The function $P_{qg}(x)$ is called splitting function and it is easy to show that

$$
dW_{23} \quad \xrightarrow{\substack{\theta_{23} \to 0 \\ E_3 \to 0}} \quad \frac{d\theta_{23}^2}{\theta_{23}^2} \frac{dE_3}{E_3} \cdot \tag{50}
$$

In Eq.(48) we exploited the relation:

$$
x_2 x_3 (1 - \cos\theta_{23}) = 2(1 - x_1) \implies x_1 = 1 - \frac{x_2 x_3}{2}(1 - \cos\theta_{23}). \tag{51}
$$

Then, after performing the change of variable from $x_1$ to $\cos\theta_{23}$, the integration over the $\delta$-function in Eq.(41) yields:

$$
\begin{aligned}
\int \frac{dx_1}{1 - x_1} dx_2 dx_3 \delta(2 - x_1 - x_2 - x_3) \\
= \int \frac{d\cos\theta_{23}}{1 - \cos\theta_{23}} dx_2 dx_3 \delta\left[1 - \left(1 - \frac{x_3}{2}(1 - \cos\theta_{23})\right) x_2 - x_3\right] \\
= \int_{-1}^{1} \frac{\cos\theta_{23}}{1 - \cos\theta_{23}} \int_0^1 dx_3 \frac{1}{1 - \frac{x_3}{2}(1 - \cos\theta_{23})}
\end{aligned}
$$

Now, it turns out that it is possible to combine the real and the virtual contribution together paying attention to the fact that the two contributions leave on different phase spaces. The result is as follows:

$$
\begin{aligned}
\sigma_R + \sigma_V \;=\; & \sigma_0 \int_{-1}^{+1} \frac{d\cos\theta_{23}}{1 - \cos\theta_{23}} \int_0^1 dx_3 P_{qg}(x_3) \left[\frac{1}{1 - \frac{x_3(1 - \cos\theta_{23})}{2}} - 1\right] \\
& + (2 \to 1) + finite\ terms
\end{aligned} \tag{52}
$$

The first term in the squared bracket comes entirely from the real contribution, while the $-1$ is the contribution of the divergent part of the virtual term. Actually, knowing that the total is finite we could have predicted it. In the small $x_3$ (soft) and small $\theta_{23}$ (collinear) limits the whole squared parenthesis goes like $x_3 \theta_{23}^2$, so regularizing the divergences of the other factors (see Eq. (50)). The finiteness of the inclusive result that we have seen in the previous section is however a property of the theory related directly to the unitarity that force the conservation of probabilities. In summary, the total of the virtual and the real contribution has to be finite (and their inclusive sum is just $\sigma_0 \alpha_s / \pi$), but now we know something more: first, the virtual is divergent in every point of its (Born) phase space, second, the limiting configurations in which the real becomes divergent (extra radiated parton soft or collinear to another parton) are the ones in which the real and the virtual become indistinguishable, third, the two divergences cancels locally, ie for any and irrespective of the Born level partonic configuration. With these properties we can think to implement the local cancellation of the divergences and make predictions at differential level including the first order radiative corrections. This will have an impact on the kind of variables we can define in perturbation theory. Let's state it more carefully. In every computation that is not just the total cross section, the differential cross section is convoluted with a phase space function ($F$) that defines the physical quantity we want to compute (including experimental cuts):

$$
\sigma_B = \int_m d\sigma_B \qquad d\sigma_B = d\phi^{(m)} |M_m^{tree}|^2 F^{(m)}(p_i) \tag{53}
$$

going at Next-to-Leading Order (NLO) we have:

$$
\begin{aligned}
d\sigma_V &= d\phi^{(m)} |M_m^{loop}|^2 F^{(m)}(p_i) \tag{54} \\
d\sigma_R &= d\phi^{(m+1)} |M_{m+1}^{tree}|^2 F^{(m+1)}(p_i) \tag{55}
\end{aligned}
$$

so that:

$$\sigma_{NLO} - \sigma_B = \int_m d\sigma_V + \int_{m+1} d\sigma_R \qquad (56)$$

In order for the InfraRed (IR) and Collinear (C) cancellation to take place also at differential level it is necessary and sufficient that the value of the observable we consider, $F$ in the formulas above, is insensitive to a soft emission or to the situation in which the emitted parton is collinear to another parton. In formulas, one must have that for:

$$p_i \to 0 \qquad \text{or} \qquad p_i \parallel p_j \qquad (57)$$

the function $F$ is such that:

$$F^{m+1}(..., p_i, ..., p_j, ...) \simeq F^m(..., p_i + p_j, ...) \qquad (58)$$

This condition guarantees the local cancellation of the singularities. As expected of course, the computed quantity cannot resolve long distance phenomena, but having cancelled the logarithmic enhancements from real and virtual, non perturbative physics will be power suppressed! The power of the suppression will depend on the physical observable. The implementation of the formula in Eq.(56) is however a highly non trivial task, but thanks to another fundamental property of QCD amplitudes this task has been strongly simplified. In the divergent limits in Eq.(57), the amplitudes factorize, in the sense that they can be approximated by the product of the (process dependent) amplitude with a final state parton less times a universal factor that represents the emission of the extra parton. Apart from colour and spin correlation factors (computable algorithmically), the amplitude with a parton less is just the leading order amplitude. The universality of the radiation factors can be exploited to write an auxiliary cross section to be subtracted from the real matrix element and so cancelling its divergences. One builds such a cross section combining the Born matrix element and the radiation factors, which bring the dependence on the radiation variables. To not alter the physical cross section one has to add back this subtraction and one can do so after having integrated the auxiliary cross section over the radiation degrees of freedom, obtaining a function that depends only on the Born phase space variables, as for the virtual contribution. So that, combining the two, one can also check the explicit cancellation of the divergences in the virtual matrix element. In formula, what is done in practice is to build two separately finite and integrable integrands as follows [10]:

$$\sigma_{NLO} - \sigma_B = \int_m \left[ d\sigma_V + \int_1 d\sigma_{Aux} \right] F^{(m)}(p_i) + \int_{m+1} \left[ d\sigma_R F^{(m+1)}(p_i) - d\sigma_{Aux} F^{(m)}(p_i) \right] \quad (59)$$

In the formula above we have made explicit the function that represents the measurement constraint, $F$. Note that the auxiliary cross section is in general a function of the whole real phase space variable. Nevertheless, it multiplies the observable function for the leading order kinematics $F^{(m)}$. Indeed, this must be the case in order not to alter the distributions at the differential level. In the language of Monte Carlo integration, we call a phase space point an event and we assign a weight to it accordingly to the integrand function. In particular, for the real events, the weight must be given by the real cross section $d\sigma_R F^{(m+1)}(p_i)$ only. The idea underlying Eq.(59) is that in correspondence of each real event we build a "counter-event" in the leading order phase space whose weight is given by the auxiliary cross section $d\sigma_A F^{(m)}(p_i)$. In the singular limits, event and counter-event coincide leading to the local cancellation of divergences and making the real integration finite. On the other hand, the extra weights in the leading order phase space due to the counter-events are exactly balanced by the integrated auxiliary cross section over the radiation degrees of freedom. This means that the procedure is effectively based on a mapping of the real phase space variables into the ones of the Born cross section plus (three) radiation variables, which allows one to pick a counter-event for each real event. Moreover, these radiation variables are those we integrate over to build the second piece in the first integrand in Eq.(59). Thanks to the universality of the factorization property and the possibility to write general mappings, the main ingredients of this

construction can be built once and for all and the auxiliary cross sections can be built algorithmically. Various subtraction formalisms have been proposed [11, 12] and they can be considered the prequel of the so called NLO revolution.

The emerging underlying theoretical picture of an high energy event is then as follows. Perturbation theory can be used to describe the hard process, ie the scattering process with the highest momentum transfer of the whole event. We learned from the running of the strong coupling that at very high energy scales it gets smaller and smaller, so that one can truncate the perturbative expansion at a given finite order to get a fully differential prediction for any InfraRed and Collinear safe observable. Then, perturbation theory can still be used to approximate the large number of subsequent emissions (almost soft and collinear) at lower and lower energies. Finally, when the original energy has been degraded and the emissions happen at energies close to the scale of the breakdown of the perturbation theory a new non-perturbative regime starts in which the interaction is really strong and binds quarks and gluons into hadrons. This last hadronization process, in which a multitude or shower of partons is packed into well separated and colourless hadrons, cannot certainly be described by perturbation theory and has to be modelled. This means that the hadronization process can be based on theoretically well motivated assumptions, but it also contains many parameters and ultimately no justification for their values. Nevertheless, in QCD the scaling of hadron multiplicity with the total energy of the $e^+e^-$ annihilation can be computed with the only assumptions that hadron formation happens at low scale, of the order of 1GeV, and that perturbation theory is valid up to slightly above that scale. Relating hadron multiplicity to the average number of partons at the end of the parton shower cascade one can predict:

$$\langle n \rangle = \frac{\sqrt{96\pi}}{\beta\sqrt{\alpha_s(Q^2)}} + \left( \frac{1}{4} + 10\frac{n_f}{27\beta} \right) \log \alpha_s(Q^2) + \mathcal{O}(1). \tag{60}$$

Such a dependence on the collision energy of the number of charged particles produced, is in good agreement with data, as shown in Fig.(18).
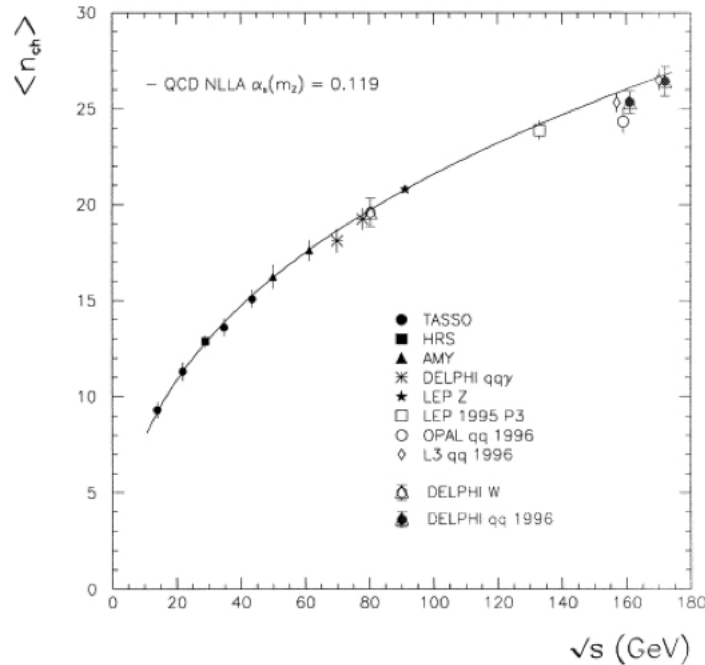


**Fig. 18:** Data-theory comparison for the average hadron multiplicity in a $e^+e^-$ collision as function of the collider energy.

## 1.5 QCD phenomenology at differential level

### 1.5.1 Shape variables

The first example of an InfraRed and Collinear safe variable for $e^+e^-$ annihilation into hadrons is the Thrust distribution. It is defined by the formula:

$$T = \max_n \frac{\sum_i |p_i \cdot n|}{\sum_i |p_i|} \tag{61}$$

where the $p_i$ are the three momenta of the outgoing hadrons. Event by event, the value of the variable $T$ is computed searching for the direction that maximize the ratio in Eq.(61). It is easy to convince yourself that a two jet like event, i.e. an event with just two hadrons or just two well collimated back to back sprays of hadrons, has $T = 1$, while an isotropic event ends up with $T = 1/2$. In turn, the value of this variable reflects the shape of the distribution of the hadrons in the final state. From this the name of shape variable for the Thrust. It is also easy to check that a soft or a collinear emission cannot alter significantly the Thrust value. Going at very high energy the strong coupling becomes smaller and smaller and so does the radiative correction to the average Thrust. The average Thrust is then predicted to tend to 1, approaching the leading order back-to-back configuration. This behaviour can be seen in Fig.(19) also in comparison with the experiments. In the fit reported in Figure(19) also power corrections
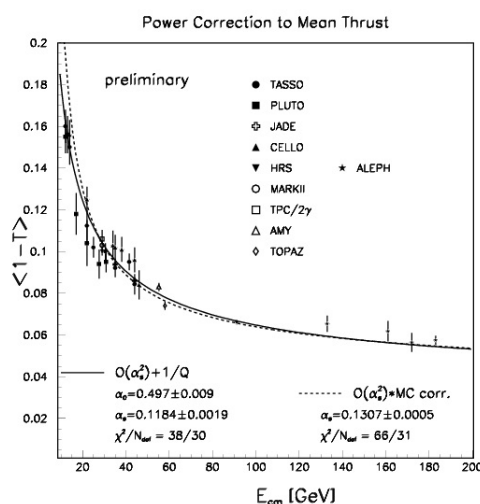


**Fig. 19:** Collection of average Thrust measurements in $e^+e^-$ collisions at different collider energies. Actually, they are reported in terms of the variable $t = 1 - T$. As predicted by the QCD, the Thrust variable approaches 1 for increasing energies (so, correspondingly $t$ tend to vanish). The solid line represents a fit which includes power correction of order $\Lambda/Q$ to be compared to the dashed line fit (which does not include them).

to the (soft) hadronization process of order $\Lambda/Q$, are taken into account. Indeed, the emission of a pion with few hundred MeV transverse momentum, involving the strong coupling at such small scales, has probability 1. Using Eq.(61) the effect on the average Thrust measurement at LEP energies induced by such emission can be easily estimated to be of about $8\%$. This effect is parametrised fitting many data with different hadronization models. Going more differentially, perturbative QCD also predicts that the Thrust distribution must become more and more peaked in 1 with increasing energies as it is well verified experimentally (see Fig.(20), where the variable $t = 1 - T$ is represented). Thrust distribution can be used to measure the strong coupling constant. The fit range is the result of a balance among: size of the perturbative corrections, quality of data and size of the hadronization corrections that are usually estimated comparing predictions obtained using different Monte Carlo programs. A typical choice at
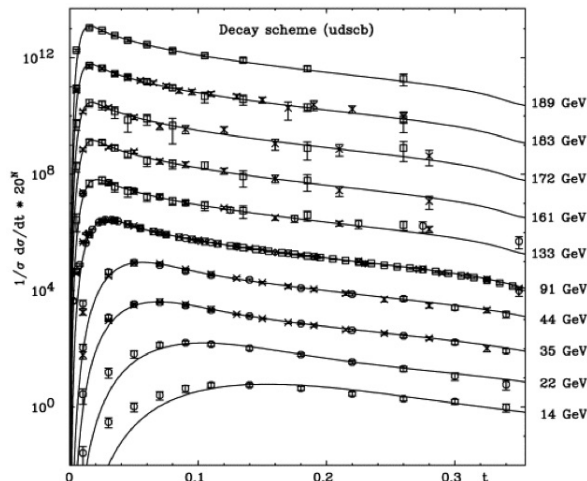
**Fig. 20:** Measurements of normalized Thrust distributions wrt to the variable $t = 1 - T$ for collisions at several energies. As predicted by QCD, the shape of the distribution becomes more peaked towards small values of $t$ as the energy is increased.

LEP is shown in Fig.(21). The Thrust variable is only one example of a number of shape variables, more or less sensitive to the radiative corrections and so to the strong coupling constant. We will not discuss them further here and just point out that for the InfraRed and Collinear safety of the shape variables, it is the linearity wrt the particle momenta to play the main role, as in Eq.(61).

### 1.5.2 Jet cross sections

Experimentally, high energy $e^+e^-$ collisions into hadrons produce for the large part two jet-like events, with two collimated back-to-back sprays of high energy hadrons, see Fig.(22). Then there is a lower number of events with three jets of particles, a much lower number of events where four jet structures can be identified and so on. This is in line with the expectations of perturbative QCD: we can imagine that the jet structures we see are in a one to one correspondence, and so they are the evolution, of the high energy partons that took part to the primary hard scattering. At high energy the coupling is relatively small and so the lower rate measured for increasing number of jets reflects the exponent of the coupling constant in the matrix elements squared for the production of such multi-parton final states. This is an heuristic argument, but for quantitative studies, one needs of a precise definition of a jet, which in turn is a prescription for grouping particles together. In particular, to the aim of comparing perturbative QCD predictions in terms of "partons" to measurements of collimated sprays of "paticles" we need a prescription that can be applied to both. Furthermore, for the theory prediction to be meaningful, one must be sure that the definition of a jet is inclusive enough wrt to the low energy particles in such way not to spoil the cancellation of long distance logarithmic divergences. In general, a jet algorithm must fulfill the following requirements: InfraRed (IR) and Collinear (C) safety, they should be simple to implement in both experimental analyses and theory predictions, and they should require small hadronization (non-perturbative) corrections. A way to proceed is to define:

1. a distance between particles, $d_{ij}$;
2. a merging scheme to merge particles;

Then, one has to compute all the distances between particles (protojets) and merge the ones with minimum distance. This procedure is iterated until a fixed minimum distance among all protojets is reached.
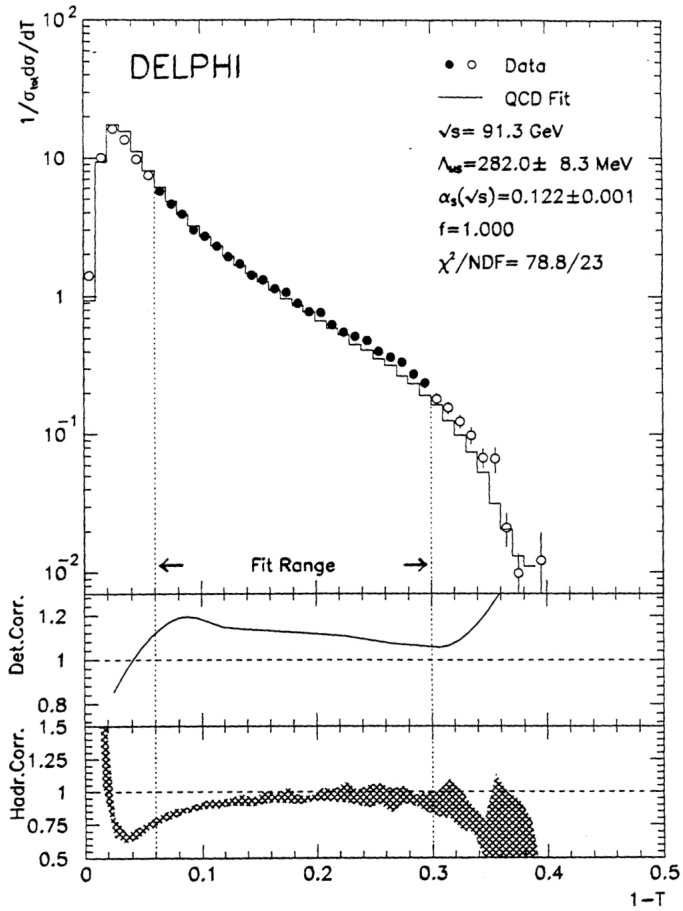
Figure 3:

**Fig. 21:** DELPHI dataset of measurements of the Thrust distribution to extract the strong coupling constant. The fit range is selected to minimize the impact of the detector and hadronization corrections.
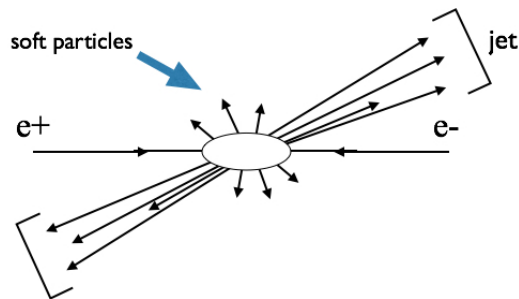


**Fig. 22:** Sketch of a two-jets event in $e^+e^-$ annihilation.

Such a procedure is called an iterative clustering algorithm and unambiguously assigns particles to jets. As for IR and C safety, what matters is that the distance definition goes to zero if $E_i \to 0$ or $\theta_{ij} \to 0$. The number of jets will have, of course, a predicted dependence on the value of the fixed minumin distance (resolution) value.

The original JADE algorithm [13], for example, starts from a distance defined as $d_{ij} = 2E_i E_i (1 - \cos\theta_{ij})$ that for massless $i, j$ particles represents the invariant mass squared of the pair. This is a dimensional variable, one can normalize it to the total energy of the event squared ($Q^2$) building the variable $y_{ij} = d_{ij}/Q^2$ and fix the resolution variable to some relatively small value $y_{cut}$ so that the algorithm will proceed up to the situation in which $y_{ij} > y_{cut}$ for every pair of protojets $i$ and $j$. Note, however, that with this algorithm soft particles turn out to be strongly correlated so that, even if they are at large angle, they are merged in the same jet and eventually can unnaturally produce a sort of sparse jet. With the kT algorithm [14], the distance measure is instead set to $d_{ij} = 2 \mathrm{Min}\{E_i^2, E_j^2\}(1 - \cos\theta_{ij})$ that, in the small angle limit, tends to represent the transverse momentum of one particle wrt the direction of the other. Note that with this definition the distance $d_{ij}$ is diagonal wrt particle energy. Soft particles are merged with the hard particle closest in angle (soft fragments are likely to be merged with their parent) avoiding unnatural assignments with creation of soft and wide angle jets. As for the merging scheme, sensible choices are simply to add the four momenta of the protojets (E-scheme) or first add and then rescale the space component of the sum to make again the new protojet massless (E0-scheme). Finally, once finished the assignment of all the particles to each jet one has to choose the way to assign the momentum to the final jets. A typical choice is to take the sum of the four momenta of the particles entering each jet. QCD predictions for jets rates, like their dependence on the algorithm and for each algorithm the increase of multijet rates by increasing the jet resolution ($y_{cut} \to 0$), have been nicely confirmed at LEP, for example. Furthermore, using a shower Monte Carlo it is possible to compare parton vs hadron jet rates to provide an estimation of non perturbative corrections. This, for example, motivates the introduction of the E0-scheme over the E-scheme, as fairly shown in the left panel of Fig.(23). In this figure the continuous line represents results at hadron level that reproduce well the data. Hadronization corrections turn out to be smaller for the kT algorithm (named D-scheme in the right panel Fig.(23)). Indeed, the kT algorithm is theoretically favoured also because the logarithmic unbalanced corrections that appear for small $y_{cut}$, can be resummed to all orders in perturbation theory. We conclude this section
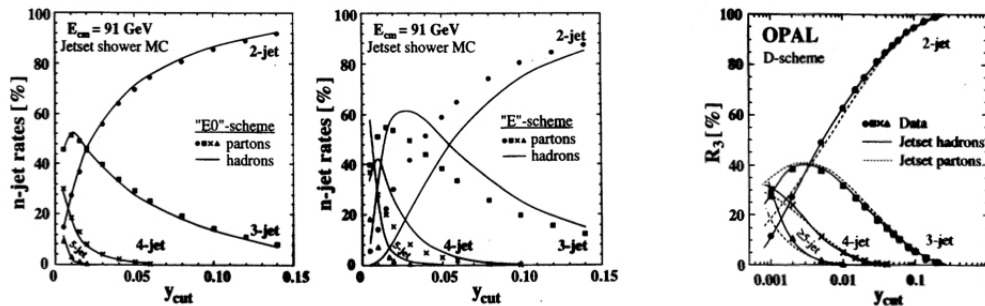


**Fig. 23:** Studies on jets reconstruction algorithms at LEP. Jets are reconstructed both at parton level (dots on left, dashed line on the right) and at hadron level (solid line). In the left panel, the E-scheme and the E0-scheme are compared: the latter has clearly better performances wrt to the hadronization corrections. In the right panel, the results for the kT algorithm.

showing in Fig.(24) [1] the nice combination of inclusive and differential measurements that prove the running of the strong coupling constant. In this figure there are also measurements extracted from deep inelastic lepton-hadron and hadron-hadron scattering, that will be the subject of the next sections, and the agreement of all of them with the predicted running is extremely good.

## 2   Hadrons in the initial state

The presence of hadrons in the initial state of a collision greatly complicates the picture. Nevertheless, for scattering processes characterized by a large momentum transfer, a description in terms of the parton
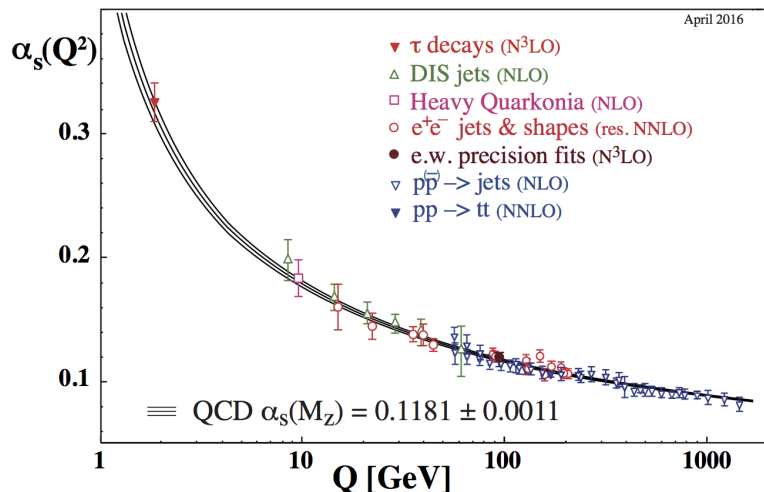
**Fig. 24:** Compilation of measurements of the strong coupling constant in a wide range of energies and for different physical processes. The solid lines represent the perturvative QCD prediction with its uncertainty bands.

model formulas is possible. Here "large" has to be understood as much larger than the characteristic hadronic binding energy. In this regime, it can be effectively thought that a single parton (quarks and gluons), and not the whole hadron, is taking part to the scattering process. Translating in a time-picture: if the scattering occurs on a time scale much smaller than the characteristic time of the processes that regulate the hadron dynamics, the system cannot respond to the interaction in a coherent way and a single parton is struck while the rest is frozen. So that, one assumes in general that a flux of composed particles (like atoms, nuclei or positronium bound states) is equivalent to a flux of their constituents carring just a fraction of the total moment with a distribution dictated by the intensity of the interaction among the constituents. Boosting the bound state particles along one direction produces a spread of the longitudinal component of the momentum while the transverse component remains the same as for the particle at rest (negligible after the boost). According to this picture, the parton formulas are convolutions of parton density functions (pdf's) and partonic scattering cross sections (with hat to distinguish them from particle cross sections). In case of hadron-hadron collision it reads:

$$\sigma_{H_1 H_2}(p_1, p_2) = \sum_{i,j} \int dx_1 dx_2 f_i^{H_1}(x_1) f_j^{H_2}(x_2) \widehat{\sigma}_{ij}(x_1 p_1, x_2 p_2) \tag{62}$$

while for lepton-hadron collision one has:

$$\sigma_H(p, \gamma^*) = \sum_i \int dx f_i^H(x) \widehat{\sigma}_{i\gamma^*}(xp). \tag{63}$$

In summary, $x$ in the formulas above is the fraction of the total momentum carried by the constituent partons and the variance of the fraction depends on the interaction strength. This scheme will work as far as we have large scattering angles (at small angle partons could scatter coherently) or, simply, as far as we can neglect the transverse momentum of the partons. With these formulas it is possible to make predictions at leading order for any scattering process involving hadrons in the initial state once the parton densities have been extracted from some measurement.

## 2.1   Naive parton model and Deep Inelastic Scattering

Let's start considering electron-hadron Deep Inelastic Scattering (DIS), namely the process

$$e^-(k) + P(p) \to e^-(k') + X$$

where we label with $X$ any possible hadronic final state resulting from the fragmentation of the incoming proton as shown Fig.(25). Indeed, in a typical DIS experiment, one mainly measures the scattered leptons, inclusively wrt to the possible hadronic final states. The kinematics of the DIS scattering is
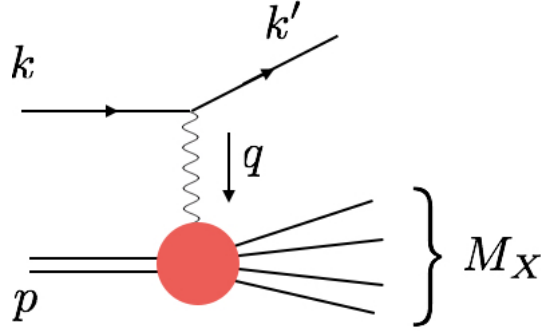


**Fig. 25:** Kinematics of electron-proton DIS scattering.

usually described in terms of the following variables: the square of the total energy in the CoM frame $S = (k+p)^2 = 2k \cdot p + M_P^2 \simeq 2k \cdot p$ (neglecting the proton mass $M_P$ wrt to the energy of the incoming electron), the time-like momentum transfer $q = k - k'$, its virtuality taken with opposite sign to have the positive quantity $Q^2 = -q^2 > 0$, and the invariant scalar product $\nu = p \cdot q$. In terms of these variables, it is customary to introduce the dimensionless fractions:

$$x = \frac{Q^2}{2\nu} \qquad \text{and} \qquad y = \frac{q \cdot p}{k \cdot p} = \frac{2\nu}{S} \tag{64}$$

which are related to the energy transfer and the scattering angle of the outgoing electron in the lab frame. Then, one has that the invariant mass of the recoiling system $X$ is given by:

$$M_X^2 = (p + q)^2 = Q^2 \frac{1 - x}{x}. \tag{65}$$

Note that, measuring the scattering angle and the energy of the outgoing electron, all the invariants defined above can be computed. What is observed in nature is that in the limit of large $Q^2$ and finite $x$ the differential cross section scales with energy as:

$$\frac{d\sigma}{dxdy} \sim \frac{1}{Q^2} f(x, y) \tag{66}$$

with no hadronic scale involved!
Going to the parton language, one has to consider the elementary scattering process of a virtual photon off a massless quark as represented in Fig.(26). Again, we will use the hat to indicate partonic momenta,
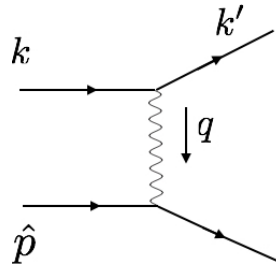


**Fig. 26:** Lowest order partonic elementary process contributing to the electron-proton DIS.

kinematic invariants and cross sections. The partonic variables are now:

$$\hat{s} = (k + \hat{p})^2 = 2\,k \cdot \hat{p} \qquad\qquad \hat{y} = \frac{q \cdot \hat{p}}{k \cdot \hat{p}} \tag{67}$$

and we impose the massless condition for the final state quark, $(q + \hat{p})^2 = 2\,q \cdot \hat{p} - Q^2 = 0$. It is easy to show that the partonic cross section is then:

$$\frac{d\hat{\sigma}_i}{d\hat{y}} = q_i^2 \frac{\hat{s}}{Q^4} 2\pi\alpha_{em}^2 \left[1 + (1 + \hat{y})^2\right] \tag{68}$$

where $q_i$ are the partonic electric charge in units of the electron charge. Now, neglecting parton transverse momentum and assuming that the proton is a beam of partons carring each a certain fraction $\tilde{x}$ of the total hadron momentum, we have $\hat{p} = \tilde{x}\,p$, $\hat{y} = y$. From the massless quark condition $(2\,\tilde{x}\,p \cdot q - Q^2 = 0)$ we find that the particle level and measurable $x = Q^2/2p \cdot q$ variable, coincides with the fraction of momentum $\tilde{x}$ carried by the scattered parton, so that we write:

$$\frac{d\sigma}{dxdy} = \sum_i f_i^H(x) \frac{d\hat{\sigma}_i}{d\hat{y}} \tag{69}$$

that inserting the expression for the partonic cross section gives:

$$\frac{d\sigma}{dydx} = \frac{2\,\pi\,\alpha_e^2\,S\,x}{Q^4} \left[1 + (1 + y)^2\right] \sum_i q_i^2 f_i^H(x). \tag{70}$$

The expression above predicts the mentioned scaling behaviour and also the full $y$ dependence. It also shows that, of course, deep inelastic scattering is a good place to extract the parton densities $f_i^H(x)$, that are hadron specific. Indeed, the measured quantity is:

$$F_2^H(x) = x \sum_i q_i^2\, f_i^H(x) \tag{71}$$

Assuming isospin symmetry, i.e. that a neutron ($n$) is like a proton ($p$) with up and down quarks exchanged, we can combine data for electron scattering off protons and neutrons (deuterons) to extract $u(x)$ and $d(x)$:

$$F_2^p(x) = x \left(\frac{4}{9} u_p(x) + \frac{1}{9} d_p(x)\right) \tag{72}$$

$$F_2^n(x) = x \left(\frac{4}{9} u_n(x) + \frac{1}{9} d_n(x)\right) = x \left(\frac{4}{9} d_p(x) + \frac{1}{9} u_p(x)\right). \tag{73}$$

Note, however that now the interaction can create also quark pairs inside the hadron (very different from a non relativistic bound state like an hydrogen atom), so that indeed we are measuring $u(x) + \bar{u}(x)$ and $d(x) + \bar{d}(x)$, because the photon cannot distinguish the sign of the charge. Furthermore, each distribution can also diverge for very low momentum fraction, but still one has that sum rules related for example to the fermion number conservation:

$$\int_0^1 dx\,(u_p(x) - \bar{u}_p(x)) = 2 \qquad\qquad \int_0^1 dx\,(d_p(x) - \bar{d}_p(x)) = 1. \tag{74}$$

must hold. To get the difference among $u$ and $\bar{u}$ we measure the structure functions related to charged current electroweak interactions like for neutrino DIS or $W^\pm$ production in hadronic collisions. As a result of this exercise to extract $u$, $\bar{u}$, $d$ and $\bar{d}$ parton densities, one can look at the total momentum of the proton obtained summing over all the fractions:

$$P = \sum_i \int_0^1 dx\,x\,p_i \qquad p_i = u,\,\bar{u},\,d,\,\bar{d}. \tag{75}$$

This sum turns out to be smaller then 1 and it amounts to about 1/2. We deduce once more that in the proton there are not only quarks but also something else that must interact rarely with photons (eventually indirectly), i.e. the gluons. To properly quantify the impact of the gluons in the description of the content of the proton probed at high momentum transfer we have to include radiative corrections.

## 2.2   Radiative corrections

The inclusion of the radiative corrections changes the parton model formulas given above. This is the reason why the formulas above are said the naive parton model formulas, while after including radiative corrections we have the so called improved parton model formulas. For hadron-hadron collisions for example we have:

$$\sigma_{H_1 H_2}(p_1, p_2) = \sum_{i,j} \int dx_1 dx_2 f_i^{H_1}(x_1, \mu) f_j^{H_2}(x_2, \mu) \widehat{\sigma}_{ij}(x_1 p_1, x_2 p_2, \mu). \tag{76}$$

This formula has a similar form wrt the naive parton model formula, but the various factors have a different meaning. First of all, we introduce an arbitrary scale $\mu$, which enters in all the factors in the rhs. In principle, we could distinguish among two different arbitary scales: the renormalization scale and another arbitrary scale to be called the factorization scale. However, as we will see in the following, in general it will be better to set them close to each other and in the range of the hard scale of the process, so we will use just one scale for the moment. The partonic cross section now has to be interpreted as a short distance cross section calculable in perturbation theory:

$$\hat{\sigma}_{ij}(x_1 p_1, x_2 p_2, \mu) = \sum_l \hat{\sigma}_{ij}^l(x_1 p_1, x_2 p_2, \mu)(\alpha_s(\mu))^l. \tag{77}$$

Furthermore, the dependence of the pdfs, $f_i^H(x, \mu)$, upon $\mu$ is mild and calculable with the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equation:

$$\frac{d}{d \log \mu^2} f_i^H(x, \mu) = \int_x^1 \frac{dz}{z} \sum_j P_{ij}(\alpha_s(\mu), z) f_j^H(x/z, \mu) \tag{78}$$

where the integral kernels $P_{ij}$ are the Altarelli-Parisi splitting functions, also calculable in perturbation theory with a series expansion of the form:

$$P_{ij}(\alpha_s(\mu), z) = \frac{\alpha_s(\mu)}{2\pi} P_{ij}^0(z) + \left(\frac{\alpha_s(\mu)}{2\pi}\right)^2 P_{ij}^1(z) + ... \tag{79}$$

Finally, for small variations, the $\mu$ dependence in the pdfs and in the short distance cross section compensate. As stated above, the scale $\mu$ has to be chosen close to the hard scale of the process, this avoids the presence of large logarithmic corrections in the short distance cross section $\hat{\sigma}_{ij}$ (similarly to the renormalisation logarithm in $e^+ e^-$ annihilation). There will be, of course, divergences associated with radiation. We expect that soft and collinear divergences associated to gluon radiation off the final state will cancel (among the real and virtual contributions) as for $e^+ e^-$ annihilation into hadrons. For the initial state the situation is different: the collinear divergences of the real and virtual parts are associated to different kinematical configurations, in particular to final states with different invariant masses, and eventually do not cancel each other! This is a very different situation with respect to $e^+ e^-$ annihilation. Let's extract the most divergent part of the amplitude squared related to the initial emission as represented in Fig.(27).

We start from the kinematics. Imagine that the initial parton with momentum $\hat{p}$ is travelling along the $z$ axis, then, to parametrize the divergences, we decompose the momenta as follows:

$$l = (1 - x)\hat{p} + l_\perp + \xi\eta \tag{80}$$
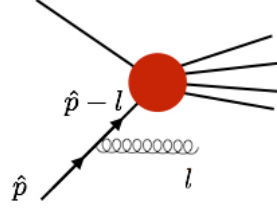
**Fig. 27:** Gluon radiation off an incoming quark line in a generic scattering process involving partons. In the strictly collinear limit, the amplitude exactly factorizes in a universal factor associated to the gluon emission process times the amplitude corresponding to the process without the radiation of the extra gluon.

with:
$$\hat{p} = (p_0, 0_\perp, p_0), \qquad \eta = (p_0, 0_\perp, -p_0), \qquad l_\perp = (0, \vec{l}_\perp, 0). \tag{81}$$

From the condition $l^2 = 0$ one has:
$$\xi = \frac{\vec{l}_\perp^2}{2\,\hat{p} \cdot \eta\,(1 - x)} \tag{82}$$

and for the vanishing denominator of the propagator one has:
$$(\hat{p} - l)^2 = -2\,\hat{p} \cdot l = -2\,\hat{p} \cdot \eta\,\xi = -\frac{\vec{l}_\perp^2}{1 - x}. \tag{83}$$

For the phase space factor one has:
$$\frac{d^3 l}{2\,l^0\,(2\,\pi)^3} = \frac{d^2 l_\perp}{2\,(2\,\pi)^3}\frac{dx}{1 - x}. \tag{84}$$

The whole amplitude for the scattering process, that we generically indicate with $A$, is:
$$A = M \frac{\hat{\not{p}} - \not{l}}{(\hat{p} - l)^2} \not{\epsilon}\, u(\hat{p}). \tag{85}$$

In the above formula, we have stripped out the colour matrix $t^A$ which will contribute with the colour factor $C_F$ and we have denoted with $M$ all the remaining part of the amplitude attached to the incoming quark line where the gluon emission occurs (which corresponds to the red blob and the other external legs in Fig.(27)). In the strict collinear limit $l_\perp \to 0$, the squared of the momentum $\hat{p} - l$ in the denominator is vanishing, exposing a singularity at the level of the amplitude squared which we would naively expect to behave as $d^2\vec{l}_\perp/\vec{l}_\perp^4$. However, this is not the case because when $l$ becomes parallel to $p$ also the scalar product of $p$ and the gluon polarisation vector $\epsilon$ (that as we have seen is transverse for any physical gauge) vanishes. Indeed, the conservation of angular momentum provides a more general argument: there is no helicity flip in the gluon emission in a quark-quark-gluon vertex, so that the vectorial nature of the gluon provides suppression. With a bit of Dirac algebra one can show that:
$$(\hat{\not{p}} - \not{l})\not{\epsilon}\, u(\hat{p}) \simeq -\frac{2\,\epsilon \cdot l_\perp}{1 - x} u(\hat{p}) + \not{\epsilon}\not{l}_\perp u(\hat{p}) = O(l_\perp). \tag{86}$$

The net effect is to make milder the behavior of the singularity which is actually logarithmic. Then, taking all together, the divergent part of the amplitude squared, the flux factor, the phase space factor, the colour factor, and integrating over the phase space, one has (from now on $l_\perp$ will denote the modulus of the bidimensional momentum $\vec{l}_\perp$):
$$\sigma_R^{(1)} = C_F \frac{\alpha_s}{2\,\pi} \int \sigma^{(0)}(x\hat{p}) \frac{1 + x^2}{1 - x} \frac{dl_\perp^2}{l_\perp^2} dx. \tag{87}$$

In this formula $\sigma^{(0)}$ represents the amplitude squared of the process after the initial gluon emission ($|M|^2$) with the momentum of the entering parton $\hat{p}$ reduced by a factor $x$ by the emission, so that also the (invariant) center of mass energy of the collision is of course reduced by the same amount. This result actually is a consequence of the more general factorization property of the amplitudes in a gauge theory, which we mentioned in a previous section. The divergent part of the virtual correction can be recast in a similar form and all together the real and the virtual add to:

$$\sigma^{(1)} = C_F \frac{\alpha_s}{2\pi} \int \left( \sigma^{(0)}(x\hat{p}) - \sigma^{(0)}(\hat{p}) \right) \frac{1+x^2}{1-x} \frac{dl_\perp^2}{l_\perp^2} dx. \tag{88}$$

So, we see that the soft divergence ($x \to 1$) cancels, while the collinear one ($l_\perp \to 0$) does not! Once again we can imagine that a lower cutoff scale $\lambda$ of about 1 GeV regulates this divergence and that below that energy scale we enter another regime of the strong interaction. However, even in that case, our ability to make predictions in the perturbative regime is challenged because this construction implies a strong sensitivity to this IR cutoff $\lambda$. Note, that higher order corrections worsen the situation because the effective expansion parameter will be of order 1. Think of multiple ($n$) radiations over large $l_\perp$ gap, they will contribute to the event probability with a factor:

$$\left[ \alpha_s(Q^2) \int_{\lambda^2}^{Q^2} \frac{dl_\perp^2}{l_\perp^2} \right]^n \tag{89}$$

where $Q$ represents the hard scale of the process and each factor in the square parenthesis evaluates to:

$$\alpha_s(Q^2) \int_{\lambda^2}^{Q^2} \frac{dl_\perp^2}{l_\perp^2} = \alpha_s(Q^2) \log \frac{Q^2}{\lambda^2} \simeq \alpha_s(Q^2) \frac{1}{\alpha_s(Q^2)} = \mathcal{O}(1). \tag{90}$$

Regularisation of the collinear divergence will definitely not be enough, we will need of the resummation of the leading remaining effects. The problem is, indeed, solved with a renormalisation, similarly to the way we did the ultraviolet renormalisation for $e^+e^-$ and it will work at the same time for all the observables receiving QCD corrections in hadron initiated collisions. In summary, the collinear not cancelled divergence is related to the vanishing of a denominator representing a massless propagator that is going on-shell. Furthermore the amplitude exactly factorises in the collinear limit into two amplitudes, one with the initial radiation and another one with the hard collision. The vanishing propagator implies the presence of a long distance physics phenomenon in the problem. Indeed, the collinear emission might well happen quite far from the hard interaction and it can be considered as another process among the ones that regulates the life of the hadron. This makes a big difference with respect to the naive parton model because, due to radiative corrections, the transverse momentum of the partons inside an hadron is not limited to $\Lambda$ and in principle can be very large. We will now adsorb this collinear divergence into the definition of the parton distribution functions that make up the hadron. First, we rewrite Eq.(88) introducing an infrared cutoff $\lambda$ to regulate the low $l_\perp$ emissions and we adopt the plus notation for the soft singularity getting a more compact form. Indeed, we observe that in Eq.(88), the function with a non-integrable singularity:

$$\frac{1+x^2}{1-x} \tag{91}$$

multiplies the difference among a regular function of the variable $x$ and the same function evaluated in $x = 1$, in a such a way that the non integrable singularity is regularized. In analogy, a plus distribution version of the function in Eq.(91) is defined by the action on a generic function $f$ as follows:

$$\int_0^1 \left( \frac{1+x^2}{1-x} \right)_+ f(x) dx = \int_0^1 \frac{1+x^2}{1-x} \left( f(x) - f(1) \right) \tag{92}$$

so that we can recast Eq.(88) in the form:

$$\sigma^{(1)} = \frac{\alpha_s}{2\pi} \log \frac{Q^2}{\lambda^2} \int \sigma^{(0)}(x\hat{p}) P_{qq}^0(x) dx. \tag{93}$$

The introduction of the plus distribution is useful also because the initial splitting functions are always the same for every process with any number of final state particles, in other words they are universal. In Eq.(93) we have also collected the colour factor and the plus distribution into the function $P_{qq}^0$. No surprise, this is once again the splitting function we have introduced in Eq.(49), while we were talking about $e^+e^-$ into hadrons. Here there are two small differences: in Eq.(93) the $x$ variable is the fraction of the original quark momentum carried by the quark after the splitting, while in Eq.(49) the fraction refers to the outgoing gluon and in fact to go from one expression to the other we have to exchange $x \to 1 - x$, then the presence of the plus prescription that comes from the inclusion of the virtual correction. Note also that we have changed notation for the subscript, talking about the initial state we call $P_{ab}$ the splitting probability to produce a parton of kind $a$ from the decay of a parton of kind $b$, or even, to find a parton $a$ inside a parton $b$, with fraction $x$ of the momentum of parton $b$. Now we insert into the game the other (arbitrary) scale $\mu$, that we mentioned at the beginning of the section, to split the logarithm. Neglecting terms of order $\alpha_s^2$, we can write:

$$\sigma^{(0)}(\hat{p}) + \sigma^{(1)}(\hat{p}) = \int dx \left( \delta(1-x) + \frac{\alpha_s}{2\pi} \log \frac{\mu^2}{\lambda^2} P_{qq}^0(x) \right) \hat{\sigma}(x\hat{p}, \mu^2) \tag{94}$$

with:

$$\hat{\sigma}(x\hat{p}, \mu^2) = \hat{\sigma}^0(x\hat{p}) + \frac{\alpha_S}{2\pi} \log \frac{Q^2}{\mu^2} \int dz P_{qq}^0(z) \sigma^{(0)}(zx\hat{p}) \tag{95}$$

plus other finite terms. The scale $\mu$ is the factorisation scale and again, as we argue from Eq.(95), it will be better to choose it of the same order as $Q$ to avoid large logarithmic corrections in $\hat{\sigma}(x\hat{p}, \mu^2)$. Now we convolute the corrected short distance cross section with the parton density getting:

$$\sigma(p) = \int dy dx f_q(y) \left( \delta(1-x) + \frac{\alpha_S}{2\pi} \log \frac{\mu^2}{\lambda^2} P_{qq}^0(x) \right) \hat{\sigma}(xyp, \mu^2). \tag{96}$$

Finally, we make the last manipulation introducing a delta function to get:

$$\sigma(p) = \int dz \widetilde{f}_q(z, \mu^2) \hat{\sigma}(zp, \mu^2) \tag{97}$$

with:

$$\widetilde{f}_q(z, \mu^2) = \int dy dx f_q(y) \left( \delta(1-x) + \frac{\alpha_S}{2\pi} \log \frac{\mu^2}{\lambda^2} P_{qq}^0(x). \right) \delta(z - xy) \tag{98}$$

The last step has been the adsorption of the large logarithmic correction into a redefinition of the parton density. From the universality of the splitting probabilities, it can be shown that this parton density redefinition does not depend upon the specific hard process, and so it's universal! Furthermore, there is a variety of arguments showing that the above construction holds to all orders in perturbation theory (factorisation theorem). In turn, we have to consider a parton as having a structure that depends upon the scale at which we are probing it ($\mu \sim Q$).

Nevertheless, we are now left with these large corrections that depend upon unknown low scale dynamics. The key observation here is that although low scale dynamics of parton density functions can only be measured, their scale dependence is predictable in perturbation theory. Indeed, neglecting terms of order $\alpha_S^2$ we have that:

$$\frac{d}{d\log\mu^2} \widetilde{f}_q(z, \mu^2) = \frac{\alpha_S}{2\pi} \int dy dx \widetilde{f}_q(y, \mu^2) P_{qq}^0(x) \delta(z - xy) + \mathcal{O}(\alpha_S^2) \tag{99}$$

that is the Altarelli-Parisi, or Dokshitzer-Gribov-Lipatov-Altarelli-Parisi equation [15]. Once the pdf's are measured in a certain set of measurements performed at a given energy scale, they can be used as the initial conditions for the Altarelli-Parisi equations. Then, the solution to these equations, corresponding to the effective resummation of the logarithms associated to multiple collinear emissions, provides the pdf's to all other scales that can be used to make predictions for processes at different energy scales.

## 2.3 Altarelli-Parisi splitting probabilities and evolution equations

Indeed, the program outlined at the end of the last section needs of the knowledge of the splitting functions for all the possible elementary processes and of the solution of the following system of integro-differential equations:

$$
\begin{aligned}
\frac{d}{d\log\mu^2} f_q(x,\mu^2) &= P_{qq} \otimes f_q + P_{q\bar{q}} \otimes f_{\bar{q}} + P_{qg} \otimes f_g \\
\frac{d}{d\log\mu^2} f_g(x,\mu^2) &= P_{gq} \otimes f_q + P_{g\bar{q}} \otimes f_{\bar{q}} + P_{gg} \otimes f_g
\end{aligned}
\tag{100}
$$

where we have used the $\otimes$ symbol to represent the convolution integral:

$$
f \otimes g \equiv \int dy\, dz\, f(y) g(z) \delta(x - yx) = \int_x^1 \frac{dz}{z} f\left(\frac{x}{z}\right) g(z).
\tag{101}
$$

The computation of the lowest order splitting functions is not difficult and can be done following the steps of the previous paragraph also for the other splitting processes. The results are given in Fig.(28). These



$$
P_{qq}(z) = C_F \left(\frac{1+z^2}{1-z}\right)_+
$$

$$
P_{gg}(z) = 2C_A \left[\frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z)\right]
$$
$$
+ \delta(1-z)\frac{1}{6}(11C_A - 2n_f)
$$

$$
P_{qg} = T_R[z^2 + (1-z)^2]
$$

$$
P_{gq}(z) = C_F \left[\frac{1 + (1-z)^2}{z}\right]
$$

**Fig. 28:** Lowest order Altarelli-Parisi splitting kernels.

functions fulfil a number of properties: first, $P_{qq}$ and $P_{gg}$ exhibit the soft singularity at $z = 1$, while of course $P_{gq}$ has it for $z = 0$, all of them are positive definite for $z < 1$, the real parts are connected by final state parton exchange, it's easy to prove also the crossing symmetry $z \to 1/z$ keeping into proper account the number of states while building the colour/spin averages. Furthermore, the evolution does not spoil the sum rules of course. Considering the proton, we report again the flavour sum rules:

$$
\int_0^1 \left[f_u(x,\mu^2) - f_{\bar{u}}(x,\mu^2)\right] = 2
\tag{102}
$$

$$
\int_0^1 \left[f_d(x,\mu^2) - f_{\bar{d}}(x,\mu^2)\right] = 1.
\tag{103}
$$

further, from momentum conservation we now must have:

$$\sum_a \int_0^1 dx \, x \, f_a(x, \mu^2) = 1. \tag{104}$$

Differentiating with respect to the scale $\mu$ one obtains the following conditions respectively:

$$\int_0^1 [P_{qa}(x) - P_{\bar{q}a}(x)] = 0 \qquad \forall \, a \tag{105}$$

$$\sum_a \int_0^1 dx \, x \, P_{ab}(x) = 0 \qquad \forall \, b \tag{106}$$

and the splitting probability functions quoted in Fig.(28) satisfy them. As a power series in the strong coupling, the splitting functions are known up tp Next-to-Next to Leading Order (NNLO). They contain terms proportional to $\log \mu^2$ but also to $\log 1/x$ and $\log(1-x)$. The leading order DGLAP evolution equation sums up the $(\alpha_S \log \mu^2)^n$ contributions, the Next to Leading order (NLO) sums the $\alpha_S(\alpha_S \log \mu^2)^{n-1}$ terms and so on. Small $x$ resummation has also been performed predicting a power low behaviour for the pdf's. In general the precision of the experimental data demands that at least NLO (and preferably NNLO) DGLAP evolution has to be used in comparisons between theory and experiment.

Finding a stable solution of the system of Eq's(100) is a non trivial numerical task that is usually addressed with the Runge-Kutta method. Nevertheless, for extreme situations we can predict the leading behaviour of the evolution. Let's consider the large $x$ limit first, $x \to 1$. In this case the dominant splittings are $P_{qq}$ and $P_{gg}$ so that the equations decouple and the leading terms in these two splitting functions are the two plus distributions that can be both represented as:

$$P_{aa} \simeq \frac{2 \, C_a}{(1-z)_+} \tag{107}$$

with $C_a$ equal to $C_F(C_A)$ for the quark q(gluon g). This is the limit of soft gluon radiation. Starting from the known pdf's at the scale $Q_0$, the evolution takes the form:

$$f_a(x, Q^2) \simeq f_a(x, Q_0^2) \exp \left\{ \int_x^1 dz \frac{2 \, C_a}{(1-z)_+} \int_{Q_0^2}^{Q^2} \frac{dq^2}{q^2} \frac{\alpha_S(q^2)}{2 \, \pi} \right\}. \tag{108}$$

From the definition of the plus distribution one has that:

$$\int_x^2 dz \frac{1}{(1-z)_+} = -\int_0^x \frac{dz}{1-z} = \log(1-x) \tag{109}$$

and inserting the running coupling we find:

$$f_a(x, Q^2) \simeq f_a(x, Q_0^2)(1-x)^{p_a} \qquad \text{with} \qquad p_a = \frac{C_a}{\pi \, b_0} \log \frac{\alpha_S(Q_0^2)}{\alpha_S(Q^2)}. \tag{110}$$

Then, we see that in the large $x$ limit, the pdf's with our approximations vanish with a power-law behavior. Increasing the scale the exponent slightly increases, so that the higher is the scale the flatter is the behaviour of the distribution. In the small $x$ limit instead, only the second term in $P_{gg}$ matters. The inclusion of the splitting described by $P_{gq}$ does not catch the leading singularity because it cannot give rise to a chain of all enhanced contributions. This is the limit driven by multiple soft gluon exchange so that only the equation for the gluons is relevant, with:

$$P_{gg} \simeq \frac{2 \, C_A}{z}. \tag{111}$$

In this case the solution takes the form:

$$xf_g(x, Q^2) \simeq xf_g(x, Q_0^2) + \exp\left\{\sqrt{\frac{2\,C_a}{\pi\,b_0}\log\frac{\alpha_S(Q_0^2)}{\alpha_S(Q^2)}\log\frac{1}{x}}\right\}. \tag{112}$$

Inspecting the exponent and knowing the running coupling, this expression predicts a steepness increasing with $Q^2$ at small $x$.

To compare with phenomenology, we go back to the proton $F_2$ structure function and now we move from the naive parton model prediction:

$$F_2(x) = x \sum_i q_i^2 \, f_i(x) \tag{113}$$

to the one including the leading radiative corrections (with $\mu^2 = Q^2$) resumed in the pdf evolution and get:

$$F_2(x, Q^2) = x \sum_i q_i^2 \, f_i(x, Q^2). \tag{114}$$
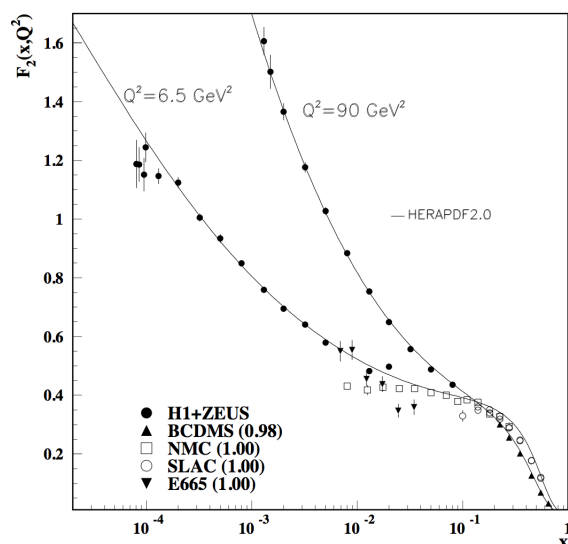


**Fig. 29:** Fit of the proton $F_2$ structure function as function of $x$ for two different scale $Q^2 = 6.5\,\text{GeV}^2$ and $Q^2 = 90\,\text{GeV}^2$. As predicted by the DGLAP evolution, with increasing energy, the pfds become steeper at small $x$ and flatter at large $x$.

In Fig.(29) [1] we can see the effects of the scale evolution mentioned above: the faster descent at high $x$ and the greater steepness at low $x$. The scaling violations together with the QCD analysis is also shown in Fig.(30). The best fit of the pdf's is performed combining a large amount of data for all kind of available collisions and at very different energies. In Fig.(31) a sketch of the kinematic domain of the available data used in pdf fits is shown, while in Fig.(32) a solution obtained by the NNPDF [16] collaboration is shown for two scale values, similar results are obtained from other collaborations (see for example ABM [17], JR [18], MSTW [19], MMHT [20] and other collaborations).

## 3   Hadronic collisions and jets

### 3.1   Minimum bias

Proton-proton and proton-antiproton colliders are mainly discovery machines, see Fig.(33). In these
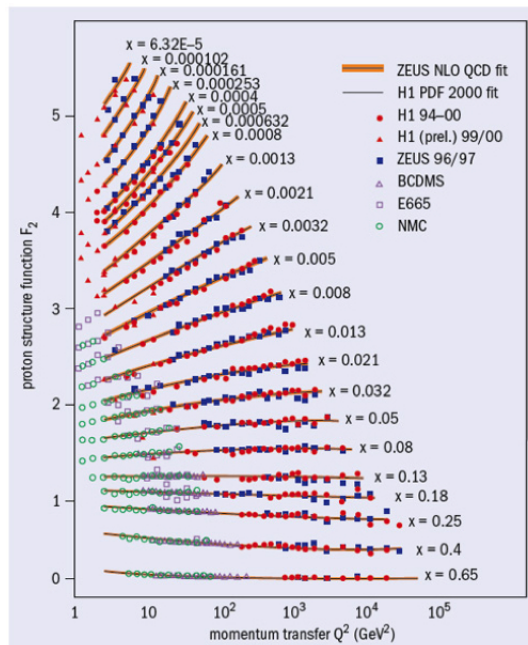
**Fig. 30:** Fits of the proton $F_2$ structure function as a function of the momentum transfer $Q^2$ for different values of the momentum fraction $x$. The simple scaling low with $Q^2$ at fixed $x$ predicted by the parton model breaks down due to radiative corrections that well explain the measured scaling violation.
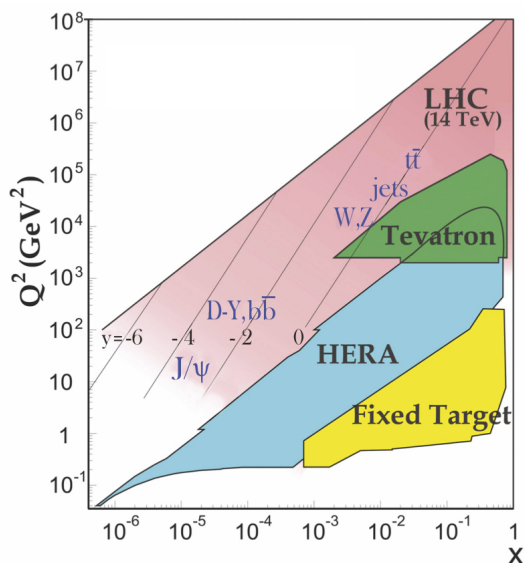


**Fig. 31:** Available data sets for the pdf fit in the plane $x - Q^2$.

experiments, QCD is ubiquitous, and this situation of course allows one to make many tests of the theory. Furthermore, a good knowledge of QCD effects is of course essential to establish signal and background rates. The variables used to describe hadron production in hadron-hadron collisions are represented in Fig.(34). The transverse plane is the plane orthogonal to the beams and the azimuthal angle is the azimuth around the beam direction. The transverse momentum is the projection of momentum on the transverse plane, $k_\perp = k|\sin\theta|$ and the transverse energy is the quantity $E_T = E|\sin\theta|$. Sometimes the transverse
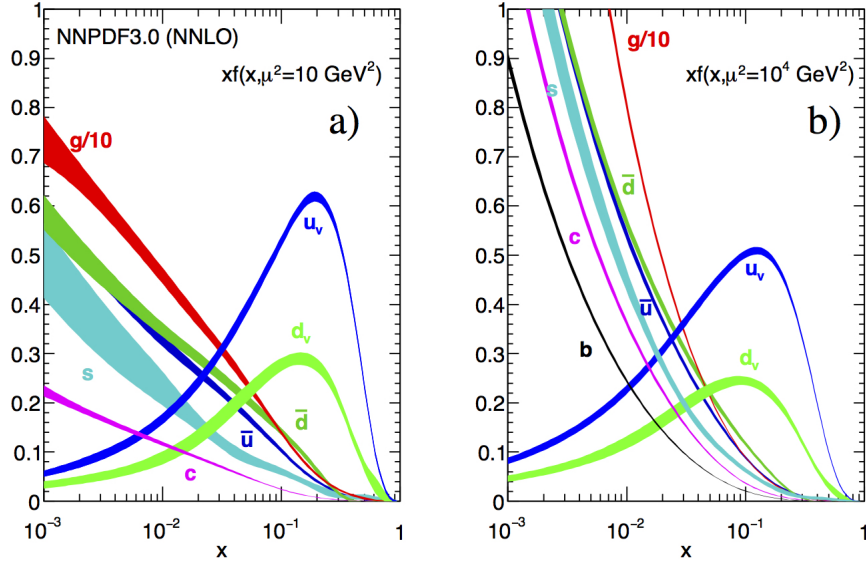
**Fig. 32:** Examples of pdf fits including DGLAP evolution at NNLO by the NNPDF collaborations. On the left, the pdfs at a scale $Q^2 = 10\,\mathrm{GeV}^2$ while on the right $Q^2 = 10^4\,\mathrm{GeV}^2$.

| Accelerator | SppS | TEVATRON | LHC |
|---|---|---|---|
| Energy | 200÷900GeV | 1.96TeV | 13TeV |
| Discovery | W/Z,Jets | top | Higgs |

**Fig. 33:** Summary of the main discoveries achieved at the proton-antiproton (SppS and Tevraton) and proton-proton colliders (LHC).
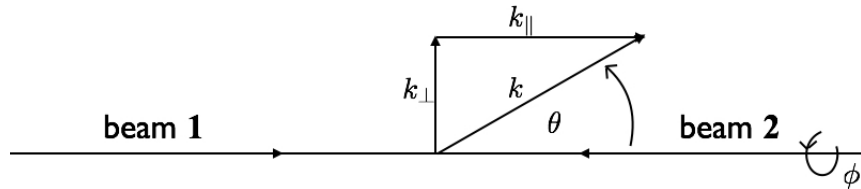


**Fig. 34:** Representation of the kinematical variables used in hadron-hadron collisions.

mass $m_T = \sqrt{k_\perp + m^2}$ is also used. The rapidity is defined by the formula:

$$y = \frac{1}{2} \log \frac{k^0 + k_\parallel}{k^0 - k_\parallel} \tag{115}$$

where $k^0$ is the energy of the detected particle. Under a boost in the beam direction it gets an additive contribution related to the boost velocity, so that rapidity differences are invariant under such boosts. This is particularly useful for the description of hadron-hadron collisions where there is always a boost among the hadronic and partonic cms. For massless particles the rapidity reduces to:

$$y = \frac{1}{2} \log \frac{1 + \cos\theta}{1 - \cos\theta} = -\log \tan \frac{\theta}{2}. \tag{116}$$
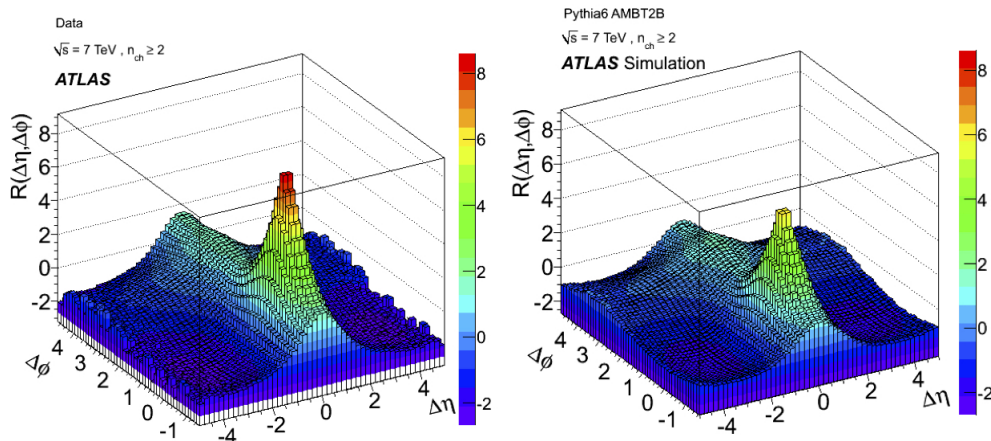
**Fig. 35:** Study of the angular correlations among hadrons in the Minimum Bias (main cuts: $k_\perp > 100\,\text{MeV}$ and $|\eta| < 2.5$) by the ATLAS collaboration. The data distribution (on the left) is compared to a simulated sample using the PYTHIA events generator (on the right). As stated in the main text, the simulation programs do not reproduce well the peculiar features of the data.

This variable refers directly to the scattering angle in hadronic cms frame and, as such, can be defined also for massive particles. In this case it is called pseudorapidity ($\eta$). Note however that $\eta$ differences are not anymore boost invariant so that it is not immediate (as for the case of rapidity) to compare pseudorapidity distributions, for example among the results of hadron collisions in fixed target experiment with collider results, and to this aim one has to reinsert back the mass of the measured particles. As we have seen in the first section, the cross section for hadronic collisions reflects the presence of a typical hadronic scale:

$$\sigma \simeq \frac{1}{(\text{Few hundred MeV})^2} \tag{117}$$

and it increases very slowly with the center of mass energy. These features are not computable in QCD although this behaviour is consistent with QCD. Consider quark-quark scattering down to a minimum exchanged virtuality $Q^2$, one has that:

$$\frac{d\sigma}{dQ^2} \sim \frac{1}{Q^4} \implies \int_{Q^2_{min}}^{Q^2_{max}} \frac{d\sigma}{dQ^2} dQ^2 \sim \frac{1}{Q^2_{min}} \tag{118}$$

so that, the typical scale of the non perturbative phenomena, naturally appears and QCD is in line with the fact that these kind of low energy interactions dominate. The bulk of the scattering events is then soft and rarely we assist to a collision characterized by a relatively large momentum transfer. We collectively call the soft collisions Minimum Bias (MB). A typical collision event at the LHC produces on average $80 \sim 100$ charged particles with huge fluctuations. This number grows only (a bit more then) logarithmically with the cms energy of the colliding protons. The transverse momentum distribution for a species of hadron with mass $m$ has an exponential descent behavior ($\sim \exp\left(-a\sqrt{m^2 + k_\perp^2}\right)$ with an average value of $\langle k_\perp \rangle \sim 600\,\text{MeV}$ which, again, grows slowly with energy. As for rapidity distributions, they are rather flat with a total of $5 \sim 6$ particles per unit of central rapidity. Keep in mind that for a particle of mass $\sim 1\,\text{GeV}$ produced at $E_{CM} = 13\,\text{TeV}$ the maximum absolute rapidity is about 9. Again, all these features are non computable in QCD, but are compatible with QCD. To understand this one can just start from heavy quark pair production with a mass $m \gg \Lambda$. Reducing the value of the mass, one gets a broadening of the rapidity distribution and the result that the typical transverse momentum is of the order of the mass of the produced object. The features of the minimum bias constitute the first study performed at hadron colliders, which starts considering the angular correlation among the

produced hadrons in a sufficiently large number of collisions, collecting particles up to very low values of the transverse momentum. In Fig.(35) [21], the left panel shows the result of a study by the ATLAS collaboration obtained collecting hadrons with $k_\perp > 100\,\text{MeV}$ and $|\eta| < 2.5$ (plus some other less relevant cuts). We can clearly see: 1. the emergence of a jet like structure at $\Delta\eta = \Delta\phi = 0$, 2. the emergence of the recoil of one particle against the other as a back to back jet like structure at $\Delta\eta = \pi$ for all $\Delta\phi$ and the 3. decay of resonances or other colour structures at $\Delta\eta$ of about 2 and for all $\Delta\phi$. Nevertheless, these three features are not well reproduced by simulation programs (see for example the right panel in Fig.(35)) and more investigation is ongoing on this point.

## 3.2 Jets

For jets there is now a very different situation with respect to $e^+e^-$ annihilations, where almost all events are back to back. From high energy partonic collisions we expect something similar, although now the higher the energy of the jets, the smaller the fraction of events. To identify jets something has to be done to distinguish them from the fluctuations of the minimum bias. Indeed, trigger is crucial to make discoveries! Think about the discovery of the $Z$ boson at the UA2 experiment with proton-antiproton collisions at $630\,\text{GeV}$ cm energy. In that case one has:

$$\frac{\sigma(Z)}{\sigma_{TOT}} \cdot Br(Z \to l^+l^-) \sim \frac{\frac{1}{3}\left(\frac{1}{300\text{GeV}}\right)^2}{\left(\frac{1}{300\text{MeV}}\right)^2} \cdot 0.1 \approx \frac{1}{3}10^{-7}. \tag{119}$$

Considering that one year has about $3 \cdot 10^7$ seconds, to have 1 $Z$ boson decaying to charged leptons per year, you need 1 proton collision per second, while to claim the discovery you need more, O(1000)! This means that you need to register events at a rate of the order of few thousands per second. This is not a problem today, with the modern systems of data acquisition, thirty or forty years ago it was different. Furthermore, as stated in the previous section, most of the collisions involved an overwhelming rate of soft QCD reactions and particles which we are not interested in and that can represent a real challenge for the acquisition electronics in terms of memory resources and time response. Then, what is needed is an efficient system to choose the events to store (the trigger). As a general guideline, one stores events with a final state characterized by particles with large transverse momentum, that is for sure the signal of a short distance interaction. One could think to trigger events on the base of the presence of very high energetic particles in an event, but that is not a very good idea because minimum bias events can have very high energetic particles at small angles. It is better to use a "transverse energy trigger" (energy suppressed by the angle):

$$E_T = \sum_i E_i |\sin\theta_i| \tag{120}$$

But still, minimum bias events at UA2 counted $\sim 25$ charged as well neutral hadrons, so having an average transverse energy $\langle E_T \rangle \approx 50 \cdot 500\,\text{MeV} \approx 25\,\text{GeV}$. To stay away from minimum bias events the experiment considered events with $E_T > 70\,\text{GeV}$ finding that all events are as in Fig.(36), i.e. jet-like! Most of the events have two jets, firmly establishing that jets in hadronic collisions are a property of nature. From the definitions of rapidity and transverse momentum it is easy to show that measuring $y_1$, $y_2$ and $p_\perp$ we fix the values of the fractions $x_1$ and $x_2$. We can make theory predictions using the improved parton model formula for two jet differential cross section:

$$\frac{d^3\sigma}{dy_1 dy_2 dp_\perp^2} = \frac{1}{16\pi S^2} \sum_{i,j,k,l} \frac{f_i(x_1,\mu)}{x_1} \frac{f_j(x_2,\mu)}{x_2} \overline{\sum} |M_{i,j,k,l}|^2 \frac{1}{1+\delta_{kl}}. \tag{121}$$

and parton densities extracted from DIS measurements. The formula above predicts that jets are back to back in azimuth and it also predicts the jets angular distribution in the cm frame of the two jets (or equivalently the $p_\perp$ dependence of the jets). This differential prediction is nicely verified experimentally,
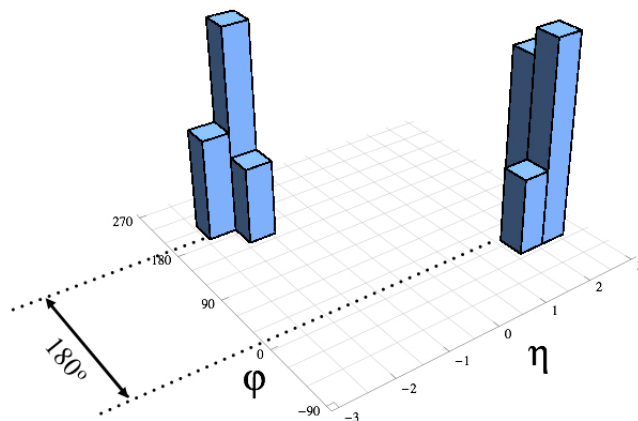
**Fig. 36:** Two jets event at UA2 experiment as emerging after imposing the hard cut on the transverse energy $E_T > 70\,\text{GeV}$ to separate the hard scattering events from the minimum bias.
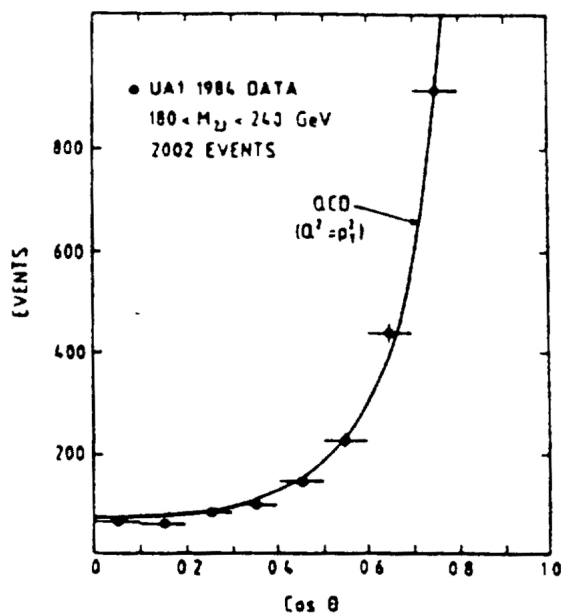


**Fig. 37:** Measurement of the jets scattering angle in the jets cm frame by the UA1 experiment. The solid line is the QCD prediction given by the improved parton model formula.

as it can be seen in Fig.(37). Furthermore, note that as a good first approximation, the partonic scattering cross sections stay with colour factors dictated as shown in Fig.(38). So that, the two jet cross section can be written in terms of a generalized structure function $F$ as:

$$\frac{d^3\sigma}{dx_1 dx_2 d\cos\theta} = \frac{F(x_1)}{x_1}\frac{F(x_2)}{x_2}\frac{d\hat{\sigma}_{gg\to gg}}{d\cos\theta} \tag{122}$$

with:

$$F(x) = f_g(x) + \frac{4}{9}\sum_q \left[f_q(x) + f_{\bar{q}}(x)\right]. \tag{123}$$

78
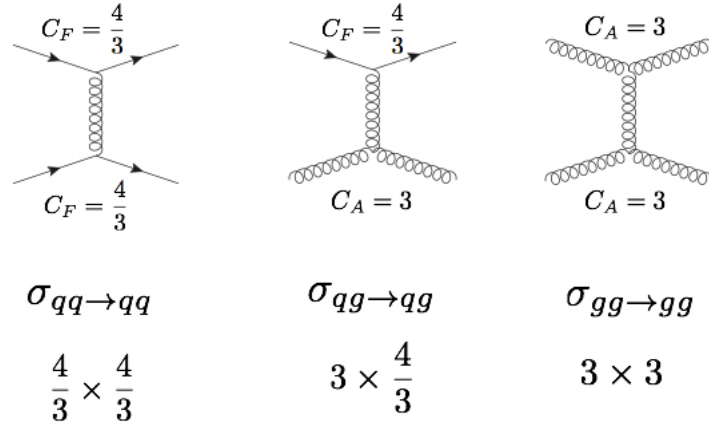
**Fig. 38:** Main (lowest order) partonic subprocesses contributing to the jets production at hadron collision and their corresponding colour structure.

In Fig.(39), we can see the comparison among theory and experiment. Note the necessity to include the gluon pdf to get agreement.
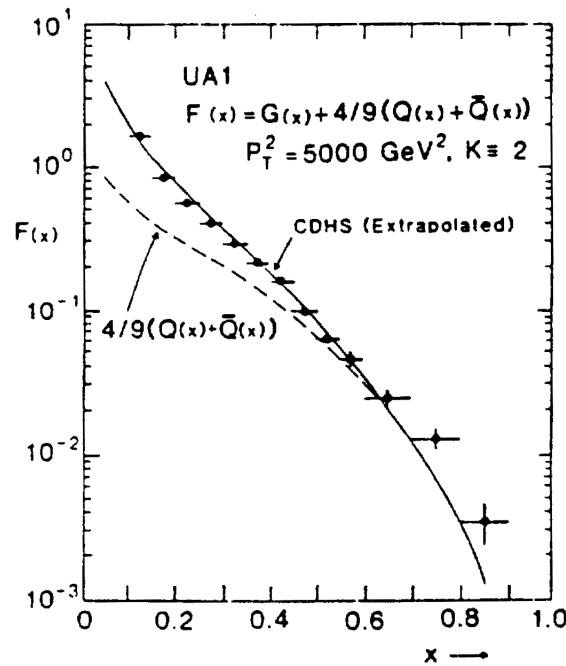


**Fig. 39:** Measurement of the jet generalized structure fuction $F(x)$ as function of the momentum fraction $x$ by the UA91 experiment and comparison with the theory prediction. The dashed line corresponds to the QCD excluding the gluon contribution. The inclusion of the gluon pdf (solid line) is required to have agreement with the data.

## 3.3   Jets algorithms

By the way, it is only when we go to higher orders in perturbation theory that we start getting more solid predictions for the normalization of the cross sections with a reduced dependence from the un-avoidable but unphysical scales (renormalization and factorization scales). Such predictions allow to perform more reliable tests of QCD comparing theory and experimental measurements of jet differen-

tial distributions [22]. However, not all the measurements that can be done at an hadron collider can be compared to QCD perturbation theory. QCD can provide perturbative predictions only for InfraRed and Collinear safe variables. For jet physics, as we will see in a moment, this means that care must be given for theoretically viable jet definitions. At very high energies like at the Tevatron or the LHC, for the large part of the interesting events the situation is not as clear as the one in Fig.(36). Looking at the $\eta, \phi$ plane, one often has a situation more like the one in Fig.(40). with the energies of the particles

$$R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$$
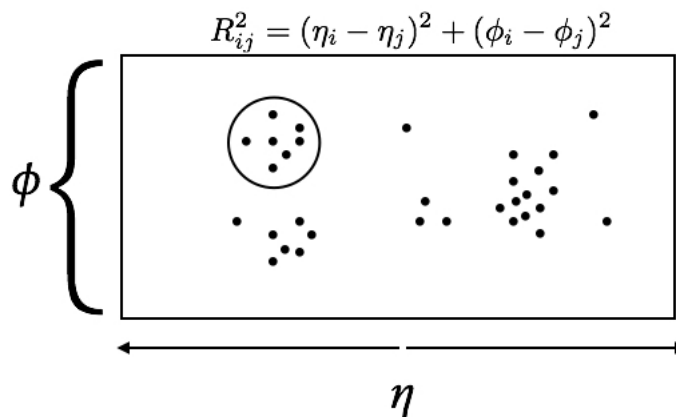


**Fig. 40:** Distribution of the reconstructed particles in the $\eta, \phi$ plane for a typical collision event at the Tevatron or the LHC. There are sparse groups of particles and the assignment particle-jets is not unambiguously clear.

(or small groups of particles in a calorimeter cell) that are very different from each other. The basic idea is always that, in very high energy collisions, jets can be reconducted to the original partons that came out from the hard scattering. Whenever shower and (non perturbative) hadronization corrections are expected to be small, the same algorithm can be logically applied to both partonic and particle level data. It could happen that for experimentally well motivated jet definitions, through which it is possible to make very good measurements, it is not possible to make higher order predictions of jet observables to compare with, without introducing sensitivity to long distance phenomena. This is indeed the case for several cone algorithms. A cone algorithm is one in which a cone is drawn around the jet axis and the particles contained in that cone are collected and assigned to the jet. The crucial point is how to fix the jet axis. Consider for example for the so-called highest-$E_T$-seed algorithm. Once you fix the cone radius $R$, this works as follows: 1. select the highest $E_T$ particle as the jet axis, 2. build the cone and remove the particles in it, 3. restart from 1. One can perform measurements of the jet activity in hadron-hadron collisions using this algorithm, but it is not hard to see that the jet number is a Collinear unsafe observable and so the perturbative prediction will get at each order logarithmic enhanced corrections that depend on the low energy dynamic. The problem can be visualized thinking to a system of three partons with transverse energies such that $E_{T1} > E_{T2} > E_{T3}$ and $E_{T1} < 2\,E_{T2}$. Furthermore, assume that the distances are such that $d_{21}, d_{31} < R$. This configuration gives rise to a single jet formed with the three partons, see the left panel in Fig.(41). Now you can think that the virtual corrections, with the three partons in the final state, will again fall into the same jet bin of the Born level process. When computing the real part, however, one has to consider the possibility of a collinear splitting that transforms parton 1 into two partons with lower energies. In this case, the algorithm will form two jets as shown in the right part of the Fig.(41). This is a phenomenon that a "good" jet algorithm should avoid in order to allow for theory-experiment comparisons beyond the leading order accuracy. One can modify the algorithm as follows: 1. all particles above a certain threshold are seeds, 2. combine all the particles in a cone to define a new proto-jet axis with

$$\phi_c = \frac{\sum_{i\in c} E_{Ti}\phi_i}{\sum_{j\in c} E_{Tj}} \qquad\qquad \eta_c = \frac{\sum_{i\in c} E_{Ti}\eta_i}{\sum_{j\in c} E_{Tj}} \qquad\qquad (124)$$
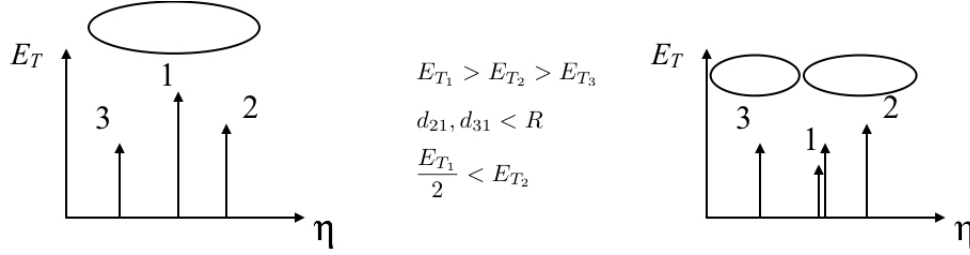
**Fig. 41:** A three partons configuration which is identified as a mono-jet event according to jet highest-$E_T$-seed algorithm (left panel). Collinear real emission from the parton 1 may lead to the production of two partons with lower energies such that, now, the algorithm identifies two distinct jets (right panel). The two collinear particles are separated in the two jets, spoiling the cancellation of the collinear singularity among the real and the virtual contribution which produces a mono-jet configuration as the leading order one.

and check the particles in the new cone: if they are the same, then a proto-jet has been formed (stable cone), 3. if proto-jets have small (large) overlap split (merge) them in some way introducing some technical parameters. In this way, jets number is now Collinear safe but unfortunately it is an InfraRed unsafe observable, that is a soft emissions change the number of jets as can be argued inspecting Fig.(42). An alternative to this problem could be the insertion of extra seeds at the mid-point of the stable cones, but
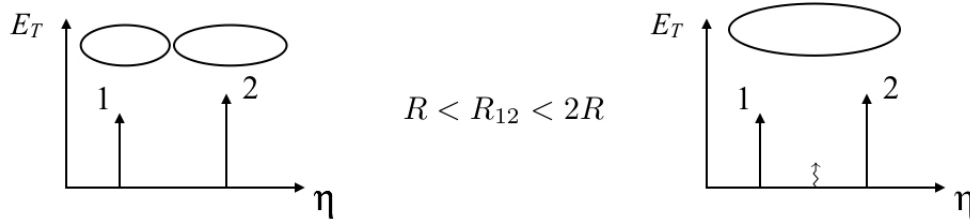


**Fig. 42:** First variant of the highest-$E_T$-seed algorithm described in the main text: the jets number is Collinear but non Infrared safe. Consider a leading order two-jets configuration as given in the left panel. Then, a soft real emission which occurs around the mid-point between the original two partons leads to the identification of a single jet (right panel). Indeed, there is now a large overlap between the two stable cones and they are merged into a single jet.

this is not a solution that turns out to be valid for higher order computations. A jet identification algorithm based on a cone definition that is fully InfraRed and Collinear safe is the seedless cone algorithm that works as follows: 1. consider a subset of particles, 2. combine momenta and search for a stable cone, 3. make it for all the subsets, 4. use split-merge as before. Nevertheless, it might become quite unpractical for very high multiplicity. Furthermore, it could happen that single particles with sufficiently high energy are not assigned to any jet (dark towers). Another solution is represented by the clustering algorithms that implement a sequential recombination along the line of the algorithms discussed for $e^+e^-$ annihilation into hadrons. For the case of hadron collisions, the clustering algorithm requires the definition of two distances, that among each pair of proto-jets, $d_{ij}$, and the distance of each of them from the beams, $d_{iB}$. The algorithm works as follows: 1. for each pair compute $d_{ij}$ and for each proto-jet compute $d_{iB}$, 2. if the smaller distance is one of the $d_{ij}$ combine their momenta to form a new proto-jet ($p_{ij} = p_i + p_j$) while, if the smalled is one of the $d_{iB}$, the proto-jet $i$ is promoted to a jet, 3. restart from point 1. This kind of algorithm is fully exhaustive and unambiguous (no need of split-merge procedures), it's also fast and does not generate dark towers. Introducing the quantities (caloremeter distances):

$$R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2 \tag{125}$$

the distances are defined in terms of an exponent $p$ and another real parameter $R$, that acts as the radius of the cone algorithm, as:

$$d_{iB} = (q_{T_i})^p \qquad d_{ij} = \text{Min}\left\{ (q_{T_i})^p, (q_{T_i})^p \right\} \frac{R_{ij}^2}{R^2}. \qquad (126)$$

The choice of the value for the parameter $p$ has an impact on the jet contours and the sensitivity to the detailed structure of the soft component of the event. For these reasons the preferred value (the standard) at the LHC experiments is $p = -1$ that define the anti-kt algorithm [23], other possible choices are the original kt algorithm [24] with $p = 1$ and the Cambridge/Aachen algorithm [25] with $p = 0$. The relevance of the second distance introduced, the one from the beams, appears when we analyse InfraRed and Collinear safety of the jet definition. Collinear safety is guaranteed by the fact that $R_{ij} \to 0$ in the collinear limit, while for InfraRed safety the point is that soft real radiation is either assigned to a jet or forms a jet by itself. In the latter case, the "soft" jet, being below any measurable threshold, is combined with the virtual counterpart, without spoiling the cancellation between real and virtual contributions. We conclude this section by observing, once again also for hadron collisions, that even if we are able to build InfraRed and Collinear safe algorithms for jet identification, the number of jets we identify and how they are distributed depends upon how we search for them, i.e. the algorithm we use (and its parameters). Anyway, this does not prevent us to test the perturbative regime of QCD probing the theory in high energy scattering processes among the elementary constituents.

### 3.4 Jets phenomenology

Let's start by considering the inclusive jet cross section at the LHC. In perturbation theory it is computed through the formula:

$$\frac{d^2\sigma}{dy\,dp_\perp^2} = \sum_{i,j,k} \int dx_1 dx_2 f_i(x_1, \mu) f_j(x_2, \mu) \frac{d^2\hat{\sigma}_{ij \to kX}}{dy\,dp_\perp^2}. \qquad (127)$$

While integrating each matrix element (and each subtraction counterterm needed order by order in perturbation theory) over the proper phase space, the event kinematics is passed through a jet reconstruction algorithm, i.e. there is a convolution with one of the jet finder functions described in the previous section, then the experimental constraints of the measurements (cuts) are applied. The error associated to the perturbative part of the prediction can be estimated from the variations of the unphysical scales plus the error related to pdf's and strong coupling determination. One then can estimate non perturbative corrections using Monte Carlo generators, first evaluating the ratio (bin-by-bin) of the Monte Carlo cross sections with and without hadronisation, and then multiplying by this ratio the NLO parton-level cross sections. The uncertainty related to this procedure is estimated as the maximum spread of the correction factors obtained using different Monte Carlo programs. In Fig.(43), the excellent agreement is shown among theory [26] and experiment [27] for the doubly differential distribution of the inclusive jet cross section. Left and right panels refer to different choices for the $R$ parameter characterizing the size of the jet in the anti-kt algorithm, and also within this variation the agreement is excellent over many orders of magnitude! Recently, the Next-to-Next to leading order computation for two jet production at hadron colliders has been performed in the leading color approximation [28] that amounts to selectively compute all the terms that are enhanced by colour factors and for this reason are supposed to dominate. In Fig.(44) it is shown the excellent agreement among theory and experiment over almost seven orders of magnitude for the invariant mass distribution of the dijet pair in different rapidity bins, at the LHC.
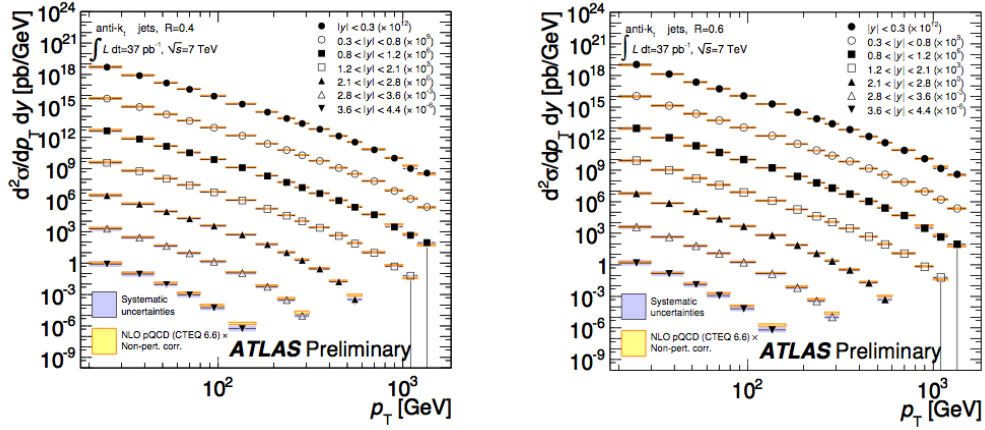
**Fig. 43:** Data-theory comparison of the double-differential inclusive jets cross section (in $p_T, y$) by the ATLAS collaboration. The jets are reconstructed according to the anti-kt algorithm setting $R = 0.4$ (left panel) and $R = 0.6$ (right panel).
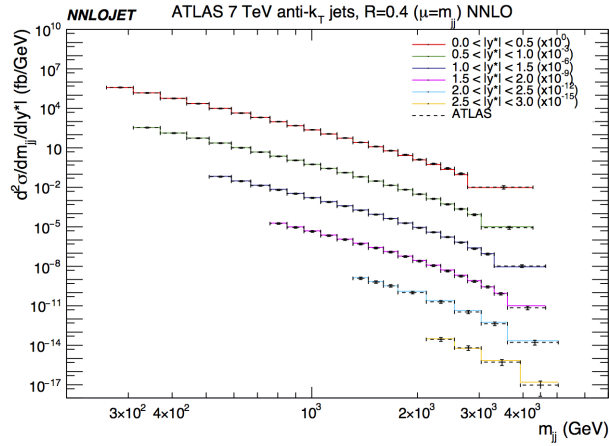


**Fig. 44:** Comparison between data and the state of the art NNLO prediction (in the leading colour approximation, using the NNLOJET code) for the double-differential inclusive jet cross section (in $m_{jj}, y$) by the ATLAS collaboration. The agreement is excellent over seven orders of magnitude.

## 4 QCD at higher and all orders

The purpose of this section is the introduction to those methods to treat QCD radiation in the perturbative regime that allow to reach the highest accuracy in testing this theory at colliders. Very few technical details are provided. The emphasis is more on the ideas underlying these methods, with the aim to give an orientation on what can be expected from their use and what cannot.

### 4.1 Higher order corrections for LHC processes

The exact computation of a hard scattering cross section including all the contributions up to a given order in the coupling constant goes under the name of fixed order computation. Although the problem has been solved at NLO, the complications that show up going at higher orders is formidable and we are still quite far from automation at NNLO. This is the frontier from the computational point of view. These computations are revealing new mathematical structures and a big effort is currently required also to the

mathematicians, so that the situation is evolving fast. Note also that, in our knowledge, there is no higher order computation performed so far that has not become relevant for a comparison among theory and experiment. Indeed, we can say that every time a new higher order computation has been performed a new phenomenological challenge for the Standard Model is ready. Recently, an N3LO QCD calculation has been also completed. It is the cross section for the Higgs boson production in gluon fusion at the LHC [29, 30], see Fig.(45) for representative diagrams up to the NLO. In the Standard Model there is
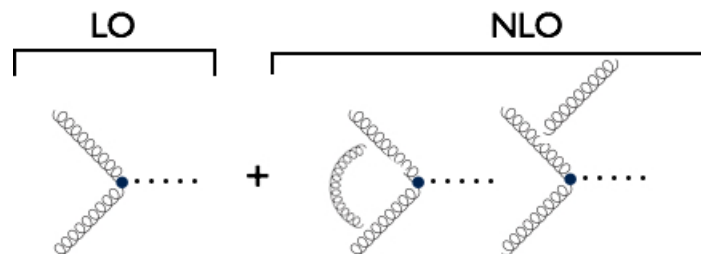


**Fig. 45:** Selection of diagrams contributing to the production of a Higgs boson in the gluon fusion channel up to NLO in the strong coupling. The interaction between the gluons and the Higgs is given by the effective vertex obtained pinching the dominant one-loop diagram given by the top quark triangle (large $m_t$ approximation).

not a vertex coupling directly the Higgs boson to gluons. This coupling in the SM proceeds, with very good approximation, via a loop of top quarks. From such top loop diagrams, where both gluons and a Higgs boson are attached, it is possible to deduce an effective theory including the $Hgg$ vertex shown in Fig.(45). Going to NNLO, one has to include three new cathegories of diagrams: the ones with two radiated partons (double real), the ones with one radiated and one virtual (real-virtual one loop diagrams) and the ones with two virtual diagrams (double virtual or two loops). Needless to say that going to N3LO the categories increase and they range from three radiated real partons up to three virtual partons. For the time being this computation in [29, 30] has been performed at the inclusive level only, that is only the total cross section has been predicted at N3LO. The experiments at CERN on the other hand measure the cross section (even differentially) but only in a certain fiducial volume. Therefore, the comparison among theory and experiment requires an extrapolation of the measured result over the unexplored region, that is done with Monte Carlo tools. In Fig.(46) [29], the result of the N3LO computation as a function of the common choices for the renormalization and factorization scales is shown. Here one can appreciate the flattening of the scale dependence going from one order to the next. We usually say that scale variations provide the uncertainty from the lack of knowledge of missing higher orders. Nevertheless, there is not a unique prescription on how to do it. This is a point debated since long. Given that these scales are fictitious and the full theory does not depend on them, what is obvious is that the scale dependence of a prediction at a certain order must exactly match the scale dependence of the collection of all the higher order contributions up to infinity with the opposite sign. For this reason, one can reasonably think that scale variations provide a taste of the size of the missing higher orders. At the differential level, NNLO computations in hadronic collisions are available for some time for the Higgs boson and single gauge boson production and more recently also for every pair of bosons and for top anti-top production. These computations allow to make very precise tests of the Standard Model at the LHC and at the same time, of course, set the scene to disentangle any possible new physics signal at this machine.

## 4.2 Resummation

At the differential level, fixed-order predictions might be affected by logarithmic enhancements which show up when a particular restricted region of the space space is considered. Such logarithmic enhancements can make meaningless the fixed order prediction in that phase space region so that apparently we assist to a break down of the perturbation theory. For a reliable comparison with experiments, whenever
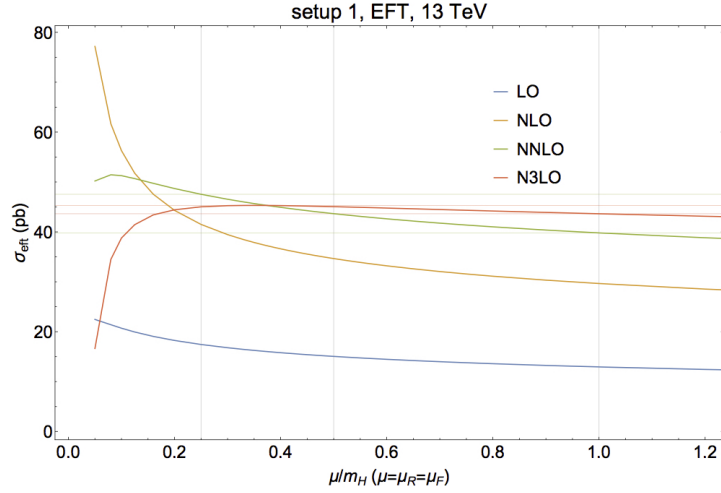
**Fig. 46:** Comparison of predictions with increasing accuracy up to N3LO for the Higgs production cross section in the gluon fusion channel as a function of the scale $\mu = \mu_F = \mu_R$. The inclusion of higher order corrections not only improves the estimate of the total rate but nicely reduces the scale variation. At N3LO, the dependence is almost flat in a wide range of scales, giving a solid confirmation of the reliability of this result.

possible, a consistent procedure of "resummation" of the logarithmic enhancements to all orders is required. We have already encountered the UV $\log$ and the Renormalisation Group Equation that allows to effectively resum this $\log$ producing the observed running coupling. Indeed, the resummed expression for the running coupling has to be used in order to connect, for example, the phenomenon of hadronic decay of the $\tau$ lepton and the hadronic decay of the $Z$ boson. Starting from $\alpha_s(m_\tau) \sim 0.36$ the running nicely predicts:

$$\alpha_s(M_Z) = \frac{\alpha_s(m_\tau)}{1 + b_0\,\alpha_s(m_\tau)\log\frac{m_Z^2}{m_\tau^2}} \sim 0.12 \tag{128}$$

which is in good agreement with measurements. The reduction of $\alpha_s$ by factor of 3 implies, roughly speaking, that:

$$b_0\,\alpha_s(m_\tau)\log\frac{m_Z^2}{m_\tau^2} \sim 2 \tag{129}$$

and proves that:

$$\alpha_s(M_Z) \neq \alpha_s(M_\tau)\left(1 - b_0\,\alpha_s(m_\tau)\log\frac{m_Z^2}{m_\tau^2}\right) < 0. \tag{130}$$

The expansion of the running coupling to the first order, given by the expression in rhs in the above formula, is even negative, a completely non sense result! So that the resummation of the $\log$ is mandatory. Then we have encoutered the collinear $\log$ in collisions with an hadron in the initial state that is effectively resummed by the Altarelli-Parisi equations. The common feature in the above examples is that they are both multiscale problems, with two (or more) scales, far apart each other. In the first case, they are given by $m_\tau$ and $M_Z$, while in the second case the scale of the hard process $Q$ and the scale of the strong coupling regime of QCD. In these situations, multiple emission processes connecting the two scales spoil the convergence of the perturbative expansion and cannot be neglected. From another perspective, large logarithmic contributions can remain in the computation, despite the fact that more inclusive quantities (as the total accepted signal) are still finite, if the experimental cuts produce an unbalanced cancellation of real and virtual contributions, with a suppression or emphasis of the real contribution. Think for example once again to the jet cross section in $e^+e^-$ annihilation into hadrons at NLO. Inclusively, the

real and the virtual contributions together give a finite result, but if we restrict too much the value of the jet resolution parameter $y_{cut}$, the three jet cross section fraction increases more and more becoming grater then 1, while at the same time the two jet fraction becomes negative! This situation is of course untenable and it signals the unbalance we meant above. Fixed order perturbation theory has to be used only far from these kinematical regions. But, what to do if experimentally one can perform good measurements also assuming such small values of the resolution parameter? This problem, on the theory side, can happen for every observable which is sensitive to soft and/or collinear radiation and the situation does not improve that much including only a finite number of higher orders contributions. Consider for example the computation of the transverse momentum distribution of a vector boson produced in hadron-hadron collisions. Illustrative diagrams of this process are shown in Fig.(47).

The scale of the process is of course the mass of the vector boson ($M$). The schematic behaviour of the perturbative expansion for the $p_T$ distribution of the boson is shown in Fig.(48). At leading order ($\alpha_S^0$), this distribution is just a $\delta$ function selecting zero $p_T$. Experimentally the measurement will have a certain resolution and the result will be given as an histogram as for the theory prediction. At Next-to-Leading order ($\alpha_S$), the logarithm ($L$) of the ratio $M^2/p_T^2$ shows up with a positive diverging behaviour in the low $p_T$ limit. Note as well the presence of a characteristic feature of the fixed-order computation. Since the area under the histogram represents the total cross section (which is a Collinear and InfraRed safe observable), the first bin is negative and if we restrict the bin it tends to become divergent to compensate the divergent positive behaviour of the integral of the rest of the distribution. At the next order ($\alpha_S^2$) the distribution tend to minus infinity approaching low $p_T$ and up to two more powers of the same $\log$ are present. Once again, the first bin is large and positive and compensates to produce a finite total rate. One can show that at order $n$ all powers $m$ of the log with $0 < m < 2n - 1$ manifest. This situation repeats also for other measurements that are very well defined experimentally like the small mass limit of a jet in $e^+e^-$ annihilation, the small jet radius in hadron collisions or the large transverse momentum distribution of massive quarks produced at high energy and so on. Predictions for all these observables need resummation of large logs ($L$) in the kinematical region where indeed $\alpha_s L^2 \sim 1$. The contributions have a well defined structure and can be schematically organized as shown in Fig.(49).

We do not resum the whole series, just classes of contributions. The sum of all the terms in the first column in Fig.(49) is called the Leading Log (LL) resummation, the second column the Next-to-Leading Log (NLL) resummation and so on. This time the next term in the series, the next column, will provide a relatively lower contribution because it will be suppressed by one power of $L$ with respect to the previous one. Note that resummed results can be combined with the fixed order "horizontal" contributions via a so called matching procedure that removes double counting of the logarithmic contributions. However, whether the series converges or not (and if it has an exponential form) depends on the quantity being measured. For examples, the $\log(y_{cut})$ contribution in $e^+e^-$ annihilation does not exponentiate for the JADE algorithm but it does for the kT algorithm! In general, for the computation of the coefficients of the series, it is not needed to compute the whole correction at a given order, it is enough to select and compute those contributions that catch the soft and collinear behaviour up to the desired order. Furthermore, it is not needed to compute explicitly all the coefficient in a certain column, from a subset of coefficients it is possible to deduce the all order expansion. Resummation for basic variables measurable at the LHC, such as for example the transverse momentum of the Higgs boson and vector gauge bosons, have been
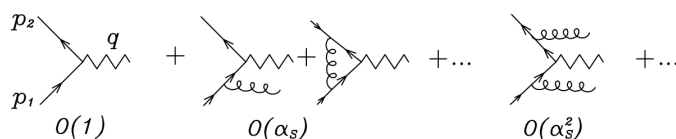


**Fig. 47:** Illustrative diagrams contributing to the production of a vector boson in hadron-hadron collisions.
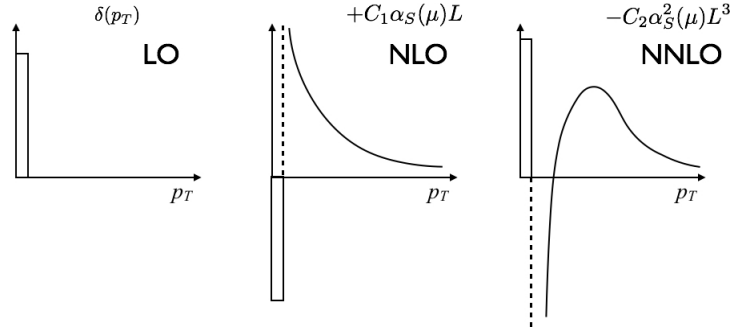
**Fig. 48:** Fixed-order predictions including an increasing content of radiative corrections for the differential $p_T$ distribution of the vector boson. A logarithm $L = \log M^2/p_T^2$ shows up at NLO, its square at NNLO and so on. The predicted behavior in the small-$p_T$ limit is divergent and so completely unphysical.



**Fig. 49:** Structure of the logarithms which occurs at each order in perturbation theory.

performed recently including up to N³LL+NNLO corrections [31]. The result for the Higgs boson is shown in Fig.(50).

While the fixed order results have no physical meaning in the small $p_T$ region, the resummed one has an acceptable behaviour in agreement with the experimental results (not shown in the figure) that present a maximum and then goes down at lower $p_T$. Note also the strong reduction of the scale variation that is a typical benefit of the procedure for that part of distributions that are pathological for the fixed order computation.

## 4.3 Parton shower

As we have seen in the previous section, the resummation of higher order logarithmic enhancements contributing to single variable distributions has reached an high level of accuracy for the cases studied. There is another possible strategy which allows to resum only the leading soft and collinear logarithmic enhancements but, on the other hand, is fully differential. This means that it is not limited to the single inclusive distribution and can be used for all the observables for a given process. Starting from a hard 2 to 2 process, it is possible to build an approximation for all the diagrams in which subsequent multiple emissions have distributed the available energy to a shower of partons. To give an idea about how this works we will consider again $e^+e^-$ annihilation into hadrons. In QCD we aim at a description as given in Fig.(51).

We start from the $q\bar{q}$ pair produced at high energy. It happens at a certain fixed starting hard scale $Q$.
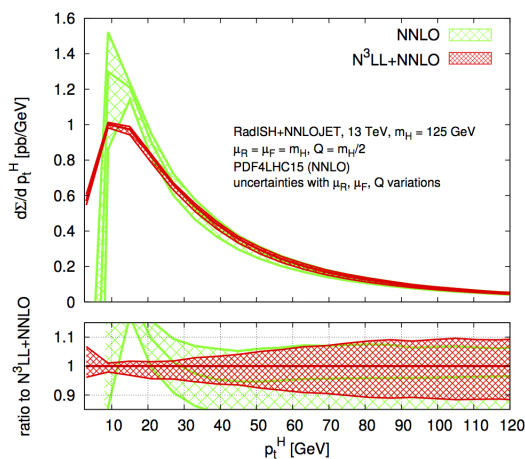
**Fig. 50:** Comparison between the NNLO fixed-order prediction and the N³LL+NNLO one for the differential $p_T$ distribution of the Higgs boson in gluon fusion channel. As expected, at high $p_T$ there is an excellent agreement, while in the small $p_T$ region, the resummed prediction gives a reliable and physical result, while the fixed order prediction is divergent. Moreover, resummation nicely reduces the scale variation, especially in the small $p_T$ limit.
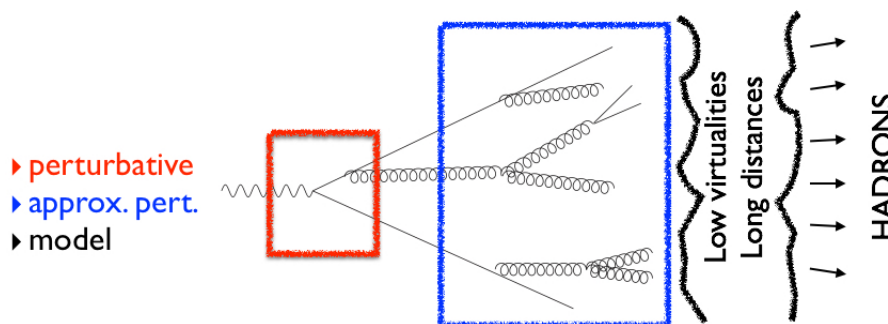


**Fig. 51:** Sketch of the QCD evolution from the hard scattering to the final state hadrons in the simple case of $e^+e^-$ collisions. The main behaviour of the shower of partons can be captured by an approximate perturbative QCD description, which is implemented in the so-called Parton Shower programs.

Then we look for an iterative formula to approximately describe the sequence of parton splittings starting from the real contribution. What we are going to build is an effective description of the shower process in terms of splitting functions and propagators and not the full matrix element squared for every possible history, which is a task still beyond the possibilities of any available computer. Nevertheless, it can be shown that such construction catches the main features of the QCD evolution of the external partons that took part to a given hard process, from the hard scale of the process down to scales a bit above $1\,\mathrm{GeV}$. To start with, we know that collinear emission from each leg is enhanced and we can use Eq.(47) and Eq.(48) to describe the emission probability as:

$$d\sigma_R \sim \sigma_0 \sum_i \frac{d\theta^2}{\theta^2}\, dx\, P_{ij}(x). \tag{131}$$

Of course, this probability still diverges. We observe that to describe the collinear limit we could have used also other variables that are good as well. Indeed, we get the same limiting behaviour if the new variable is proportional to $\theta^2$. Consider for example the invariant mass of the internal propagator $q^2$ in a splitting, whose expression in the collinear limit is $q^2 = x(1-x)\theta^2 E^2$, or the transverse momentum associated to the emission, $k_\perp^2 = x^2(1-x)^2\theta^2 E^2$. We would get identical results for the collinear limit,

but different extrapolations away from it. Let us consider the virtuality of the internal line so that we have:

$$d\sigma_R \sim \sigma_0 \sum_i \frac{dq^2}{q^2} \, dx \, P_{ij}(x). \tag{132}$$

Still, of course, the probability for a given emission path, which is obtained by reiterating (i.e. exponentiating) this formula, can well exceed 1! While we want that the sum of the probabilities of all shower configurations must yield 1. Indeed, approaching the strict collinear limit, the fixed order theory prediction is challenged and the enhancement, phenomenologically, are not really there as such. In turn, we need to damp this emission probability with a factor that has to restore unitarity keeping into account the missing virtual corrections and, at the same time, resumming the logarithmic enhancement. To address this point, while discussing the basic idea, we make a further simplification by assuming that we have just one kind of parton (the gluon for example) and so one kind of splitting $P_{gg}$. At some point of the shower the number of gluons will certainly prevail on the number of quarks, nevertheless our simplifying assumption will obviously not be a good approximation for quark initiated processes, and it has to be considered just as an illustrative example. The reader interested in the details of the most general case can go through excellent reviews of the subject [32, 33]. To restore unitarity we normalise the emission probability involving an internal propagator with virtuality $q^2$ given in Eq.(132) multiplying by the inclusive probability of no emission of radiation implying and internal propagator virtuality from the maximum available virtuality $Q^2$ down to the virtuality $q^2$:

$$d\sigma_{first\ emission} = P_{gg}(z)dz\frac{dq^2}{q^2}\Delta_g(Q^2, q^2). \tag{133}$$

In the formula above, we do not include the leading order cross section $\sigma_0$ because we are now considering emission from a single final state parton (gluon in our simplified exercise), so that we are building the iterative factor that describes the emission probability from each leg in a Markov chain like process. The factor $\Delta_g$ is called a Sudakov form factor. Of course, this "no emission" probability has to be related to the emission probability, it has to be a function of it. Given that for smaller and smaller $q^2$ the emission probability diverges, in this limit the inclusive "no emission" probability has to go to zero rapidly, so regularising the divergence. Now, let us compute $\Delta$ (here comes the resummation!). In the dominant collinear limit, the probability of emission implying an internal propagator virtuality among $q^2$ and $q^2 + dq^2$ is given by:

$$dW_g = \frac{dq^2}{q^2} \int_{z_{min}}^{z_{max}} dz \, P_{gg}(z) \tag{134}$$

so that, for such an infinitesimal range of virtuality $dq^2$ the "no emission" probability is given by the unitary condition:

$$dP(\text{no emission}) = 1 - dW_g = 1 - \frac{dq^2}{q^2} \int_{z_{min}}^{z_{max}} dz \, P_{gg}(z). \tag{135}$$

Now, to get the (resummed) "no emission" probability over the finite range from $q^2$ up to $Q^2$, we do not integrate this expression, but instead we take the product over "all" the infinitesimal paths, which we consider as independent events. In doing so, the probability of "no multiple emission" of every number of gluons turns out to be:

$$\Delta_g(Q^2, q^2) = \prod dP(\text{no emission}) = \exp\left\{ -\int_{q^2}^{Q^2} \frac{dk^2}{k^2} \int_{z_{min}}^{z_{max}} dx \, P_{gg}(z) \right\}. \tag{136}$$

Note that in the limit of small virtualities we know that all order emissions are relevant and that factorization holds. This product has an exponential form and produces the damping behaviour we were looking for. Indeed, it damps the emission probability at low virtualities and it does not count much at
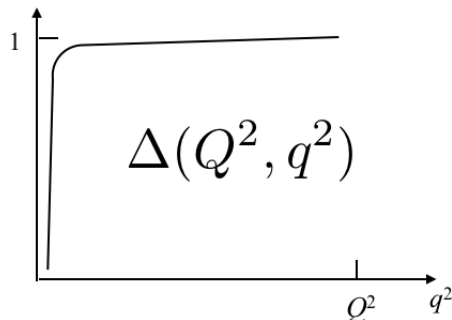
**Fig. 52:** Characteristic behaviour of the Sudakov form factor or "no emission" probability. It rapidly vanishes in the $q \to 0$ limit, while saturates to 1 at high virtualities.

large virtuality. A typical form is shown in Fig.(52). In our simplified QCD model, the parton shower is then produced iteratively generating radiation according to the probability distribution:

$$d\sigma_{first\ emission} = P_{ji}(z)dz\frac{dq^2}{q^2}\exp\left\{-\int_{q^2}^{Q^2}\frac{dk^2}{k^2}\int_{z_{min}}^{z_{max}}dzP_{ji}(z)\right\}. \tag{137}$$

At each iteration of this formula one updates the upper $Q^2$ scale to the extracted $q^2$ of the previous step, until the virtuality of the internal line reaches the cut-off scale. We stress that by construction unitarity is restored, so that the event simulation will not alter the value of the original cross section at the starting point ($\sigma_0$ at the origin). A number of observations have to be done. In real life event generators, the qualitative reasoning to describe the iterative procedure implemented in a parton shower given above are supplemented by a large number of technicalities we have not discussed. They go from the implementation of the momentum reshuffling after each emission to the management of multiple splitting processes. Another complication is related to the treatment of hadrons in the initial state. In that case one has to take into account the structure of the colliding hadrons, but the shower is worked out with the same logic (in that case one speaks of backword evolution and space-like shower). Furthermore, for a parton shower implementation there is also a number of choices to do, like: the evolution variable (that we have set to the the virtuality of the internal propagator), the scale to be used in the strong coupling, the cut-off scale. As for the evolution variable, it has a relevant impact. Indeed, note that the iterative multiple emission description obtained starting from the collinear approximation discussed above cannot represent adequately multiple soft emissions. It is true that damping the virtuality of the internal propagator or the tranverse momemtum of the emission also the soft radiation goes out, but we also know that the divergent behaviour of soft radiation has a universal different factorized form. From direct computation of QCD amplitudes, it turns out that for soft radiation there is a coherent effect such that, if one chooses as evolution variable the angle or the transverse momentum, the procedure automatically catches also the leading logarithmic enhancement associated to the soft radiation. This is a very good news because, if also soft radiation has to be more and more collinear, we can imagine that, although in principle it has a long range, in QCD it stays relatively close, so that the hadronization can be considered and modelled as a local process. An important application of a parton shower at the LHC is the study of the jet shape. Such a study requires of course the resummation of large soft and collinear logarithmic enhancements. A relevant variable in this case is the fraction of the jet transverse momentum contained within a ring of tickness $\Delta r$ around the jet core, $\rho(r)$ (averaged over an ensemble of jets), with:

$$r_i = \sqrt{(\Delta_{i,jet}y)^2 + (\Delta_{i,jet}\phi)^2} \tag{138}$$

$$r_a = r - \Delta r/2 \qquad r_b = r + \Delta r/2 \tag{139}$$

$$\rho \;=\; \frac{1}{N}\sum_{jets}\frac{1}{\Delta r}\frac{\sum_{r_a<r_i<r_b}p_{Ti}}{\sum_{r_i\leq R}p_{Ti}}. \tag{140}$$

In Fig.(53), we can see the general agreement among the measurement of $\rho(r)$ distribution and parton shower Monte Carlo programs. Another intersting variable is the fraction of the jet transverse momentum
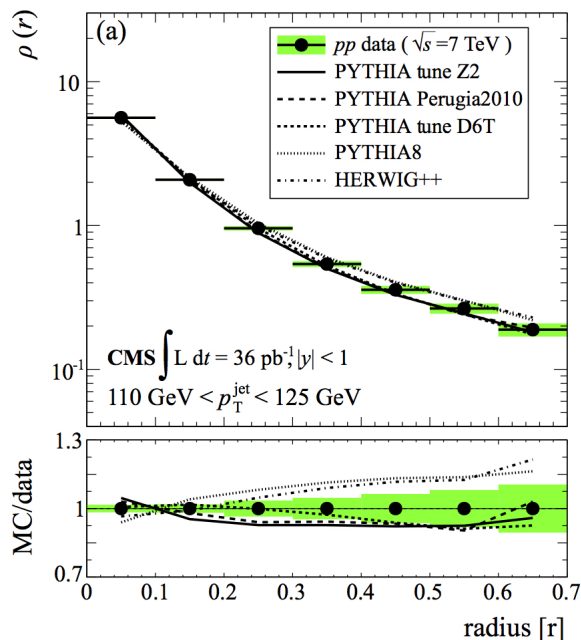


**Fig. 53:** MC/data comparison for the jet shape variable $\rho$ representing the fraction of the jet transverse momentum in a ring of tickness $\Delta r = 0.1$, measured by the CMS collaboration.

contained within a circle of radius $r$ around the jet core, $\Psi(r)$. In Fig.(54) the ATLAS measurement for $1 - \Psi(0.3)$ is shown. In this case, we see that in principle this variable could be used to distinguish jets
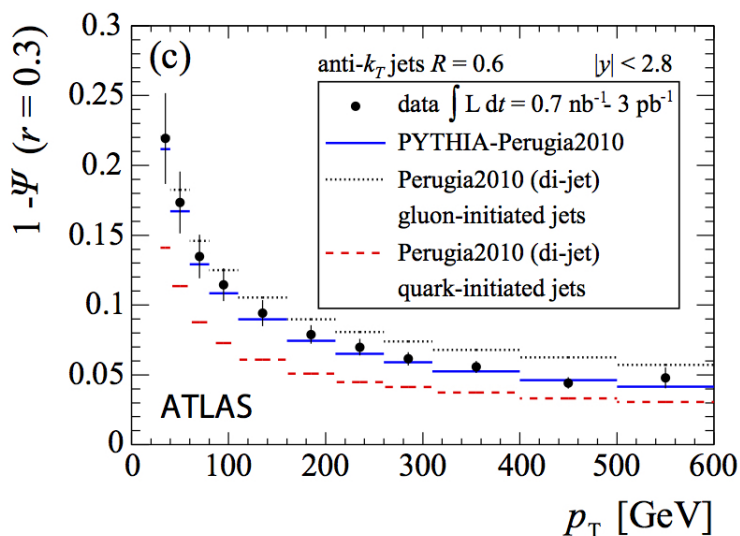


**Fig. 54:** MC/data comparison for the jet shape variable $1 - \Psi(r)$ with $\Psi(r)$ representing the fraction of the jet transverse momentum in a circle of radius $r$ (for the tipical value $r = 0.3$), measured by the ATLAS collaboration.

initiated by a gluon or a quark.

## 4.4 Merging, matching and both

Shower Monte Carlo programs effectively handle and resum the multiple collinear and soft QCD emissions and they provide the full event simulation, including the conversion from partons to hadrons. For this reason they are an essential tool to study the sensitivity of the experiments in high energy physics. On the other hand, they are able to generate cross sections for the requested hard process that are correct only at leading order. Furthermore, the soft and collinear approximations underlying the parton shower (PS) evolution may fail to reproduce the full pattern of hard wide-angle emissions at higher orders. One would like to include in a parton shower simulation all relevant hard processes while taking advantage of available higher order corrections. We are indeed on this path. The first step forward has been the merging of multi-jet final states preserving LO+LL accuracy, done combining together all the LO matrix elements with up to $n$ partons in the final state. The problem is how to regularize the divergent cross sections in the soft and collinear limits of the corresponding matrix elements (for $x > 2$, $x$ being the number of partons in the finale state) and to generate hard multijet configurations in a way that is not very sensitive to the regularization prescription. Here we will not give any proofs, but we will just outline the basic idea about how it works in practice. To this aim, let us consider again, for the sake of simplicity, the case of $e^+e^-$ annihilation at a cm energy $Q$. There have been more then one proposal on how to reach this goal [34, 35], here we limit to show how the Catani, Krauss, Kuhn and Webber (CKKW) algorithm [34] works. The first step is the introduction of a merging scale, so: 1. choose a scale $Q_1$ which fixes the kT algorithm resolution parameter $y_{cut} = Q_1^2/Q^2$. This parameter acts as a cut-off regulator for the computation of the $m$-jet cross section using the exact tree level matrix element with $m$ partons in the final state. This means that any $m$-partons configuration that would be clustered into a number of jets lower than $m$ is given zero weight. Repeat the procedure for every $m$ from 2 to $n$ (typically $n = 5$ or 6), so that for each multiplicity $m$ and partonic subprocess $i$ (for each multiplicity there is finite number of them increasing with multiplicity) you now have a corresponding finite numerical value of the cross section $\sigma_{m,i}^{(0)}$, where the (0) apex refers to the fact that we are considering leading order cross sections. 2. Select a jet multiplicity and a partonic subprocess with probability:

$$P(m,i) = \frac{\sigma_{m,i}^{(0)}}{\sum_{kj}^{k=n} \sigma_{k,j}^{(0)}}.$$

(141)

3. Generate an event with probability given by the corresponding exact matrix element. 4. Reconstruct the PS probability of this event first using the kT algorithm backward, recombining partons until only 2 remain, and then building an event weight combining Sudakov form factors (as propagators) and splitting probabilities (as vertices) with the strong coupling evaluated at the scale of the branching process for each branch. 5. Accept or reject the event according to the combined PS weight. If the event is accepted, assign it a color configuration and start the shower from each leg of the $m$ partons event. The initial scale of the shower should be set to the scale value of the node at which that parton was created. Furthermore, during the shower one has to veto radiation with transverse momentum larger then $Q_1$. If the event is rejected go back to point 3. Using the steps listed above, it has been proved that the results do not depend on the cutoff scale $Q_1$. Of course, the algorithm needs modifications when there are hadrons in the initial state, but the logic is the same: the idea is again to regularize the matrix elements separating the phase space into hard (ME domain) and soft (PS domain) and then to reject or accept an event, and eventually start the shower, according to the PS weight of the event. In Fig.(55) and (56), we can appreciate the adequate description given by the shower Monte Carlo programs for both inclusive and differential multijet cross sections at the LHC. However, one has not to expect too much from such a PS description. Note, for example, that a 5% uncertainty on the strong coupling induces a 35% uncertainty on the 6 jets production rate, on top of which one has to consider scale variations (both renormalization
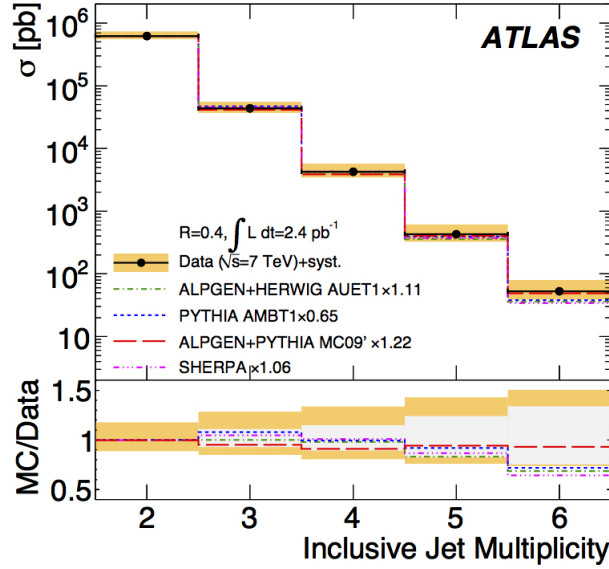
**Fig. 55:** Measurements of the total cross section for different inclusive jet multiplicities by the ATLAS collaboration compared to different MC generators.

and factorisation) and hadronization effects. So, in turn, the description of multijets final states with merging shower MC programs might be considered more as an exercise to tune those tools than a test for perturbative QCD.

The matching of NLO computation with PS LL resummation has been a big progress of the recent years. The first proposals have been the MC@NLO [36] and POWHEG [37] methods. Let us take for example the POWHEG method. The acronym stands for positive weight hardest emission generator. The technical problem to solve was how to avoid the double counting of radiation described by the NLO computation and the shower process. In particular, we want to keep the description of the hard and wide angle radiation as given by the NLO computation and exploit the resummation performed by the shower Monte Carlo program for the soft and collinear radiation. In this way, the full differential description offered by the PS programs will gain the normalization and a formal accuracy, for sufficiently inclusive observables, driven by the NLO calculation. The interested reader can look into the original reference for the proofs [37], here we just give a sketch of the construction in practice. We start considering the first emission in a parton shower approach:

$$d\sigma_{\text{first emission}} = \sigma_0 \, dz \frac{dk_\perp^2}{k_\perp^2} P_{ij}(z) \Delta_j(Q^2, q^2). \tag{142}$$

To have NLO accuracy on inclusive variables, we start generating the "underlying" Born kinematic configuration with probability given by the parton level NLO computation. So we promote $\sigma_0$ to:

$$\bar{B}(\Phi_B) = B(\Phi_B) + V(\Phi_B) + \int d\Phi_{rad}[R(\Phi_B, \Phi_{rad}) - C(\Phi_B, \Phi_{rad})] \tag{143}$$

In this formula, $B$ and $R$ are the Born and the Real matrix element squared, $C$ is the counter-term that makes finite the Real cross section and $V$ is the combination of the Virtual matrix element and the integrated counterterm, along the lines of the discussion in the first section. Actually, one should partition the Real matrix element into pieces that are each singular in just one kinematical radiation region and build as many of the last piece in the formula above as singular regions. Their simultaneous treatment is addressed with Monte Carlo techniques and it does not pose a problem. For ease of notation,
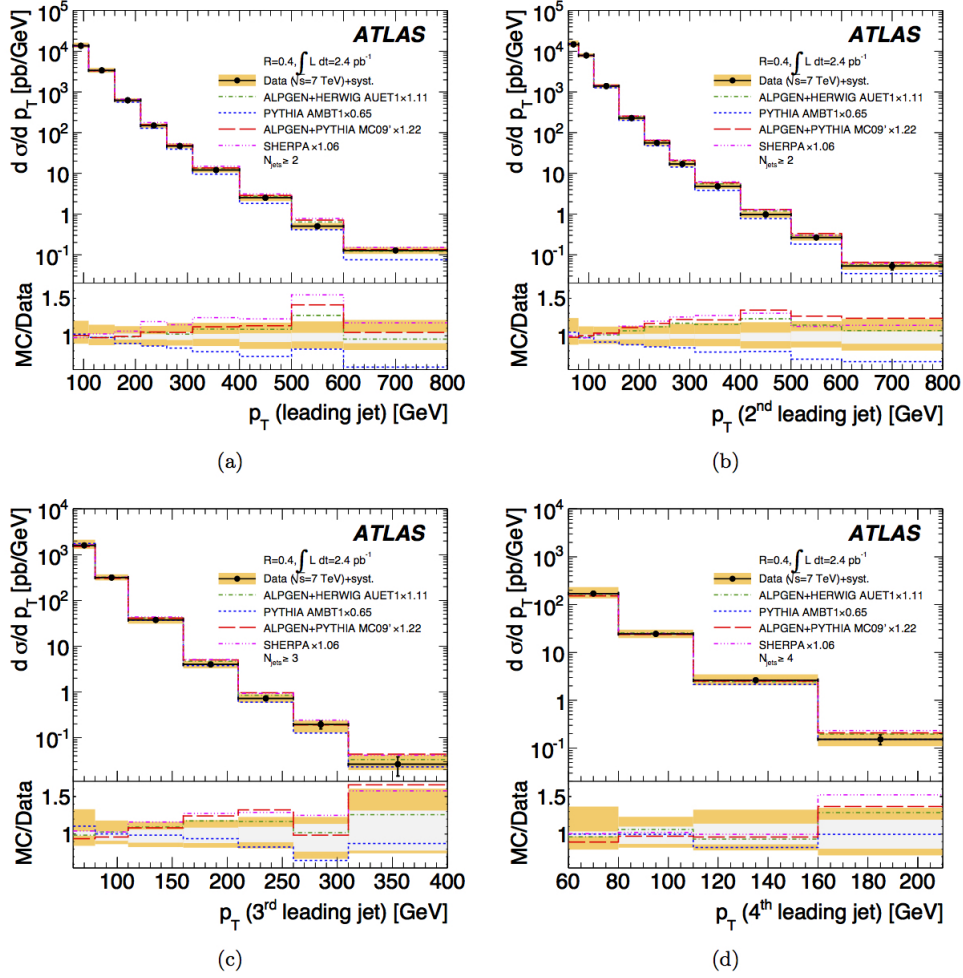
**Fig. 56:** Measurements of the $p_T$ distribution of the leading jet (a), of $2^{nd}$ leading jet (b), of $3^{rd}$ leading jet (c), of $4^{th}$ leading jet (d) by the ATLAS collaboration compared to different MC generators.

here we leave just one of them. Furthermore, with $\Phi$ we indicate the phase space variables for a Born configuration ($\Phi_B$), a Real one ($\Phi_B, \Phi_{rad}$) or just the radiation variables ($\Phi_{rad}$). To match the Leading Logarithmic resummation performed by the parton shower, while at the same time avoiding double counting, one has to generate the first radiation using the formula

$$d\sigma = \bar{B}(\Phi_B)d\Phi_B \, d\Phi_{rad} \frac{R(\Phi_B, \Phi_{rad})}{B(\Phi_B)} \exp\left\{ \int \frac{R(\Phi_B, \Phi'_{rad})}{B(\Phi_B)} \theta(k'_T - k_T) d\Phi'_{rad} \right\}, \qquad (144)$$

then assign a color configuration and shower the event vetoing emissions harder then the first one. We see that the damping Sudakov form factor is built upon the ratio of Real and Born matrix elements and not just by its divergent limit represented by the splitting functions. On the other hand the exponential resummation is certainly valid in such limits so that for the method to work properly one has to ensure that soft and collinear configurations are the only sources of enhancement for the ratio $R/B$. In the first method proposed for the NLO+PS matching (MC@NLO), the procedure is a bit different. In that case, the mapping used in the NLO calculation to connect a Real configuration to a Born one plus the radiation variables is chosen to be the one implemented in the Parton Shower program to describe the emission. Then the PS weight associated to the configuration of the real event (called Hard in the MC@NLO framework) is subtracted from the value computed with the exact Real matrix element and added back to the weight of the corresponding (mapped) Born configuration (called Soft). Note however that, being all the

contributions finite, it is always possible to built unweighted event samples before showering. The two procedures MC@NLO and POWHEG are then different in the way they treat subleading contributions and, in general, an estimate of the size of the uncertainty related to the matching procedure can be represented by the differences among the results found simulating the same process with the two tools. As an example of the results obtained at NLO+PS accuracy, we show the case of top anti-top production in proton proton collisions at $14\,\mathrm{TeV}$ [38]. As shown in Fig.(57), the MC@NLO generator reproduces very well the NLO distributions that are inclusive over extra QCD radiation as the rapidity and the transverse momentum of the top quark. In the case of variables that are sensitive to radiation attached to the Born
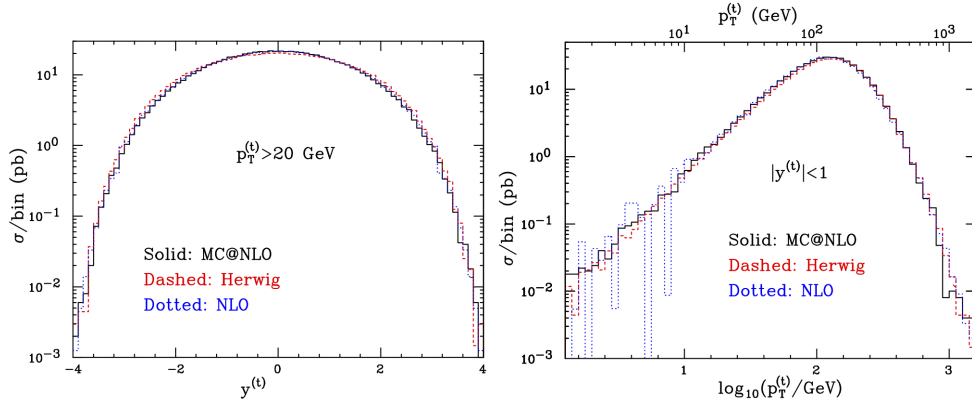


**Fig. 57:** The NLO+PS(MC@NLO) prediction for the single inclusive observables $y^{(t)}$ (left) and $p_T^{(t)}$ (right) (rapidity and transverse momentum of the top) in top-antitop production at LHC14 is compared with the PS(Herwig) and the NLO fixed-order predictions. Here, a very good agreement is found among all the prediction.

process, as the transverse momentum of the top anti-top pair in Fig.(58), we observe a complementary behaviour of the NLO and MC approaches regardless of the cuts on the rapidities and transverse momenta of the top quark. In the tail of the transverse momentum distribution of the top pairs, the NLO
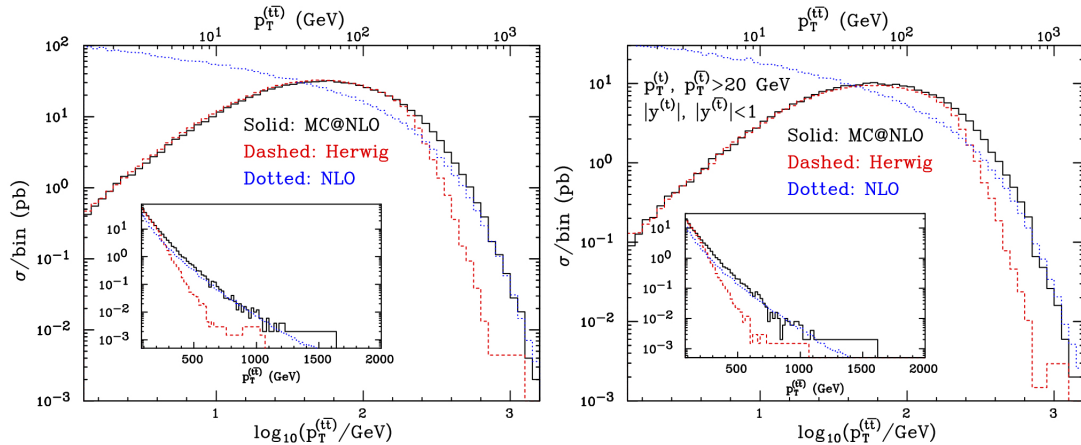


**Fig. 58:** The NLO+PS(MC@NLO) prediction for the transverse momentum of the top pair, $p_T^{(t\bar{t})}$, in top anti-top production at LHC14 with (right) and without (left) cuts on the top quarks, is compared with the PS(HERWIG) and the NLO fixed-order predictions. As expected, in the hight-$p_T$ tail there is agreement between the NLO+PS and the NLO, while the PS prediction fails in this $p_T$ region. On the other hand, at small $p_T$ values the NLO+PS shows up the characteristic LL resummed result as given by the PS prediction.

cross section is much larger than the MC one, simply because hard emissions are correctly treated only in the former. For small transverse momentum of the top pair, the difference between the two histograms

shows the effect of all-order resummation, clearly, no meaningful comparison between NLO and data can be attempted in this region. In turn, we assist to a smooth interpolation performed by MC@NLO. For their precision, flexibility and availability the NLO+PS generators are the standard at the LHC.

A step further has been the combination of the two improvements mentioned above, having the merging and the matching at the same time. There have been several proposals on how to merge NLO computations with different jets in the final state. In general, one can think that there is always the need of a merging scale as the $Q_1$ we introduced discussing the CKKW algorithm. Nevertheless, for the class of processes with only one parton in the final state at the leading order (as H/W/Z+1jet) which, of course, presents divergences when this radiation becomes unresolved, it is possible to have finite results (even including the NLO corrections) just by multiplying the event weights by an extra Sudakov form factor that suppresses the divergences. It has been shown that such a Sudakov form factor can be chosen in such a way that the formal accuracy of the inclusive generation is again at Next-to-Leading Order [39]. This method is called MiNLO, that stands for Multi-scale improved NLO computation, and has the nice feature that it does not require a merging scale. As an example of application, we show the result for the associated production of an Higgs and a Vector boson at the LHC (VH with V=W or Z) [40]. In Fig.(59), the total cross section for proton proton collisions at $8\,\mathrm{TeV}$ is shown for both VH (in black) and VH plus one jet (VHJ) with MINLO (in red). When the usual scale variation bands are considered, we see that there is a good overlap of the two results.
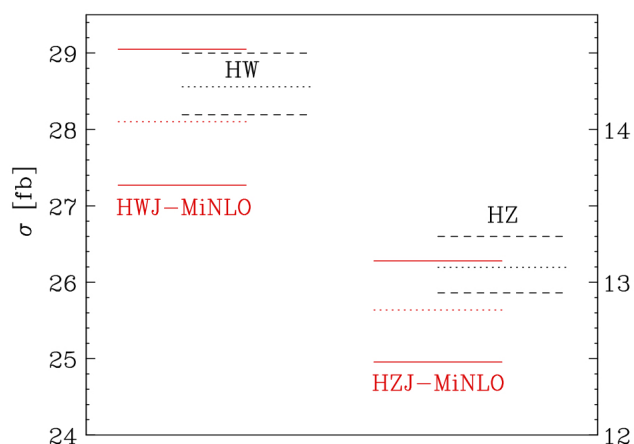


**Fig. 59:** Total cross section for associated Higgs production in proton-proton collision at LHC8. HV (V=W,Z) refers to the standard NLO computation, HVJ-MiNLO refers to the process with an extra jet treated with the MiNLO method.

In Fig.(60), we show the distribution of the transverse momentum of the WH system. For the high energy part of the distribution the two results obtained with WH and WHJ generators nicely overlap telling that the NLO correction provides only a small effect. In the low energy part of the distribution, instead, the two differ. The WHJ prediction, that contains the exact radiative correction for the first radiation, provides the most accurate result.

We conclude this section mentioning that reweighing a MiNLO NLO+PS generator for a process that has exactly one parton in the final state at the leading order (like H/W/Z+1jet, and in principle every process producing a colourless final state plus 1 jet) with a parton level NNLO computation for the fully exclusive production of such a colourless final state, one obtains an NNLO+PS matching! [41]. In Fig.(61), the Higgs boson transverse momentum distribution from gluon fusion at the LHC8 is shown.
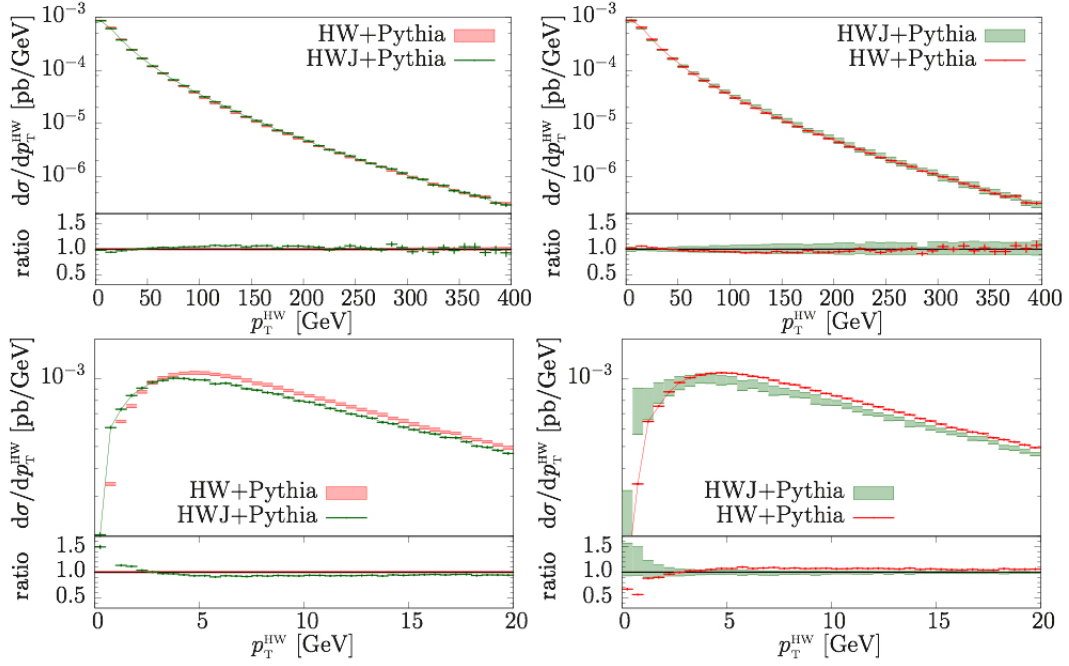
**Fig. 60:** Comparison between the HW and HWJ predictions for the $p_T$ distribution of the HW system. In both case, generated events have been showered with PYTHIA. In the hight-$p_T$ tail (upper panels), there is a good agreement while for small $p_T$s, the HWJ-MiNLO predicts a slightly softer spectrum.

We remind that the fixed order result for this distribution (obtained with the program HNNLO [42] also
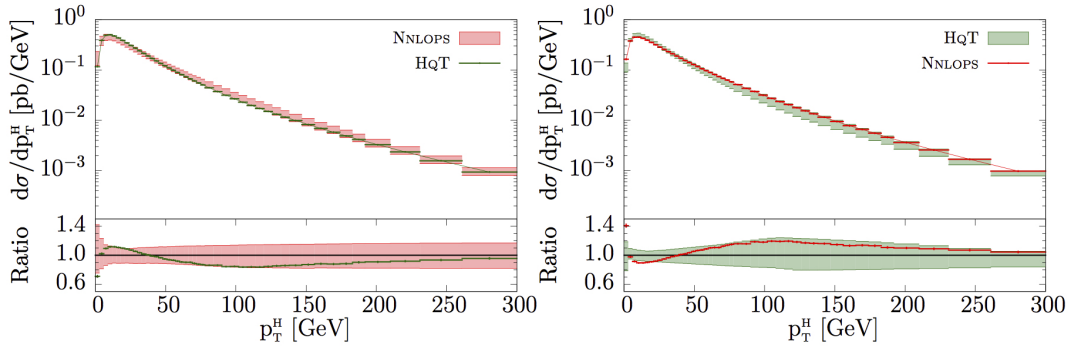


**Fig. 61:** Comparison between the NNLO+PS prediction and the resummed one for the $p_T$ distribution of a Higgs boson produced in the gluon fusion channel at LHC8.

used to perform the reweight) is divergent in the low $p_T$ region. The result obtained binning an event sample obtained with the NNLOPS generator is compared here to the result of the program HqT [43,44] that implements the resummation of the logarithmic enhancement discussed in a previous section. A reasonable agreement is found among the two when the scale variations are taken into account. Interestingly, the extension of the MiNLO and the NNLOPS techniques to the case with more partons in the final state is doable in principle, although it requires further non trivial analytic computations to build the appropriate Sudakov suppression factor, that are still missing.

## Acknowledgements

I thank the organizers of the 2018 edition of the European Schools of High-Energy Physics for the excellent environment I've found at the school, all the discussion leaders for their great job in driving the discussion on the relevant points and all the students for their active and stimulating participation to the lectures. The discussion on several topics in these lectures have been strongly inspired by other lectures, in particular the ones by S. Catani held at the Galileo Galilei Institute in Italy[1] and by P. Nason at CERN[2], that I recommend.

## References

[1] M. Tanabashi et al. Review of Particle Physics. *Phys. Rev.*, D98(3):030001, 2018.

[2] Roel Aaij et al. Observation of the doubly charmed baryon $\Xi_{cc}^{++}$. *Phys. Rev. Lett.*, 119(11):112001, 2017.

[3] A. Esposito, A. L. Guerrieri, L. Maiani, F. Piccinini, A. Pilloni, A. D. Polosa, and V. Riquer. Observation of light nuclei at ALICE and the X(3872) conundrum. *Phys. Rev.*, D92(3):034028, 2015.

[4] R. Keith Ellis, W. James Stirling, and B. R. Webber. QCD and collider physics. *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, 8:1–435, 1996.

[5] Francisco J. Yndurain. *The Theory of Quark and Gluon Interactions*. Theoretical and Mathematical Physics. Springer, Berlin, Germany, 2006.

[6] P. A. Baikov, K. G. Chetyrkin, J. H. Kuhn, and J. Rittinger. Adler Function, Sum Rules and Crewther Relation of Order $O(\alpha_s^4)$: the Singlet Case. *Phys. Lett.*, B714:62–65, 2012.

[7] P. A. Baikov, K. G. Chetyrkin, and J. H. Kühn. Five-loop fermion anomalous dimension for a general gauge group from four-loop massless propagators. *JHEP*, 04:119, 2017.

[8] Thomas Luthe, Andreas Maier, Peter Marquard, and York Schröder. Five-loop quark mass and field anomalous dimensions for a general gauge group. *JHEP*, 01:081, 2017.

[9] F. Herzog, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt. The five-loop beta function of Yang-Mills theory with fermions. *JHEP*, 02:090, 2017.

[10] R. Keith Ellis, D. A. Ross, and A. E. Terrano. The Perturbative Calculation of Jet Structure in e+ e- Annihilation. *Nucl. Phys.*, B178:421–456, 1981.

[11] S. Catani and M. H. Seymour. A General algorithm for calculating jet cross-sections in NLO QCD. *Nucl. Phys.*, B485:291–419, 1997. [Erratum: Nucl. Phys.B510,503(1998)].

[12] S. Frixione, Z. Kunszt, and A. Signer. Three jet cross-sections to next-to-leading order. *Nucl. Phys.*, B467:399–442, 1996.

[13] W. Bartel et al. Experimental Studies on Multi-Jet Production in e+ e- Annihilation at PETRA Energies. *Z. Phys.*, C33:23, 1986. [,53(1986)].

[14] S. Catani, Yuri L. Dokshitzer, M. Olsson, G. Turnock, and B. R. Webber. New clustering algorithm for multi - jet cross-sections in e+ e- annihilation. *Phys. Lett.*, B269:432–438, 1991.

[15] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys.*, B126:298–318, 1977.

[16] Richard D. Ball et al. Parton distributions with LHC data. *Nucl. Phys.*, B867:244–289, 2013.

[17] S. Alekhin, J. Blumlein, and S. Moch. Parton Distribution Functions and Benchmark Cross Sections at NNLO. *Phys. Rev.*, D86:054009, 2012.

---

[1]https : //www.youtube.com/playlist?list = PLDxsZU4NC6Z5ArDOepoRx4EfYmw6S_gne
[2]http : //cds.cern.ch/record/281780?

[18] P. Jimenez-Delgado and E. Reya. Dynamical NNLO parton distributions. *Phys. Rev.*, D79:074023, 2009.

[19] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J.*, C63:189–285, 2009.

[20] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J.*, C75(5):204, 2015.

[21] Georges Aad et al. Measurement of inclusive two-particle angular correlations in pp collisions with the ATLAS detector at the LHC. *JHEP*, 05:157, 2012.

[22] Gavin P. Salam. Towards Jetography. *Eur. Phys. J.*, C67:637–686, 2010.

[23] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-$k_t$ jet clustering algorithm. *JHEP*, 04:063, 2008.

[24] S. Catani, Yuri L. Dokshitzer, M. H. Seymour, and B. R. Webber. Longitudinally invariant $K_t$ clustering algorithms for hadron hadron collisions. *Nucl. Phys.*, B406:187–224, 1993.

[25] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997.

[26] Zoltan Nagy. Next-to-leading order calculation of three jet observables in hadron hadron collision. *Phys. Rev.*, D68:094002, 2003.

[27] Measurement of inclusive jet and dijet cross sections in proton-proton collision data at 7 TeV centre-of-mass energy using the ATLAS detector. 2011.

[28] James Currie, Aude Gehrmann-De Ridder, Thomas Gehrmann, E. W. N. Glover, Alexander Huss, and Joao Pires. Precise predictions for dijet production at the LHC. *Phys. Rev. Lett.*, 119(15):152001, 2017.

[29] Charalampos Anastasiou, Claude Duhr, Falko Dulat, Franz Herzog, and Bernhard Mistlberger. Higgs Boson Gluon-Fusion Production in QCD at Three Loops. *Phys. Rev. Lett.*, 114:212001, 2015.

[30] Bernhard Mistlberger. Higgs boson production at hadron colliders at N³LO in QCD. *JHEP*, 05:028, 2018.

[31] Wojciech Bizoń, Xuan Chen, Aude Gehrmann-De Ridder, Thomas Gehrmann, Nigel Glover, Alexander Huss, Pier Francesco Monni, Emanuele Re, Luca Rottoli, and Paolo Torrielli. Fiducial distributions in Higgs and Drell-Yan production at N³LL+NNLO. *JHEP*, 12:132, 2018.

[32] Torbjorn Sjostrand. Monte Carlo Generators. In *High-energy physics. Proceedings, European School, Aronsborg, Sweden, June 18-July 1, 2006*, pages 51–74, 2006.

[33] Stefan Höche. Introduction to parton-shower event generators. In *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*, pages 235–295, 2015.

[34] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber. QCD matrix elements + parton showers. *JHEP*, 11:063, 2001.

[35] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003.

[36] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 06:029, 2002.

[37] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[38] Stefano Frixione, Paolo Nason, and Bryan R. Webber. Matching NLO QCD and parton showers in heavy flavor production. *JHEP*, 08:007, 2003.

[39] Keith Hamilton, Paolo Nason, Carlo Oleari, and Giulia Zanderighi. Merging H/W/Z + 0 and 1 jet

at NLO with no merging scale: a path to parton shower + NNLO matching. *JHEP*, 05:082, 2013.

[40] Gionata Luisoni, Paolo Nason, Carlo Oleari, and Francesco Tramontano. $HW^\pm$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *JHEP*, 10:083, 2013.

[41] Keith Hamilton, Paolo Nason, Emanuele Re, and Giulia Zanderighi. NNLOPS simulation of Higgs boson production. *JHEP*, 10:222, 2013.

[42] Stefano Catani and Massimiliano Grazzini. An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC. *Phys. Rev. Lett.*, 98:222002, 2007.

[43] Giuseppe Bozzi, Stefano Catani, Daniel de Florian, and Massimiliano Grazzini. Transverse-momentum resummation and the spectrum of the Higgs boson at the LHC. *Nucl. Phys.*, B737:73–120, 2006.

[44] Daniel de Florian, Giancarlo Ferrera, Massimiliano Grazzini, and Damiano Tommasini. Transverse-momentum resummation: Higgs boson production at the Tevatron and the LHC. *JHEP*, 11:064, 2011.

# QCD under extreme conditions: an informal discussion

*E.S. Fraga*
Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil.

**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 2018 European School of High-Energy Physics and previously at CERN Physics Schools in Latin-America, and is aimed at students working in experimental high-energy physics.

**Keywords**
QCD; heavy-ion collision; quark-gluon plasma; 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. 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 $f_\pi$ or $\Lambda_{QCD}$, 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.
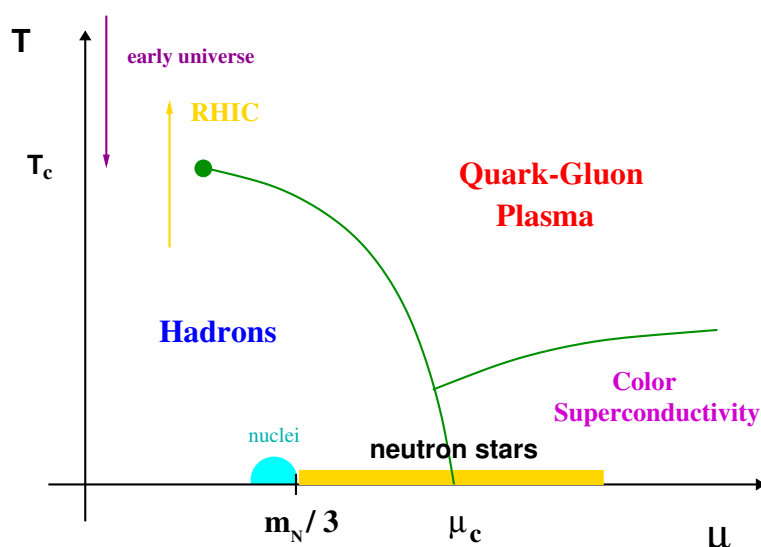


**Fig. 1:** Cartoon of a phase diagram for strong interactions. Extracted from Ref. [12]

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,

$$3p \approx \epsilon \approx \frac{\pi^2}{30} N(T) T^4 \ , \tag{1}$$

where $p$ is the pressure, $\epsilon$ 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)

$$L_{\text{univ}}(T) \approx \frac{1}{4\pi} \left( \frac{1}{1-n} \right) \left( \frac{45}{\pi N(T)} \right)^{1/2} \frac{M_{\text{Pl}}}{T^2} = \frac{1.45 \times 10^{18}}{(T/\text{GeV})^2 \sqrt{N(T)}} \text{fm} \ . \tag{2}$$

Here $M_{\text{Pl}}$ is the Planck mass, and it is clear that the system is essentially in the thermodynamic limit.
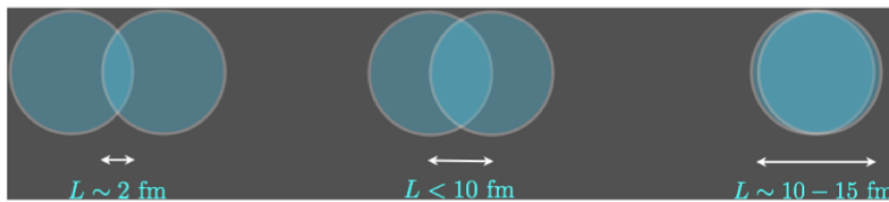


**Fig. 2:** Cartoon representing non-central heavy ion collisions and how they affect the size of the system.

On the other hand, in heavy ion collisions the typical length scale of the system is $L_{\text{QGP}} \lesssim 10 - 15$ 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 $n_0 = 0.16$ fm$^{-3}$ $= 3 \times 10^{14}$g/cm$^3$, which corresponds to squeezing $\sim 2$ solar masses into a sphere of $\sim 10$ 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:

$$\frac{dp}{dr} = -\frac{G\mathcal{M}(r)\epsilon(r)}{r^2 \left[1 - \frac{2G\mathcal{M}(r)}{r}\right]} \left[1 + \frac{p(r)}{\epsilon(r)}\right] \left[1 + \frac{4\pi r^3 p(r)}{\mathcal{M}(r)}\right] , \qquad (3)$$

$$\frac{d\mathcal{M}}{dr} = 4\pi r^2 \epsilon(r) \; ; \quad \mathcal{M}(R) = M . \qquad (4)$$

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
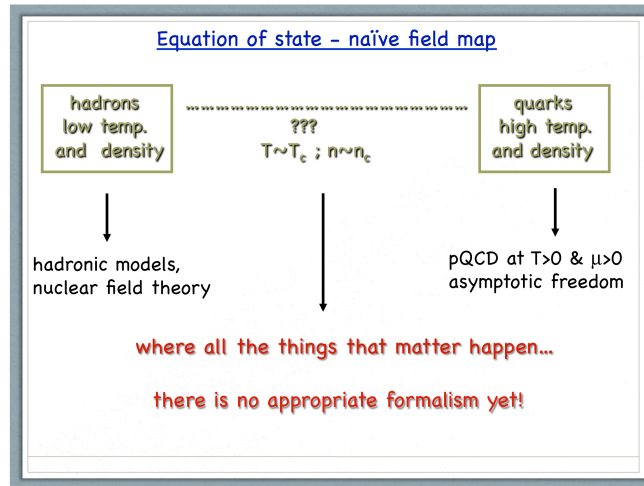
**Fig. 3:** Cartoon of the naïve field map for the equation of state for strong interactions.

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 $\mu$ playing the role of the momentum scale in a plasma, but very complicated in the opposite limit. On top of that, $T$ and $\mu$ 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 $g^2T$ [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

$$E_h^{\min} = \frac{16}{3}\pi R_h^3 B \,, \tag{5}$$

and the hadron pressure (at equilibrium)

$$p_h = \frac{\partial E_h}{\partial V} = -B + \frac{\text{const}}{4\pi R^4} = 0 \,. \tag{6}$$

Assuming the existence of a deconfining transition, the pressure in the quark-gluon plasma phase within this model is given by

$$p_{\text{QGP}} = \left(\nu_b + \frac{7}{4}\nu_f\right)\frac{\pi^2 T^4}{90} - B \,, \tag{7}$$

whereas the pressure in the hadronic phase (taking, for simplicity, a pion gas) is given by

$$p_\pi = \nu_\pi \frac{\pi^2 T^4}{90} \,, \tag{8}$$

neglecting masses. Here, we have the following numbers of degrees of freedom: $\nu_\pi = 3$, $\nu_b = 2(N_c^2 - 1)$ and $\nu_f = 2N_cN_f$ 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:

$$T_c = \left(\frac{45B}{17\pi^2}\right) \approx 144 \text{ MeV} \tag{9}$$

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.
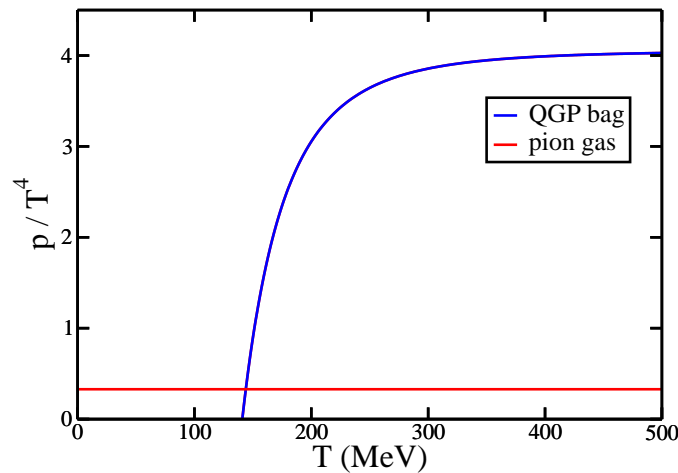


**Fig. 4:** Pressures in the bag model description.

## 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.

The basic hierarchy is the following:

**pure glue SU(N):**
- Z(N) symmetry (SSB)
- order parameter: Polyakov loop **L**
- deconfining trans.: N=2 (2nd order), N=3 (weakly 1st order)

**+ massless quarks:**
- chiral symmetry (SSB)
- order parameter: chiral condensate σ
- Z(N) explicitly broken, but rise of **L** ⇨ deconf.
- chiral trans.: N=3,2 (N_f=2) − 2nd order

**+ massive quarks:**
Z(N) and chiral explicitly broken
Yet  vary remarkably and **L** ↔ σ

**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,

$$\mathcal{L} = \frac{1}{2}\mathrm{Tr}F_{\mu\nu}F^{\mu\nu} + \bar{q}i\gamma^{\mu}D_{\mu}q \,, \tag{10}$$

$$D_{\mu} \equiv (\partial_{\mu} - igA_{\mu}) \,, \tag{11}$$

$$F_{\mu\nu} = \frac{i}{g}\left[D_{\mu}(A), D_{\nu}(A)\right] \,, \tag{12}$$

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])

$$\Omega_c = e^{i\frac{2n\pi}{N_c}}\mathbf{1} \,. \tag{13}$$

At finite temperature, one has also to impose the following boundary conditions:

$$A_\mu(\vec{x}, \beta) = +A_\mu(\vec{x}, 0) , \tag{14}$$

$$q(\vec{x}, \beta) = -q(\vec{x}, 0) . \tag{15}$$

Any gauge transformation that is periodic in $\tau$ will do it. However, 't Hooft noticed that the class of possible transformations is more general. They are such that

$$\Omega(\vec{x}, \beta) = \Omega_c \quad , \quad \Omega(\vec{x}, 0) = \mathbf{1} , \tag{16}$$

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:

$$L(\vec{x}) = \frac{1}{N_c} \text{Tr} \, \mathcal{P} \exp\left[ ig \int_0^\beta d\tau \, \tau^a A_0^a(\vec{x}, \tau) \right] , \tag{17}$$

with $L$ transforming as

$$L(\vec{x}) \mapsto \Omega_c \, L(\vec{x}) \, \mathbf{1} = e^{i\frac{2n\pi}{N_c}} L(\vec{x}) . \tag{18}$$

At very high temperatures, $g \sim 0$, and $\beta \mapsto 0$, so that

$$\langle \ell \rangle = e^{i\frac{2n\pi}{N_c}} \ell_0 \quad , \quad \ell_0 \sim 1 , \tag{19}$$

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:

$$\ell_0 = 0 , \, T < T_c \quad ; \quad \ell_0 > 0 , \, T > T_c . \tag{20}$$

Usually the Polyakov loop is related to the free energy of an infinitely heavy test quark via (confinement: no free quark)

$$\langle \ell \rangle = e^{-F_{test}/T} . \tag{21}$$

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 $\ell$ 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

$$U(N_f)_L \times U(N_f)_R \sim U(N_f)_V \times U(N_f)_A . \tag{22}$$

As $U(N) \sim SU(N) \times U(1)$, one finds

$$U(N_f)_L \times U(N_f)_R \sim SU(N_f)_L \times SU(N_f)_R \times U(1)_V \times U(1)_A , \tag{23}$$
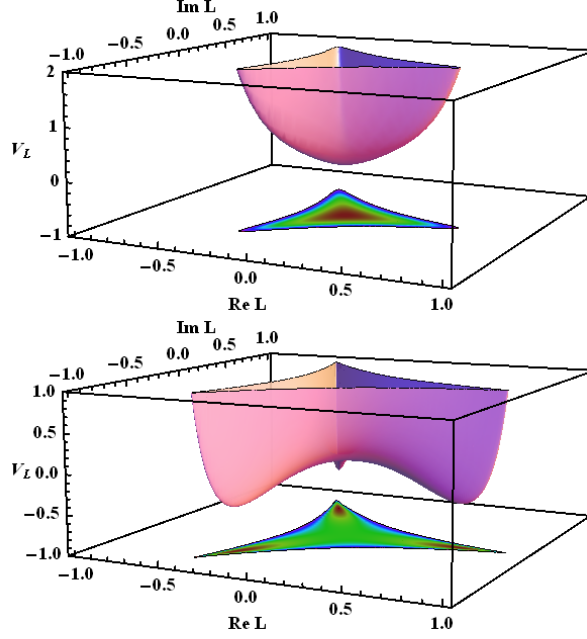
**Fig. 6:** Effective potential for the Polyakov loop for $T < T_c$ (upper) and $T > T_c$ (lower). Extracted from Ref. [32].

where we see the $U(1)_V$ from quark number conservation and the $U(1)_A$ broken by instantons.

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,

$$\mathcal{L} = \frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} + \overline{\psi}_L \gamma^\mu D_\mu \psi_L + \overline{\psi}_R \gamma^\mu D_\mu \psi_R - m_u(\overline{u}_L u_R + \overline{u}_R u_L) - m_d(\overline{d}_L d_R + \overline{d}_R d_L) , \quad (24)$$

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:

$$\langle 0|\overline{\psi}\psi|0\rangle = \langle 0|\overline{\psi}_L \psi_R|0\rangle + \langle 0|\overline{\psi}_R \psi_L|0\rangle , \quad (26)$$

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 $\alpha_s$), 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

## 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, $http : //www.claymath.org/millennium/Yang - Mills\_Theory/$.

[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. Schaefer, 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; E. S. Fraga, L. F. Palhares and P. Sorensen, Phys. Rev. C **84** (2011) 011903.

[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.

# Beyond the Standard Model

*B.C. Allanach*

Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences,
University of Cambridge, Cambridge, United Kingdom

**Abstract**

We cover some current topics in Beyond the Standard Model phenomenology, with an emphasis on collider (particularly Large Hadron Collider) phenomenology. We begin with a review of the Standard Model and some unresolved mysteries that it leaves. Then, we shall heuristically introduce supersymmetry, grand unified theories and extra dimensions as paradigms for expanding the Standard Model. The collider phenomenology of such models is too rich and complex to review, but we give some key examples of how the new states associated with the models might be inferred in Large Hadron Collider events[1]. Before concluding, we finish with a brief description of a quantum field theory approximation that can be used in some cases to reduce model dependence: effective field theory.

**Keywords**

Supersymmetry; extra dimensions; Large Hadron Collider; effective field theories; gauge unification; lectures.

## 1 Introduction

We must remember that the Standard Model of particle physics is a remarkably successful physical theory. It has been tested in literally thousands of different and diverse ways. Some of its predictions (for example the anomalous electron magnetic moment) have been verified to one part in $10^{10}$, whereas some of them (particularly the ones involving low energies and the strong interactions) have only been tested at the $10\%$ level. However, there is to date no unambiguous direct collider measurement which rules it out. The more precise predictions are sensitive to higher loops of Standard Model particles (and in principle could be affected by loops involving beyond the Standard Model particles). Going beyond the Standard Model successfully then should not upset any of these successful predictions, and so any extension is likely to only be a small perturbation, at least at the energy scales currently being probed. Let us now turn to the fundamentals that The Standard Model is built upon.

### 1.1 A basic theory: quantum field theory

Microscopically we have *quantum mechanics* and *special relativity* as two fundamental theories. A consistent framework incorporating these two theories is *quantum field theory (QFT)*. In this theory the fundamental entities are quantum fields. Their excitations correspond to the physically observable elementary particles which are the basic constituents of matter as well as the mediators of all the known interactions. Therefore, fields have a particle-like character. Particles can be classified in two general classes: bosons (spin $s = n \in \mathbb{Z}$) and fermions ($s = n + \frac{1}{2} \forall n \in \mathbb{Z}$). Bosons and fermions have very different physical behaviour. The main difference is that fermions can be shown to satisfy the Pauli "exclusion principle", which states that two identical fermions cannot occupy the same quantum state, and therefore explaining the vast diversity of atoms.

---

[1]A large portion of these notes is based on Prof. Fernando Quevedo's excellent Cambridge Part III "Supersymmetry and extra dimensions" course [1], with his permission.

All apparently elementary matter particles are fermions, for example the leptons (including electrons and neutrinos) and quarks (that make protons, neutrons and all other hadrons). Bosons on the other hand include the photon (particle of light and mediator of electromagnetic interaction), and the mediators of all the other interactions. They are not constrained by the Pauli principle. As we shall see, *supersymmetry* is a symmetry that unifies bosons and fermions despite all their differences.

## 1.2  Basic principle: symmetry

If QFT is the basic framework to study elementary processes, one tool to learn about these processes is the concept of *symmetry*.

A symmetry is a transformation that can be made to a physical system leaving the physical observables unchanged. Throughout the history of science symmetry has played a very important role in better understanding nature.

## 1.3  Classes of symmetries

For elementary particles, we can define two general classes of symmetries:

– *Space-time symmetries:* These symmetries correspond to transformations on a field theory acting explicitly on the space-time coordinates,

$$x^\mu \mapsto x'^\mu (x^\nu) \, \forall \{\mu, \nu\} = \{0, 1, 2, 3\}. \tag{1}$$

Some examples are rotations, translations and, more generally, *Lorentz- and Poincaré transformations* defining special relativity as well as *general coordinate transformations* that define *general relativity*.

– *Internal symmetries:* These are symmetries that correspond to transformations of the different fields in a field theory,

$$\Phi^a(x) \mapsto M^a{}_b \, \Phi^b(x). \tag{2}$$

Roman indices $a, b$ label the corresponding fields[2]. If $M^a{}_b$ is constant then the symmetry is a *global symmetry*; in case of space-time dependent $M^a{}_b(x)$ the symmetry is called a *local symmetry* or a *gauge symmetry*.

## 1.4  Importance of symmetries

Symmetry is important for various reasons:

– *Labelling and classifying particles:* Symmetries label and classify particles according to the different conserved quantum numbers identified by the space-time and internal symmetries (mass, spin, charge, colour, etc.). In this regard symmetries actually "define" an elementary particle according to the behaviour of the corresponding field with respect to the different symmetries.

– Symmetries determine the *interactions* among particles, by means of the *gauge principle*, for instance. It is important that *most QFTs of vector bosons are sick: they are non-renormalisable in a way that makes them lose predictivity*. The counter example to this is *gauge* theory, where vector bosons are *necessarily in the adjoint representation* of the gauge group. As an illustration, consider the Lagrangian

$$\mathcal{L} = \partial_\mu \phi \, \partial^\mu \phi^* - V(\phi, \phi^*) \tag{3}$$

which is invariant under rotations in the complex plane

$$\phi \mapsto \exp(i\alpha) \, \phi, \tag{4}$$

---

[2]Unless otherwise noted, we follow the convention that repeated indices are summed over.

as long as $\alpha$ is a constant (this corresponds to a global symmetry). If $\alpha = \alpha(x)$, the kinetic term is no longer invariant:

$$\partial_\mu \phi \mapsto \exp(i\alpha)\left(\partial_\mu \phi + i(\partial_\mu \alpha)\phi\right). \tag{5}$$

However, the covariant derivative $D_\mu$, defined as

$$D_\mu \phi = \partial_\mu \phi + iA_\mu\, \phi, \tag{6}$$

transforms like $\phi$ itself, if the gauge - potential $A_\mu$ transforms to $A_\mu - \partial_\mu \alpha$:

$$\begin{aligned} D_\mu \phi &\mapsto \exp(i\alpha)\left(\partial_\mu \phi + i(\partial_\mu \alpha)\phi + i(A_\mu - \partial_\mu \alpha)\,\phi\right) \\ &= \exp(i\alpha)\, D_\mu \phi, \end{aligned}$$

so we rewrite the Lagrangian to ensure gauge invariance:

$$\mathcal{L} = D_\mu \phi\, (D^\mu \phi)^* - V(\phi, \phi^*). \tag{7}$$

The scalar field $\phi$ couples to the gauge field $A_\mu$ via $A_\mu \phi A^\mu \phi$, similarly, the Dirac Lagrangian

$$\mathcal{L} = \overline{\Psi}\,\gamma^\mu\, D_\mu \Psi \tag{8}$$

contains an interaction term $\overline{\Psi} A_\mu \Psi$. This interaction provides the three point vertex that describes interactions of electrons and photons, illustrating how photons mediate the electromagnetic interactions.

– Symmetries can hide or be *spontaneously broken:* Consider the potential $V(\phi, \phi^*)$ in the scalar field Lagrangian above.
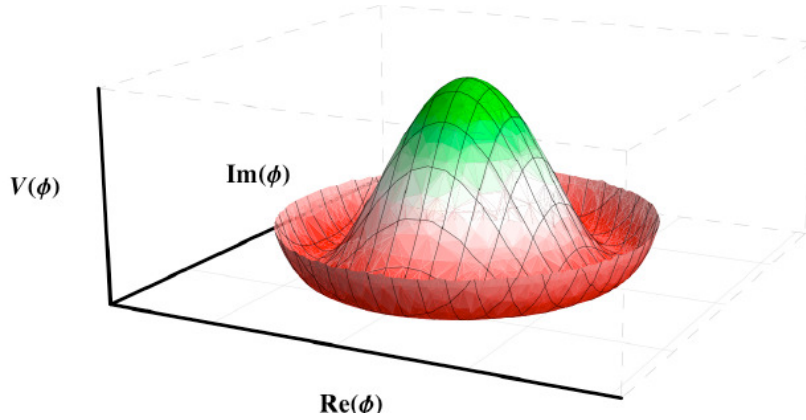


**Fig. 1:** The Mexican hat potential for $V = \left(a - b\,|\phi|^2\right)^2$ with $a, b \geq 0$. From Ref. [1].

If $V(\phi, \phi^*) = V(|\phi|^2)$, then it is symmetric for $\phi \mapsto \exp(i\alpha)\phi$. If the potential is of the type

$$V = a\,|\phi|^2 + b\,|\phi|^4 \forall a, b \geq 0, \tag{9}$$

then the minimum is at $\langle \phi \rangle = 0$ (here $\langle \phi \rangle \equiv \langle 0|\phi|0\rangle$ denotes the *vacuum expectation value (VEV)* of the field $\phi$). The vacuum state is then also symmetric under the symmetry since the origin is invariant. However if the potential is of the form

$$V = \left(a - b\,|\phi|^2\right)^2 \forall a, b \geq 0, \tag{10}$$

the symmetry of $V$ is lost in the ground state $\langle \phi \rangle \neq 0$. The existence of hidden symmetries is important for at least two reasons:

(i) This is a natural way to introduce an energy scale in the system, determined by the non vanishing VEV. In particular, in the Standard Model, the electroweak scale $M_{\text{ew}} \sim 10^2$ GeV defines the basic scale of mass for the particles of the standard model, the electroweak gauge bosons and the matter fields, through their Yukawa couplings, obtain their mass from the VEV.

(ii) The existence of hidden symmetries implies that the fundamental symmetries of nature may be larger than is apparent. This is because the only manifest symmetries we can observe are the symmetries of the vacuum we live in and not those of the full underlying theory. This opens-up an essentially unlimited resource to consider physical theories with an indefinite number of symmetries even though they are not explicitly realised in nature. The standard model is one typical example and supersymmetry and theories of extra dimensions are further examples.

### 1.4.1 The Standard Model

The Standard Model is well defined and currently well confirmed by experiments. It is based on the two classes of symmetry:

– *space-time symmetry*: Poincaré symmetry in 4 dimensions.

– *internal symmetry*: gauged $G_{SM}$=SU(3)$_c$×SU(2)$_L$×U(1)$_Y$ symmetry, where SU(3)$_c$ defines the strong interactions. SU(2)$_L$×U(1)$_Y$ is spontaneously broken by the *Higgs* mechanism to U(1)$_{em}$. The gauge fields are spin-1 bosons, for example the photon $A^\mu$, or gluons $G^{a=1,\cdots,8}$. Matter fields (quarks and leptons) have spin 1/2$\hbar$ and come in three 'families' (successively heavier copies). The Higgs boson (a particle has been discovered at the LHC whose properties are consistent with the Standard Model Higgs boson) is the spin zero particle that spontaneously breaks the SU(2)$_L$×U(1)$_Y$. The $W^\pm$ and $Z^0$ bosons get a mass via the Higgs mechanism and therefore the weak interactions are short range. This is also the source of masses for all quarks and leptons. The sub-index $L$ in $SU(2)_L$ refers to the fact that the Standard Model does not preserve parity and differentiates between left-handed and right-handed particles. In the Standard Model only left-handed fermions (and right-handed anti-fermions) transform non-trivially under $SU(2)_L$. The gauge particles have all spin $s = 1\hbar$ and mediate each of the three forces: photons ($\gamma$) for $U(1)$ electromagnetism, gluons for $SU(3)_C$ of strong interactions, and the massive $W^\pm$ and $Z^0$ bosons for the weak interactions.

## 1.5 Problems of the Standard Model

The Standard Model is one of the cornerstones of all science and one of the great triumphs of the past century. It has been carefully experimentally verified in many ways, especially during the past 20 years. However, there are still some unresolved issues or mysteries:

– The hierarchy problem. The Higgs mass is $m_h \approx 125$ GeV, whereas the gravitational scale is $M_{Planck} \sim \sqrt{G} \sim 10^{19}$ GeV. The 'hierarchy problem' is: why is $m_h/M_{Planck} \sim 10^{-17}$ so much smaller than 1? In a fundamental theory, one might expect them to be the same order. In QFT, one sees that quantum corrections (loops) to $m_h$ are expected to be of order of the heaviest scale in the theory divided by $4\pi$. The question of why the hierarchy is stable with respect to the quantum corrections is called the *technical hierarchy problem*, and is arguably the main motivation for weak-scale supersymmetry.

– The cosmological constant ($\Lambda$) problem: probably the biggest unsolved problem in fundamental physics. $\Lambda$ is the energy density of free space time. The cosmological constant problem is: Why is $(\Lambda/M_{Planck})^4 \sim 10^{-120} \ll 1$?

– The Standard Model has around 20 parameters, which must be measured then set 'by hand'. Many consider that a more satisfying fundamental theory would relate all of these parameters to less (or ideally one) fundamental parameter.

– What particle constitutes the inferred cold dark matter in the universe? It is not contained in the Standard Model. Planck and large scale structure data favour a cosmological constant-cold dark matter model, where approximately 22% of the universe's energy budget lies in dark matter, only 4% in ordinary matter, and some 74% in mysterious dark energy[3]. Neutrinos constitute a hot component of dark matter (since they are relativistic when they decouple from the thermal plasma i.e. they smooth density perturbations in the early universe on smaller scales), so they are not good candidates.
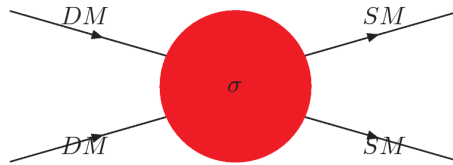


**Fig. 2:** For time $t \rightarrow$ (i.e. time increasing toward the right), this describes *annihilation*: once the particle physics model is set, a calculation tells us how much is thermally produced in the early universe. This also is a diagram for dark matter indirect detection, for example by dark matter collecting in the core of the sun and annihilating into neutrinos which could be detected by the `IceCube` experiment. For $t \leftarrow$, the diagram depicts *collider production* at (e.g.) the LHC, whereas for $t \uparrow$, it's *direct detection*, where dark matter colliding with heavy nuclei may produce measurable nuclear recoils.

– The anomalous magnetic moment of the muon: This is a particular interaction between the photon and the muon: the Dirac equation predicts a muon magnetic moment

$$\vec{M} = g_\mu \frac{e}{2m_\mu} \vec{S}, \tag{11}$$

and at tree level, $g_\mu = 2$. However, it can be measured very precisely by storing muons in a ring with magnetic fields, then measuring the *precession frequency* of their spins. The 'anomalous' part comes from loops involving various particles. Defining $a_\mu \equiv \frac{g_\mu - 2}{2}$ [2],

$$
\begin{aligned}
a_\mu^{\text{exp}} &= 11659209.1(5.4)(3.3) \times 10^{-10}, \qquad a_\mu^{\text{SM}} = 11659180.3(4.2)(2.6) \times 10^{-10}, \\
\Rightarrow \Delta a_\mu &= a_\mu^{\text{exp}} - a_\mu^{\text{SM}} = 28.8(6.3)(4.9) \times 10^{-10},
\end{aligned}
\tag{12}
$$

---

[3] A tiny negative energy density of space-time, $\Lambda \sim \mathcal{O}(10^{-3} \text{ eV})^4$.
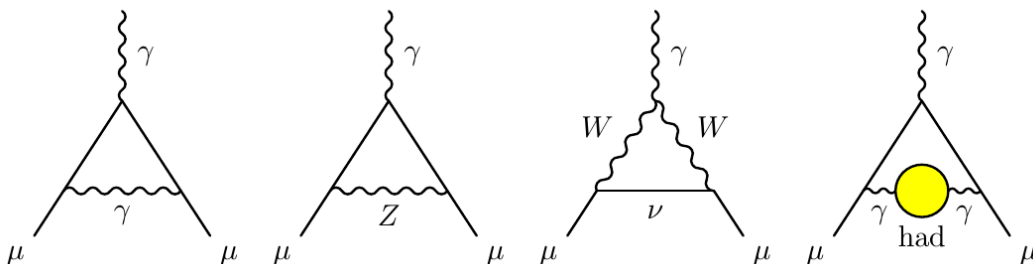


**Fig. 3:** Some SM contributions to the anomalous magnetic moment of the muon. From Ref. [2].

where the first number in brackets labels the statistical error and the second the systematic error. The measurement of $(g-2)_\mu$ thus differs with the SM prediction at around the $\sim 3.6\sigma$ level (and has done for some 20 years). There should be a new more accurate measurement from the Muon $g-2$ experiment at the Fermilab collider in 2017. If one adds new particles to the SM, it is possible that they could travel in loops in diagrams similar to those in Fig. 3, and introduce a non-standard contribution to explain the discrepancy between the SM prediction and the SM measurement.

We wish to find extensions that could solve some or all of the problems mentioned above in order to generalise the Standard Model. Experiments are a traditional way of making progress in science. We need experiments to explore energies above the currently attainable scales and discover new particles and underlying principles that generalise the Standard Model. This approach is of course being followed at the LHC. The LHC will explore physics at the TeV scale, an interesting and important régime for new physics beyond the Standard Model. Notice that directly exploring energies closer to the Planck scale $M_{Planck} \approx 10^{19}$ GeV is out of the reach for many years to come.
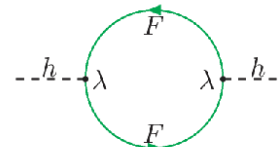
### 1.5.1 The technical hierarchy problem

The Planck mass $M_{\rm pl} \approx 10^{19}$ GeV is an energy scale associated with gravity and the electroweak scale $M_{\rm ew} \approx 10^2$ GeV is an energy scale associated with the electroweak symmetry breaking scale of the Standard Model. The hierarchy problem involves these two scales being so different in magnitude. Actually the problem can be formulated in two parts:

(i) Why is $M_{\rm ew} \ll M_{\rm pl}$ at tree level? This question is known as 'the hierarchy problem'. There are many solutions, once the SM is extended.

(ii) Once we have solved (i), we ask why is the hierarchy stable under quantum corrections? This is the 'technical hierarchy problem' and does not have many full/effective solutions, aside from supersymmetry (SUSY).

Let us now think some more about the technical hierarchy problem. In the Standard Model we know that:

– Vector bosons are massless due to gauge invariance, that means, a direct mass term for the gauge particles $M^2 A_\mu A^\mu$ is not allowed by gauge invariance ($A_\mu \to A_\mu + \partial_\mu \alpha$ for a $U(1)$ field, for example).

– Chiral fermion masses $m\bar{\psi}_L \psi_R$ are also forbidden for all quarks and leptons by gauge invariance (because, for example, $\psi_L$ and $\psi_R$ have different hypercharges). Recall that these particles receive a mass only through the Yukawa couplings to the Higgs (e.g. $H\bar{\psi}_L \psi_R$ giving a Dirac mass to $\psi$ after $H$ gets a non-zero value[4]).

– The Higgs boson is the only fundamental scalar particle in the Standard Model. There is no symmetry banning its mass term $m_h^2 H^\dagger H$ in the Standard Model Lagrangian. If the heaviest state in the theory has a mass squared of $\Lambda^2$, loops give corrections of order $\Lambda^2/(16\pi^2)$ to the scalar mass squared. The corrections come from both bosons and fermions running in loops, for example:



$$\sim -\frac{a\lambda^2}{16\pi^2}\int \frac{d^n k}{k^2 - m_F^2} + \ldots, \qquad (13)$$

where $a$ is some dimensionless $\mathcal{O}(1)$ constant. The quantum correction to the Higgs mass from this diagram are:

$$m_h^{phys\,2} = (125\text{ GeV}/c^2)^2 = m_h^{tree\,2} + \mathcal{O}(m_F^2/(16\pi^2)). \qquad (14)$$

---

[4]With $R-$parity conservation (see below), the minimal supersymmetric standard model does not give neutrinos mass. Thus one must augment the model in some way: one can do this by adding right-handed neutrinos to the model.

Experimentally, the Higgs mass is measured to be $m_h \approx 125$ GeV. The Standard Model is considered to be unnatural since the loop corrections are typically much larger: the largest are expected to be[5] $\sim \mathcal{O}(10^{17})$ GeV. Therefore even if we start with a tree-level Higgs mass of order the electroweak scale, loop corrections would bring it up to almost the highest scale in the theory: $\Lambda/(4\pi)$, since we expect $m_F \sim \mathcal{O}(\Lambda)$. This would ruin the hierarchy between large and small scales. It is possible to adjust or "fine tune" the loop corrections such as to keep the Higgs light, but this would require cancellations between the apparently unrelated tree-level and loop contributions to some 15 significant figures. This fine tuning is considered unnatural and an explanation of why the Higgs mass (and the whole electroweak scale) can be naturally maintained to be hierarchically smaller than the Planck scale or any other large cutoff scale $\Lambda$ is required.

### 1.5.2   Modifications of the Standard Model

In order to go beyond the Standard Model we can follow several avenues, for example:

– Add new particles and/or interactions (e.g. a dark matter particle).

– More symmetries. For example,

(i) Internal symmetries, for example *grand unified theories (GUTs)* in which the symmetries of the Standard Model are themselves the result of the breaking of a yet larger symmetry group:

$$G_{\text{GUT}} \xrightarrow{M \approx 10^{16}\text{GeV}} G_{\text{SM}} \xrightarrow{M \approx 10^2\text{GeV}} SU(3)_c \times U(1)_Y, \tag{15}$$

Let's take one of the simplest examples, $G_{\text{GUT}} = SU(5)$:

$$\underline{5} = \begin{pmatrix} d \\ d \\ d \\ e^+ \\ \bar{\nu}_e \end{pmatrix}_R, \qquad \underline{10} = \begin{pmatrix} 0 & \bar{u} & -\bar{u} & -u & -d \\ & 0 & \bar{u} & -u & d \\ & & 0 & -u & d \\ & & & 0 & e^+ \\ & & & & 0 \end{pmatrix}_L. \tag{16}$$

(The $\underline{10}$ is an anti-symmetric matrix; we have omitted the lower left-hand half of it because the entries are simply related to those above the diagonal). Thus, we see how quarks and leptons become unified within multiplets of $G_{\text{GUT}}$.

The GUT proposal is very elegant because it unifies, in one single symmetry, the three gauge interactions of the Standard Model. It leaves unanswered most of the open questions above, except for the fact that it reduces the number of independent parameters due to the fact that there is only one gauge coupling at large energies. This is expected to "run" at low energies and give rise to the three different couplings of the Standard Model (one corresponding to each group factor). Unfortunately, with our present precision understanding of the gauge couplings and spectrum of the Standard Model, the running of the three gauge couplings does **not** unify at a single coupling at higher energies but they cross each other at different energies: see Fig. 4. Because leptons and quarks are unified within GUT multiplets, they predict e.g. $m_e(M_{GUT}) = m_d(M_{GUT})$, which also doesn't work, and in practice further model building is required.

GUTs have heavy $X$ and $Y$ gauge boson particles of order the gauge unification scale, which arise from a GUT Higgs mechanism (in a completely analogous way to the way in which the $W^\pm$ and $Z^0$ bosons acquire their mass).They predict *proton decay*, which isn't observed at super-Kamiokande. The current constraint from super-Kamiokande is that the proton lifetime $\tau_{p \to e^+ \pi^0} > 10^{34}$ years. However, estimating $M_{GUT} \sim 10^{15}$ GeV from Fig. 4, we predict, for ordinary GUTs, a proton lifetime of

$$\tau \approx \frac{M_{GUT}^4}{\alpha^2 m_p^5} = 4.5 \times 10^{29 \pm 1.7} \text{ years,} \tag{17}$$

---

[5]This does rely on quantum gravity yielding an effective quantum field theory that acts in the usual way.
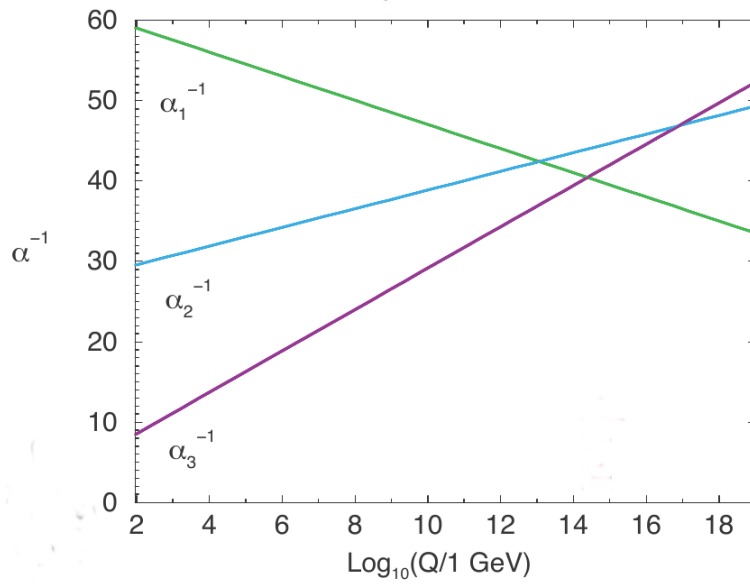
**Fig. 4:** Gauge unification doesn't work in the Standard Model: the three gauge couplings $\alpha_1, \alpha_2, \alpha_3$ should all unify at a single renormalisation scale $Q$. One needs to add some additional particles of mass below $10^{14}$ GeV in order to make this work. Experiments (LEP and LHC experiments, for example) fix the gauge couplings at the left-hand side of the figure, and renormalisation within QFT is used to evolve them to the right. From Ref. [2].
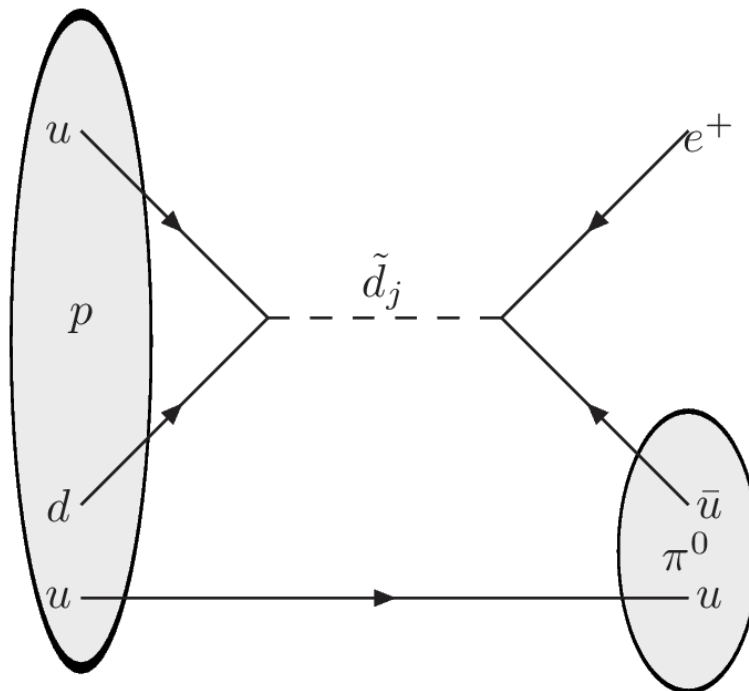


**Fig. 5:** Example $p \rightarrow e^+ \pi^0$ process from GUTs. From Ref. [2].

which easily is in contravention of the Super Kamiokande bound.

(ii) *Supersymmetry*. For a phenomenological review of supserymmetry, see Ref. [3]. Supersymmetry is an external, or space-time, symmetry. Supersymmetry solves the technical hierarchy problem due to cancellations between the contributions of bosons and fermions to the electroweak scale, defined by the Higgs mass. Combined with the GUT idea, it also solves the unification of the three gauge couplings at one single point at larger energies. Supersymmetry also provides the most studied example for dark matter candidates. Moreover, it provides well defined QFTs in which the régime of strong coupling can be better studied than in non-supersymmetric models.

(iii) Extra spatial dimensions. More general space-time symmetries open up many more interesting avenues for investigation. These can be of two types. First we can add more dimensions to space-time, extending the Poincaré symmetries of the Standard Model and the general coordinate transformations of general relativity. This is the well known *Kaluza Klein theory* in which our observation of a 4 dimensional universe is only due to the fact that we have limitations about "seeing" other dimensions of space-time that may be hidden to our observations. In recent years this has been extended to the *brane world scenario* in which our 4 dimensional universe is only a brane or surface inside a higher dimensional universe. These ideas lead to a different perspective on the hierarchy problem and also may help unify internal and space-time symmetries.

– Beyond QFT: A QFT with Supersymmetry and extra dimensions does not address the problem of quantising gravity. For this purpose, the current best hope is string theory which goes beyond the basic framework of QFT. It so happens that for its consistency, string theory requires supersymmetry and extra dimensions.

## 1.6 Supersymmetry algebra

### 1.6.1 *History of supersymmetry*

– In the 1960's, the study of strong interactions lead to the discovery of many hadrons. These were successfully organised into multiplets of $SU(3)_f$, the $f$ referring to flavour. This procedure was known as the *eight fold way* of Gell-Mann and Neeman. Questions arose about bigger multiplets including particles of different spins.

– In a famous *No-go theorem* (Coleman, Mandula 1967) said that the most general symmetry of the $S$ - matrix (which still has non-trivial scattering) is Poincaré × internal. The implication is that there is no symmetry that mixes up the internal and external symmetries in a non-trivial way, or that mixes particles of different spin, and still has scattering.

– Golfand and Licktman (1971) extended the Poincaré algebra to include spinor generators $Q_\alpha$, where $\alpha = 1, 2$.

– Ramond, Neveu-Schwarz, Gervais, Sakita (1971) derived supersymmetry in 2 dimensions (from string theory).

– Wess and Zumino (1974) wrote down supersymmetric field theories in 4 dimensions. They opened the way for many other contributions to the field. This is often seen as the actual starting point for the systematic study of supersymmetry.

– Haag, Lopuszanski, Sohnius (1975): generalised the Coleman Mandula theorem to show that the only non-trivial quantum field theories have a symmetry group of super Poincaré group in a direct product with internal symmetries.

### 1.6.2 *Graded algebra*

The Poincaré algebra consists of commutation relations between 4-momentum operators $P^\mu$ (generating translations in space and time) and $M^{\mu\nu}$, generating Lorentz boosts and rotations. Particles of the Standard Model are all irreducible representations of the Poincaré group.

To implement supersymmetry, we *extend* the Poincaré algebra non-trivially. The *Coleman Mandula theorem* stated that in 3+1 dimensions, one cannot do this in a non-trivial way and still have non-zero scattering amplitudes. In other words, there is no non-trivial mix of Poincaré and internal symmetries with non-zero scattering except for the direct product

Poincaré × internal.

However (as usual with no-go theorems) there was a loop-hole because of an implicit axiom: the proof only considered "*bosonic* generators".

We wish to turn bosons into fermions, thus we need to introduce a fermionic generator $Q$. Heuristically:

$$Q|\text{boson}\rangle \propto |\text{fermion}\rangle, \qquad Q|\text{fermion}\rangle \propto |\text{boson}\rangle.$$

For this, we require a underlined graded algebra - a generalisation of Lie algebra. If $O_a$ is an operator of an algebra (such as a group generator), a graded algebra is

$$O_a O_b - (-1)^{\eta_a \eta_b} O_b O_a = iC^e_{ab} O_e, \tag{18}$$

where $\eta_a = 0$ if $O_a$ is a *bosonic generator*, and $\eta_a = 1$ if $O_a$ is a *fermionic generator*.

For supersymmetry, the bosonic generators are the Poincaré generators $P^\mu$, $M^{\mu\nu}$ and the fermionic generators are $Q^A_\alpha$, $\bar{Q}^A_{\dot\alpha}$, where $A = 1, \ldots, N$. In case $N = 1$ we speak of a simple supersymmetry (SUSY), in the case $N > 1$, of an extended SUSY. Here, we will only discuss the more immediately phenomenologically relevant case of $N = 1$.

## 2 Introducing the minimal supersymmetric standard model (MSSM)

The MSSM is based on $SU(3)_C \times SU(2)_L \times U(1)_Y \times (N = 1 \text{ SUSY})$. We must fit all of the experimentally discovered field states into $N = 1$ 'super multiplets': just as quarks are 3 dimensional representations of $SU(3)$ (i.e. one has a red, blue and green quark all within one multiplet), the MSSM fits all of its particles into super multiplets, whose types are:

– *Chiral super multiplets*: These contain a chiral left-handed fermion and a complex scalar.
– *Vector super multiplets*: These contain a spin $1\hbar$ vector boson and a spin $1/2\hbar$ Majorana fermion.

Super multiplets are formally built up from the algebra (we omit such technical details from these lectures). Since the symmetry group is a direct product between SUSY and the SM gauge symmetries, one can perform a SUSY transformation without changing the gauge quantum numbers of the super multiplet. Spin $1\hbar$ vector bosons (e.g. the gluon) must be in the adjoint representation (for $SU(3)$ this has eight colour states) in order to make a renormalisable QFT, therefore the vector super multiplets must be in the adjoint representation. Thus, the spin $1/2\hbar$ copy must also be in the adjoint representation (thus, from our example, we predict eight colour states of spin $1/2\hbar$ fermions: the 'gluinos'). Supersymmetry imposes that the two partners $\tilde{f}_{L,R}$ and $F$ of the super multiplet should couple with the same strengths as each other to other particles, and it also imposes that they should have the same mass as each other. Since

$$m^2_{\tilde{f}_{L,R}} = m^2_F, \tag{19}$$

and the scalars $\tilde{f}_{L,R}$ and the fermion $F$ couple to the Higgs field $h$ with the same strength coupling $\lambda$:



$$\sim \mathcal{O}\left(\frac{m_h^2 \log(M_Z/m_F)}{16\pi^2}\right). \tag{20}$$

Even if $F$ is a very heavy field associated with the highest scale of new physics, Eq. 20 does not present a huge correction to $m_h^2$: it is a usual loop-level correction, adding a few percent. The really huge corrections from Eq. 14 $\propto m_F^2$ have been cancelled between the two diagrams[6] in Eq. 20. This is how supersymmetry solves the technical hierarchy problem.

Eq. 19 is not realised in nature (no one has seen a scalar version of the electron with the same mass as it, for example) and so we must bear in mind that supersymmetry must eventually be broken. However, we only wish to break it in a way that preserves it as a solution to the technical hierarchy problem: in specific models of supersymmetry breaking this can be done, but the coupling relations (that superpartners couple to other fields with the same strength as their SM partners) remain valid even after SUSY breaking. In particular, Eqs. 19 and 20 become

$$m_{\tilde{f}_{L,R}}^2 = m_F^2 + \Delta m^2, \tag{21}$$

whilst the scalars $\tilde{f}_{L,R}$ and the fermion $F$ still couple to the Higgs field $h$ with the same strength coupling $\lambda$:



$$\sim \mathcal{O}\left(\frac{m_h^2 \log(M_Z/m_F) + \Delta m^2}{16\pi^2}\right). \tag{22}$$

Thus, as long as the *splitting* between the particles in a super multiplet is small, and as long as certain SUSY relations are preserved (such as the coupling of the Higgs field to the scalar and fermionic components of a super multiplet being equal), one still obtains only reasonable corrections to the Higgs mass squared, even if the fields $F$ and $\tilde{f}_{L,R}$ are very heavy. The fact that we require $\Delta m^2/(16\pi^2)$ to be not much larger than $m_h^2 = (125\,\text{GeV})^2 \Rightarrow \Delta m^2 < \mathcal{O}(1\,\text{TeV}^2)$. This is then the main argument for why supersymmetric partners of SM particles should not be much heavier than the TeV scale, because otherwise its correction to the Higgs mass would be too large. Given that the LHC currently operators at a centre of mass energy of 13 TeV, this implies that there ought to be enough energy to pair produce such sparticles.

## 2.1 Particles

First of all, we have vector superfields containing the Standard Model gauge bosons. We write their representations under $(SU(3)_C,\ SU(2)_L,\ U(1)_Y)$ as (pre-Higgs mechanism):

– gluons/gluinos
$$G = (8, 1, 0)$$

– $W$ bosons/winos
$$W = (1, 3, 0)$$

– $B$ bosons/gauginos
$$B = (1, 1, 0),$$

which contains the gauge boson of $U(1)_Y$.

Secondly, there are chiral superfields containing Standard Model matter and Higgs fields. Since chiral superfields only contain left-handed fermions, we place charge conjugated, i.e. *anti* right handed fermionic fields (which are actually left-handed), denoted by $^c$ ($\{i, j, k\} \in \{1, 2, 3\}$ are family indices):

---

[6]Recall that loops of fermions acquire a minus sign in the sum as compared to scalars.

– (s)quarks: lepton number $L = 0$, whereas baryon number $B = 1/3$ for a (s)quark, $B = -1/3$ for an anti-quark.

$$\underbrace{Q_i = \left(3, 2, \tfrac{1}{6}\right),}_{\text{left-handed}} \underbrace{u_i^c = \left(\bar{3}, 1, -\tfrac{2}{3}\right), d_i^c = \left(\bar{3}, 1, \tfrac{1}{3}\right)}_{\text{anti (right-handed)}}$$

– (s)leptons $L = 1$ for a lepton, $L = -1$ for an anti-lepton. $B = 0$.

$$\underbrace{L_i = \left(1, 2, -\tfrac{1}{2}\right),}_{\text{left-handed}} \underbrace{e_i^c = \left(1, 1, +1\right)}_{\text{anti (right-handed)}}$$

– Higgs bosons/higgsinos: $B = L = 0$.

$$H_2 = \left(1, 2, \tfrac{1}{2}\right), H_1 = \left(1, 2, -\tfrac{1}{2}\right)$$

the second of which is a new Higgs doublet not present in the Standard Model. Thus, the MSSM is a *two Higgs doublet model*. The extra Higgs doublet is needed in order to avoid a gauge anomaly, and to give masses to down-type quarks and leptons.

Note that after the breaking of electroweak symmetry (see the Standard Model course), the electric charge generator is $Q = T_3^{SU(2)_L} + Y/2$. Baryon and lepton number correspond to multiplicative discrete perturbative symmetries in the SM, and are thus conserved, perturbatively.

Chiral fermions may generate an *anomaly* in the theory, as shown by Fig. 6. This is where a symmetry that is present in the tree-level Lagrangian is broken by quantum corrections. Here, the symmetry is $U(1)_Y$: all chiral fermions in the theory travel in the loop, and yield a logarithmic divergence proportional to

$$A \equiv \sum_{LH\ f_i} Y_i^3 - \sum_{RH\ f_i} Y_i^3 \tag{23}$$

multiplied by some kinematic factor which is the same for each fermion. If $A$ is non-zero, one must renormalise the diagram away by adding a $B_\mu B_\nu B_\rho$ counter term in the Lagrangian. But this breaks $U(1)_Y$, meaning that $U(1)_Y$ would not be a consistent symmetry at the quantum level. Fortunately,
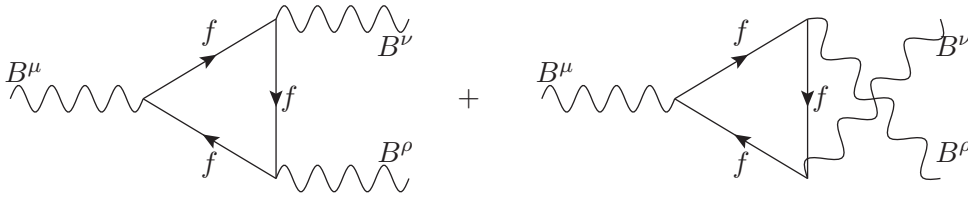


**Fig. 6:** Anomalous Feynman diagrams proportional to $\text{Tr}\{Y^3\}$. The sum of them must vanish for $U(1)_Y$ to be a valid symmetry at the quantum level. Hyper-charged chiral fermions $f$ travel in the loop contributing to a three-hypercharge gauge boson $B$ vertex. From Ref. [1].

$A = 0$ for each fermion family in the Standard Model. Contributions are from (the factors of 3 are from the different colours of the quarks, whereas the factors of 2 come from the different $SU(2)_L$ degrees of freedom):

$$\underbrace{3 \times 2 \times (\tfrac{1}{6})^3}_{Q_L} + \underbrace{3 \times (-\tfrac{2}{3})^3}_{u_R^c} + \underbrace{3 \times (\tfrac{1}{3})^3}_{d_R^c} + \underbrace{2 \times (-\tfrac{1}{2}))^3}_{L_L} + \underbrace{1^3}_{e_R^c} = 0.$$
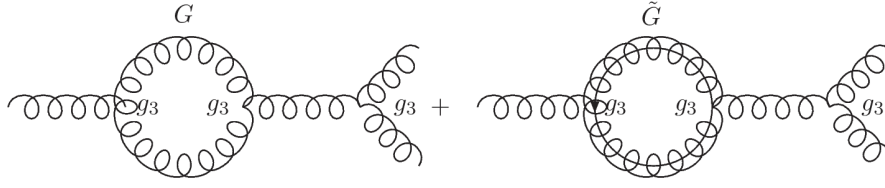
**Fig. 7:** Example Feynman diagrams leading to renormalisation of the strong coupling constant $g_3$. The left-hand diagram renormalises the QCD gauge coupling in the Standard Model, whereas in the MSSM, we have additional contributions from supersymmetric particles such as the one on the right-hand side with gluinos in the loop. There are other contributing diagrams, some involving loops of quarks and squarks, for instance.

In SUSY, we add the Higgsino doublet $\tilde{H}_1$, which yields a non-zero contribution to $A$. This must be cancelled by another Higgsino doublet with opposite $Y$: $\tilde{H}_2$.

There is another special super multiplet sometimes considered to be part of the MSSM with $B = L = 0$. This is the *gravity super multiplet*, with the spin $2\hbar$ graviton and a spin $3/2\hbar$ gravitino. Usually, after SUSY breaking (see later), the only component of the gravitino that couples with non-negligible strength is its spin $1/2\hbar$ component.

$$G = (1, 1, 1)$$

## 2.2 Interactions

– Gauge couplings are renormalised, which ends up giving them *renormalisation scale dependence*, which matches onto dependence upon the energy scale at which one is probing them:

$$\mu \frac{dg_a(\mu)}{d\mu} = \beta_a g_a^3(\mu), \Rightarrow g_a^{-2}(\mu) = g_a^{-2}(\mu_0) - 2\beta_a \ln \frac{\mu}{\mu_0} \tag{24}$$

where $\beta_a$ is a constant determined by which particles travel in the loop in the theory. For ordinary QCD it is $\beta_3 = -7/(16\pi^2)$ whereas for the MSSM, it is $\beta_3 = -3/(16\pi^2)$ because of additional contributions from squarks and gluinos to the loops, as in Fig. 7.

Eq. 24 is used to extrapolate gauge couplings measured at some energy scale $\mu_0$ (often taken to be $M_Z$, from LEP constraints) to some other scale $\mu$. With the SUSY contributions in the MSSM, the gauge couplings almost meet at a renormalisation scale $E \approx 2 \times 10^{16}$ GeV (see Fig. 8), whereas with just the Standard Model contributions, they do not meet each other at all: see Fig. 4. The meeting of the gauge couplings is a necessary condition for a Grand Unified Theory, which only has one gauge coupling (above $M_{GUT} \approx 2 \times 10^{16}$ GeV). $\alpha_1(M_Z)$ and $\alpha_2(M_Z)$ are both known with high accuracy from the LEP experiments, so we can use them to predict $M_{GUT} \sim 10^{16}$ GeV and $\alpha_s(M_Z) = 0.129 \pm 0.002$. The experimental determination[7] of $\alpha_s(M_Z) = 0.119 \pm 0.002$, so the naive prediction is some $5\sigma$ out. However, this small difference is easily explained by GUT threshold corrections (for example because the $X$ or $Y$ bosons are a factor of a few lighter than $M_{GUT}$ and change the running near the GUT scale) in explicit GUT models.

Gauge couplings are renormalised, which ends up giving them *renormalisation scale dependence*, which matches onto dependence upon the energy scale at which one is probing them (one achieves a worse approximation in a truncated perturbation series by picking the renormalisation scale to be vastly different to the energy scales probed in some process): integrating both sides,

$$\mu \frac{dg_a(\mu)}{d\mu} = \beta_a g_a^3(\mu), \Rightarrow g_a^{-2}(\mu) = g_a^{-2}(\mu_0) - 2\beta_a \ln \frac{\mu}{\mu_0} \tag{25}$$

---

[7]We quote SM gauge couplings in the $\overline{MS}$ scheme.
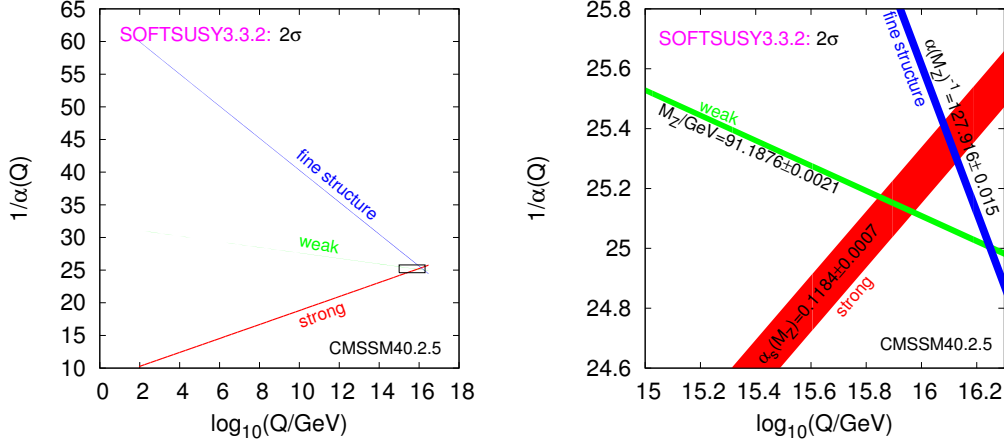
**Fig. 8:** Gauge unification in the MSSM: the thickness of the lines corresponds to the $2\sigma$ error bars. The right-hand panel shows a zoom of the unification region near $Q \sim 10^{16}$ GeV.

where $\beta_a$ is a constant determined by which particles travel in the loop in the theory. For ordinary QCD it is $\beta_3 = -7/(16\pi^2)$ whereas for the MSSM, it is $\beta_3 = -3/(16\pi^2)$ because of additional contributions from squarks and gluinos to the loops.

– A 'superpotential' is like a Lagrangian energy density for SUSY theories: it encodes some of the interactions between the chiral superfields in a way that preserves SUSY. A superpotential term $W = \lambda\Phi^3$ for a chiral superfield $\Phi = (\varphi, \psi)$ encodes both a Yukawa interaction $\mathcal{L} = -\lambda\varphi\psi\psi$ and a scalar interaction $\mathcal{L} = -|\lambda|^2|\varphi|^4$, for example.

We write down a superpotential containing all terms which are renormalisable and consistent with our symmetries. If one does this, one obtains two classes of terms, $W = W_{R_p} + W_{RPV}$. The terms in $W_{R_p}$ all conserve baryon number $B$ and lepton number $L$, whereas those in $W_{RPV}$ break either $B$ or $L$:

$$W_{R_p} = (Y_U)_{ij}\,Q_i\,H_2\,u_j^c + (Y_D)_{ij}\,Q_i\,H_1\,d_j^c + Y_E\,L_i\,H_1\,e_j^c + \mu\,H_1\,H_2 \tag{26}$$

$$W_{RPV} = \lambda_{ijk}\,L_i\,L_j\,e_k^c + \lambda'_{ijk}\,L_i\,Q_j\,d_k^c + \lambda''_{ijk}\,u_i^c\,d_j^c\,d_k^c + \kappa_i\,L_i\,H_2, \tag{27}$$

where we have suppressed gauge indices. Since superfields commute in $W$,

$$H_1^a H_1^b \epsilon_{ab} = \frac{1}{2}(H_1^a H_1^b + H_1^b H_1^a)\epsilon_{ab} = \frac{1}{2}H_1^a H_1^b(\epsilon_{ab} + \epsilon_{ba}) = 0 \tag{28}$$

The first three terms in $W_{R_p}$ correspond to standard Yukawa couplings and give masses to up quarks, down quarks and leptons, as we shall see. Writing $x = 1, 2, 3$ as a fundamental $SU(3)$ index, $a, b = 1, 2$ as fundamental $SU(2)$ indices, the first term in $W_{R_p}$ becomes

$$(Y_U)_{ij}Q_i^{xa}H_2^b u_{jx}^c \epsilon_{ab} = (Y_U)_{ij}[u_L^x H_2^0 u_{jx}^c - d_L^x H_2^+ u_{jx}^c]. \tag{29}$$

Once the neutral Higgs component develops a vacuum expectation value, $H_2^0 \equiv (v_2 + h_2^0)/\sqrt{2}$, the first term becomes $(Y_U)_{ij}v_2/\sqrt{2}u_{Li}^x u_{jx}^c + \ldots$, yielding a Dirac mass matrix $m_u \equiv (Y_U)_{ij}v_2/\sqrt{2}$ for the up quarks. The down quark and lepton masses proceed in an analogous manner. The fourth term is a mass term for the two Higgs(ino) fields.

If all of the terms in $W_{RPV}$ are present, the interaction shown in Fig. 9 would allow proton decay $p \to e^+ + \pi^0$ within seconds because

$$\Gamma(p \to e^+\pi^0) \approx \frac{\lambda'^2_{11k}\lambda''^2_{11k}}{16\pi^2\tilde{m}_{d_k}^4}M_p^5, \tag{30}$$

whereas experiments say that it should be $> 10^{34}$ years. Alternatively, we could make the RPV couplings very small to make the proton long-lived, by imposing the implied bound on $\Gamma(p \to e^+\pi^0)$:

$$\lambda'_{11k} \cdot \lambda''_{11k} < 10^{-27} \left( \frac{\tilde{m}_{d_k}}{100 \text{ GeV}} \right)^2 . \tag{31}$$

In order to forbid proton decay an extra symmetry should be imposed. One symmetry that works
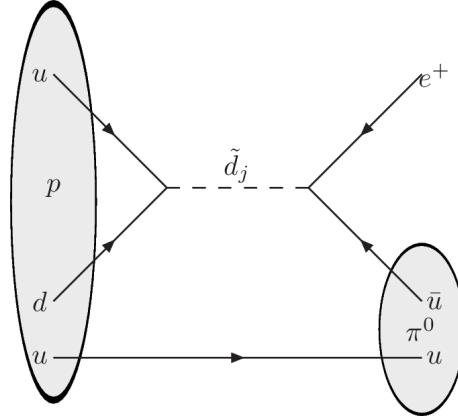


**Fig. 9:** Proton decay $p \to e^+\pi^0$ due to baryon- and lepton number violating interactions. Both $B$ and $L$ violating terms must be present for the proton to decay. The matrix element is proportional to $\lambda''_{1j1}{}^* \times \lambda'_{11j}{}^*$.

is a discrete multiplicative symmetry *R parity* defined as

$$R \equiv (-1)^{3(B-L)+2S} = \{ \ +1 : \text{Standard Model particles}, \qquad -1 : \text{superpartners} \ . \tag{32}$$

It forbids all of the terms in $W_{RPV}$, but there exist other examples which only ban some subset.

R parity would have important physical implications:

– The lightest superpartner (LSP) is stable, because it is $R-$parity odd.
– Cosmological constraints then say that a stable LSP must be electrically and colour-neutral (higgsino, photino, zino). It is then a good candidate for cold weakly interacting dark matter.
– In colliders, the initial state is $R_p = +1$, implying that superparticles are produced in pairs. When a superparticle decays, it must decay to another (lighter) superparticle plus some standard model particles.
– One ends up with LSPs at the end of the decays. These do not interact with the detector, and hence appear as unbalanced or 'missing' momentum.

Note that the terms in $W_{RPV}$ can lead to Majorana fermion structure[8]. For instance, $W = \lambda''_{112} u^c_1 d^c_1 d^c_2$: we take the $F-$ terms as usual in order to find the Lagrangian in terms of components:

$$\mathcal{L} = \frac{1}{2} \left( \lambda''_{112} \tilde{u}^*_1 d^\dagger_{1R} C d^*_{2R} - (\lambda''_{112})^* \tilde{u}_1 d^T_{1R} C^* d_{2R} \right)$$

plus supersymmetric copies, where $C$ is the charge conjugation matrix and $^T$ denotes transpose.

RPV has several potential motivations and characteristics:

---

[8]This is a familiar structure for people extending the Standard Model to include neutrino masses.
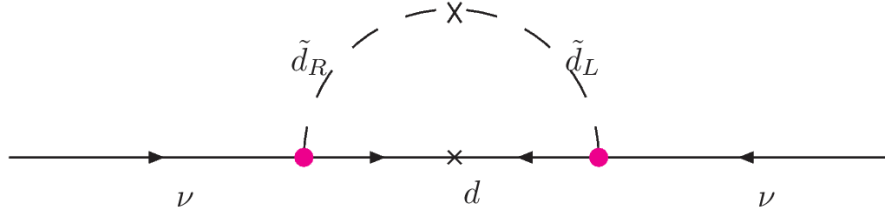
**Fig. 10:** RPV generation of neutrino masses and mixings. Here, the dots show the $L$ violating RPV couplings.

- It has many additional search possibilities[9]
- Dark matter changes character: one loses the usual neutralino dark matter candidate. However, the SUSY breaking sector always contains other fields that may be used instead, for example the gravitino or hidden sector fields. Either of these two candidates is so weakly coupled that direct or indirect dark matter detection becomes extremely unlikely, although inference of its production at colliders is still possible.
- Neutrino masses and mixings are generated by the $L$ violating couplings in diagrams like those in Fig. 10, and the mechanism of their generation is potentially testable at the LHC (unlike, for example, the seesaw mechanism of producing neutrino masses).

## 2.3 Supersymmetry breaking in the MSSM

An operator called the *supertrace* treats bosonic and fermionic parts of a super multiplet differently. It is defined as

$$\mathrm{STr}\{M^2\} \equiv \sum_j (-1)^{2j+1} (2j+1)\, m_j^2 = 0, \tag{33}$$

where $j$ represents the 'spin' of the particles in some super multiplet. This is generic for tree level directly broken SUSY. Thus, we cannot break supersymmetry directly in the MSSM, since it preserves $\mathrm{STr}\{M^2\} = 0$. Applying this to the photon, say: $-3m_\gamma^2 + 2m_{\tilde\gamma}^2 = 0$, which would predict a massless photino that hasn't been observed. Applying it to up quarks: $2m_u^2 - m_{\tilde u_L}^2 - m_{\tilde u_R}^2 = 0$, thus one up squark must be *lighter* than the up quark, again this hasn't been observed. We introduce a *hidden* sector, which breaks SUSY and has its own fields (which do not directly interact with MSSM fields) and interactions, and an additional *messenger sector* to communicate the SUSY breaking to the observable sector fields:

$$\begin{pmatrix} \text{observable} \\ \text{sector, MSSM} \end{pmatrix} \longleftrightarrow \begin{pmatrix} \text{messenger -} \\ \text{sector} \end{pmatrix} \longleftrightarrow \begin{pmatrix} \text{hidden} \\ \text{sector} \end{pmatrix}.$$

This gets around the supertrace rule. There is typically an overall gauge group

$$\big(SU(3) \times SU(2) \times U(1)\big) \times G_{\text{SUSY}} \equiv G_{SM} \times G_{\text{SUSY}},$$

where the MSSM fields are singlets of $G_{\text{SUSY}}$ and the hidden sector fields are singlets of $G_{SM}$.

We have already seen several examples of SUSY breaking theories. One popular SUSY-breaking sector in the MSSM context is that of *gaugino condensation*: here, some asymptotically free gauge coupling $g$ becomes large at some energy scale $\Lambda$. $g$ will renormalise like Eq. 24 with some beta function coefficient. Solving the equation, with $g^{-2}(\Lambda) \to 0$, we obtain $\Lambda = M \exp[g^{-2}(M)/\beta]$. $M$ could be some large scale such as the string scale, $\sim 5 \times 10^{17}$ GeV. It is easy to arrange for $\Lambda \ll M$ because

---

[9]This leads us to a conjecture: *any experimental excess can be explained by RPV SUSY*. We have not found any counter-examples to this yet. This in turn leads to Butterworth's corollary: *RPV is the last refuge of the ambulance chasing scoundrel.*.

of the exponential suppression. When the gauge coupling becomes large, and the theory becomes non-perturbative, one can obtain $\langle \tilde{g}\tilde{g} \rangle \sim \mathcal{O}(\Lambda^3)$, breaking SUSY dynamically[10].

The SUSY breaking fields have couplings with the messenger sector, which in turn have couplings with the MSSM fields, and carry the SUSY breaking over to them. There are several possibilities for the messenger sector fields, which may determine the explicit form of SUSY breaking terms in the MSSM, including (note here that $M_{\text{SUSY}}$ is the SUSY breaking in the hidden sector, whereas $\Delta m$ is the SUSY breaking that ends up in the MSSM fields):

– gravity mediated SUSY

If the mediating field couples with gravitational strength to the standard model, the couplings are suppressed by the inverse Planck mass $M_{\text{pl}}$, the natural scale of gravity. The SUSY breaking mass splitting between MSSM particles and superparticles, $\Delta m$, becomes

$$\Delta m = \frac{M_{\text{SUSY}}^2}{M_{\text{pl}}}. \tag{34}$$

We want $\Delta m \approx 1 \text{TeV}$ and we know that $M_{\text{pl}} \approx 10^{19} \text{GeV}$, so

$$M_{\text{SUSY}} = \sqrt{\Delta m \cdot M_{\text{pl}}} \approx 10^{11} \text{GeV}. \tag{35}$$

The gravitino gets a mass $m_{\frac{3}{2}}$ of $\Delta m$ order TeV from the 'super Higgs mechanism'.
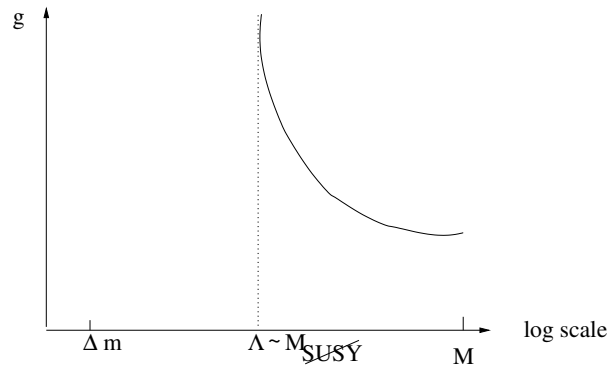


**Fig. 11:** Gaugino condensation and supergravity mediated SUSY breaking. From Ref. [1].

– gauge mediated SUSY

Messenger fields are charged under both $G_{SM}$ and $G_{\text{SUSY}}$. Gauge loops transmit SUSY breaking to the MSSM fields. Thus, $\Delta m \sim M_{\text{SUSY}}/(16\pi^2)$ is required to be of order TeV. In this case, the gravitino mass $m_{\frac{3}{2}} \sim \frac{M_{\text{SUSY}}^2}{M_{\text{pl}}} \sim \text{eV}$ and the gravitino is the LSP.

– anomaly mediated SUSY

In this case, the auxiliary fields of supergravity get a vacuum expectation value. The effects are always present, but suppressed by loop factors. They may be dominant if the tree-level contribution is suppressed for some reason.

Each of these scenarios has phenomenological advantages and disadvantages and solving their problems is an active field of research. In all scenarios, the Lagrangian for the observable sector has contributions

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{SUSY}}. \tag{36}$$

---

[10]Here, $\tilde{g}$ is the gaugino of the hidden sector gauge group, and $\beta$ is the hidden gauge group beta function coefficient.

In the second term, we write down all renormalisable symmetry invariant terms which do not reintroduce the hierarchy problem. They are of the form (where $i$ and $j$ label different fields):

$$\mathcal{L}_{\text{SUSY}} = \underbrace{m_{ij}^2 \, \varphi_i^* \, \varphi_j + m'^2_{ij}(\varphi_i\varphi_j + h.c.)}_{\text{scalar masses}} + \left( \underbrace{\frac{1}{2}M_\lambda\,\lambda\lambda}_{\text{gaugino masses}} + \underbrace{A_{ijk}\,\varphi_i\varphi_j\varphi_k}_{\text{trilinear couplings}} + h.c. \right). \qquad (37)$$

$M_\lambda, m'^2_{ij}, m^2_{ij}, A_{ijk}$ are called *soft SUSY breaking terms*: they do not reintroduce quadratic divergences into the theory. Particular forms of SUSY breaking mediation can give relations between the different soft SUSY breaking terms. They determine the amount by which supersymmetry is expected to be broken in the observable sector, and the masses of the superparticles for which the LHC is searching.
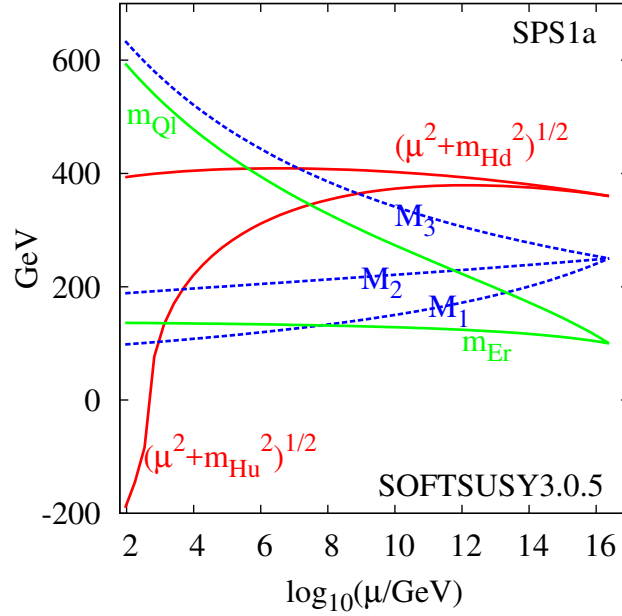


**Fig. 12:** An example of renormalisation in the MSSM using the program `SOFTSUSY` [4] to calculate the renormalisation. A particular high energy theory is assumed, which has GUT symmetry and implies that the gauginos are all mass degenerate at the GUT scale. The scalars (e.g the right-handed electron $Er$ and the left-handed squarks $Ql$) are also mass-degenerate at the GUT scale. Below the GUT scale though, the masses split and renormalise separately. When we are scattering at energies $\sim O(100)$ GeV, it is a good approximation to use the masses evaluated at that renormalisation scale $\mu \approx E$. We see that one of the Higgs mass squared parameters, $\mu^2 + M_{Hu}^2$, becomes negative at the electroweak scale, triggering electroweak symmetry breaking.

Explicitly, we parameterise all of the terms that softly break SUSY in the $R_p$ preserving MSSM, suppressing gauge indices:

$$\begin{aligned}
\mathcal{L}_{R_p}^{\text{SUSY}} &= (A_U)_{ij}\tilde{Q}_{Li}H_2\tilde{u}_{Rj}^* + (A_D)_{ij}\tilde{Q}_{Li}H_1\tilde{d}_{Rj}^* + (A_E)_{ij}\tilde{L}_{Li}H_1\tilde{e}_{Rj}^* + \\
&\quad \tilde{Q}_{Li}^*(m_{\tilde{Q}}^2)_{ij}\tilde{Q}_{Lj} + \tilde{L}_i^*(m_{\tilde{L}}^2)_{ij}\tilde{L}_j + \tilde{u}_{Ri}(m_{\tilde{U}}^2)_{ij}\tilde{u}_{Rj}^* + \tilde{d}_{Ri}(m_{\tilde{D}}^2)_{ij}\tilde{d}_{Rj}^* + \tilde{e}_{Ri}(m_{\tilde{E}}^2)_{ij}\tilde{e}_{Rj}^* + \\
&\quad (m_3^2H_1H_2 + h.c.) + m_1^2|H_1|^2 + m_2^2|H_2|^2 + \frac{1}{2}M_3\tilde{g}\tilde{g} + \frac{1}{2}M_2\tilde{W}\tilde{W} + \frac{1}{2}M_1\tilde{B}\tilde{B}.
\end{aligned}$$

Sometimes, $m_3^2$ is written as $\mu B$. Often, specific high scale models provide relations between these many parameters. For instance, the Constrained MSSM (which may come from some particular string theory

or other field theory) specifies the constraints

$$
\begin{aligned}
M_1 &= M_2 = M_3 =: M_{1/2} \\
m_{\tilde{Q}}^2 &= m_{\tilde{L}}^2 = m_{\tilde{U}}^2 = m_{\tilde{D}}^2 = m_{\tilde{E}}^2 \equiv m_0^2 I_3 \\
m_1^2 &= m_2^2 = m_0^2 \\
A_U &= A_0 Y_U, \ A_D = A_0 Y_D, \ A_E = A_0 Y_E
\end{aligned}
$$

where $I_3$ is the 3 by 3 identity matrix. Thus in the 'CMSSM', we reduce the large number of free SUSY breaking parameters down to[11] 3: $M_{1/2}$, $m_0$ and $A_0$. These relations hold at the GUT scale, and receive large quantum corrections, as Fig. 12 shows.

### 2.4 States after electroweak symmetry breaking

With two complex Higgs doublets, we count 8 real degrees of freedom. 3 of these are 'eaten' by the longitudinal components of the $W^\pm$ and $Z^0$ bosons, leaving a total of five physical Higgs fields: two $CP-$even (in mass order) $h^0, H^0$, one $CP-$odd $A^0$ and two charged Higgs' $H^\pm$. The other SUSY particles that have identical quantum numbers under QED×QCD mix after electroweak symmetry breaking: for example the bino, wino, and two neutral Higgsinos mix. Their mass eigenstates are called *neutralinos*, conventionally written in order of their masses $\chi_{1,2,3,4}^0$. $\chi_1^0$ typically has a special status in that is a good candidate for dark matter if it is the *lightest supersymmetric particle* and $R_p$ is conserved. The scalar partner of the left-handed top (called the 'left-handed stop') mixes with the right-handed stop to form two mass eigenstates: $\tilde{t}_{1,2}$. This analogously occurs for the sbottoms and staus as well. The charged Higgsinos mix with the winos to form mass eigenstates called 'charginos': $\chi_{1,2}^\pm$.

### 2.5 The Neutral Higgs Potential

Both Higgs' of the MSSM acquire vacuum expectation values:

$$
\begin{pmatrix} H_1^0 \\ H_1^- \end{pmatrix} \to \begin{pmatrix} v_1 \\ 0 \end{pmatrix}
\qquad
\begin{pmatrix} H_2^+ \\ H_2^0 \end{pmatrix} \to \begin{pmatrix} 0 \\ v_2 \end{pmatrix}
\tag{38}
$$

and to get the value of $M_W$ to match with experimental data, we require $v_{SM} = 246$ GeV. In a two-Higgs doublet model, this leads to the following construction:



$$\tan \beta = \frac{v_2}{v_1}.$$
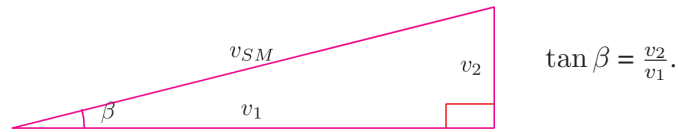
$\tan \beta$ is a parameter which changes the phenomenology of the model because the third family Yukawa couplings depend upon it, and they are comparatively large dimensionless couplings. The Yukawa terms from the MSSM superpotential are:

$$
\mathcal{L} = h_t \bar{t}_L H_2^0 t_R + h_b \bar{b}_L H_1^0 b_R + h_\tau \bar{\tau}_L H_1^0 \tau_R + \text{H.c.} + \ldots
\tag{39}
$$

$$
\Rightarrow \frac{m_t}{\sin \beta} = \frac{h_t v_{SM}}{\sqrt{2}}, \qquad \frac{m_{b,\tau}}{\cos \beta} = \frac{h_{b,\tau} v_{SM}}{\sqrt{2}},
\tag{40}
$$

after electroweak symmetry breaking and the neutral components of Higgs' are replaced by their vacuum expectation values: $H_i^0 = (v_i^0 + H_i^0)/\sqrt{2}$.

---

[11]One should really include $\tan \beta = v_2/v_1$ as well, the ratio of the two Higgs vacuum expectation values.
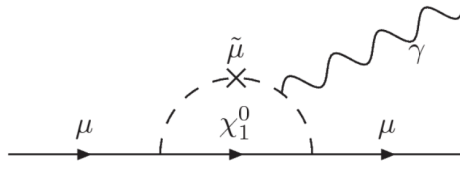
**Fig. 13:** Example one-loop diagram of sparticles contributing to $(g-2)_\mu$.

Picking out only the terms involving the neutral Higgs fields $H_1^0$ and $H_2^0$, we have the neutral Higgs potential

$$V = (|\mu|^2 + m_{H_2}^2)|H_2^0|^2 + (|\mu|^2 + m_{H_1}^2)|H_1^0|^2) - \mu B(H_2^0 H_1^0 + \text{H.c.}) + \frac{1}{8}(g^2 + g'^2)(|H_2^0|^2 - |H_1^0|^2)^2. \tag{41}$$

The vacuum minimises this potential with respect to both of the neutral components:

$$\frac{\partial V}{\partial H_2^0} = \frac{\partial V}{\partial H_1^0} = 0 \Rightarrow \mu B = \frac{\sin 2\beta}{2}(\bar{m}_{H_1}^2 + \bar{m}_{H_2}^2 + 2\mu^2), \mu^2 = \frac{\bar{m}_{H_1}^2 - \bar{m}_{H_2}^2 \tan^2\beta}{\tan^2\beta - 1} - \frac{M_Z^2}{2}. \tag{42}$$

These two conditions should be used to eliminate two of the MSSM's free parameters: often, $|\mu|$ and $B$ (although note that the sign of $\mu$ is physical and not determined by Eq. 42).

## 2.6 Pros and Cons of the MSSM

We start with a list of unattractive features of the MSSM:

- There are $\sim 100$ extra free parameters in the SUSY breaking sector, making for a complicated parameter space.
- Nearly all of this parameter space is ruled out by flavour physics constraints: SUSY particles could heavily mix in general, then this mixing could appear in loops and make the quarks mix in a flavour changing neutral current, upon which there are very strong experimental bounds. It could be that this clue is merely telling us that there is more structure to the MSSM parameter space, though (like in the CMSSM).
- The $\mu$ problem. $\mu$ in $W_{R_p}$ must be $< \mathcal{O}(1)$ TeV, since it contributes at tree-level to $m_h$. Why should this be, when in principle we could put it to be $\sim \mathcal{O}(M_{Pl})$, because it does not break any SM symmetries? (Note though that once it is set to be small at tree-level, SUSY protects it from large quantum corrections).
- As lower limits on sparticle masses increase, the extent to which SUSY solves the hierarchy problem decreases.

These SUSY problems can be solved with further model building.

We close with an ordered list of weak-scale SUSY's successes:

- SUSY solves the technical hierarchy problem.
- Gauge unification works.
- The MSSM contains a viable dark matter candidate, if $R_p$ is conserved.
- Electroweak symmetry breaks radiatively.
- A one-loop diagram involving sneutrinos and charginos (and one involving smuons and neutralinos: see Fig. 13) contribute to the anomalous magnetic moment of the muon, and may solve the discrepancy between SM predictions and experimental measurements in Eq. 12.
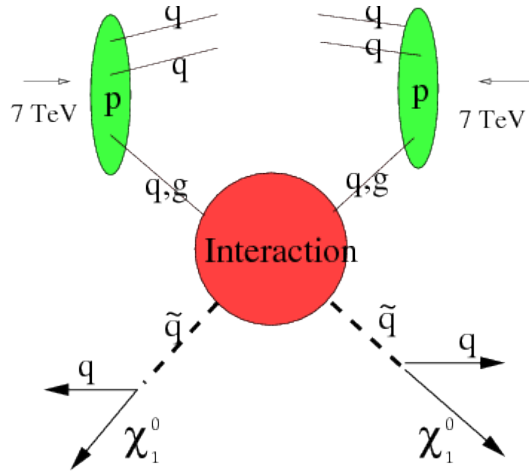
**Fig. 14:** Picture of the production of sparticles at a 14 TeV LHC.

## 2.7 LHC Production of SUSY Particles

One turns the energy of the LHC beams into mass via $E = mc^2$, hoping to produce pairs (if $R_p$ is conserved) of SUSY particles that were too heavy to have been previously produced in lower energy machines. We show a schematic in Fig. 14: occasionally, high energy constituents of the proton (called 'partons': quarks or gluons) will collide, as in the figure. The idea is that these are most likely to make strongly interacting particles, all other things being equal (in the figure, we have the example of squark production). The rest of the broken protons typically will be boosted along the beam-line. The sparticles undergo subsequent decay (in the example in the figure, into a quark - which will form a jet of hadrons and a dark matter particle: the lightest neutralino). Since we have assumed $R_p$ to be conserved, the $\chi_1^0$ is stable but since it is weakly interacting, it passes through the rest of the detector without any interactions, stealing momentum from the collision. The decays of the initial pair of sparticles may be much more complex, going through cascade decays where at each stage there is a lighter sparticle and a SM particle produced. $R_p$ conserving SUSY provides an example of how any dark matter candidate that is light enough and that (perhaps indirectly) couples to protons can be produced in LHC collisions. Jets and missing transverse momentum $\vec{p}_T^{\text{miss}}$ (sometimes this is known under the misnomer 'missing energy') form a classic SUSY search, but also jets plus varying numbers of leptons (from sparticle cascade decays) plus missing transverse momentum form another well-studied class. There is a SUSY monojet signature [5], although sparticles would likely be found in one of the other production channels first because the monojet signature is due to a strong times an electroweak matrix element. In the case of gauge mediated SUSY breaking models, the lightest neutralino may decay into a gravitino plus a photon, or a $Z^0$, and so for instance di-photon plus missing transverse momentum searches form another class. Since one obtains additional jets from showering off the initial state at the LHC, searches are often inclusive, meaning that one only selects a *minimum* number of hard jets.

Often, searches are interpreted in terms of 'simplified models': for instance, one studies gluino pair production, then assumes that each decays into 2 jets and missing transverse momentum: see Fig. 15. However, current bounds based on simplified models [7] often give much stronger bounds than in a more general MSSM set-up [8]. This is because simplified models tend to only assume a single decay mode of one sparticle (or a few decay modes of particular sparticles), whereas in full models there can be literally thousands of active decay chains, diluting the signal between many different search channels such that no one shows an excess. There are also cases of somewhat 'compressed spectra': when sparticles in decay chains are similar in mass, energy-momentum conservation means that they tend to produce fairly soft SM particles, which often fail analysis cuts. Because they are not dependent on the many MSSM
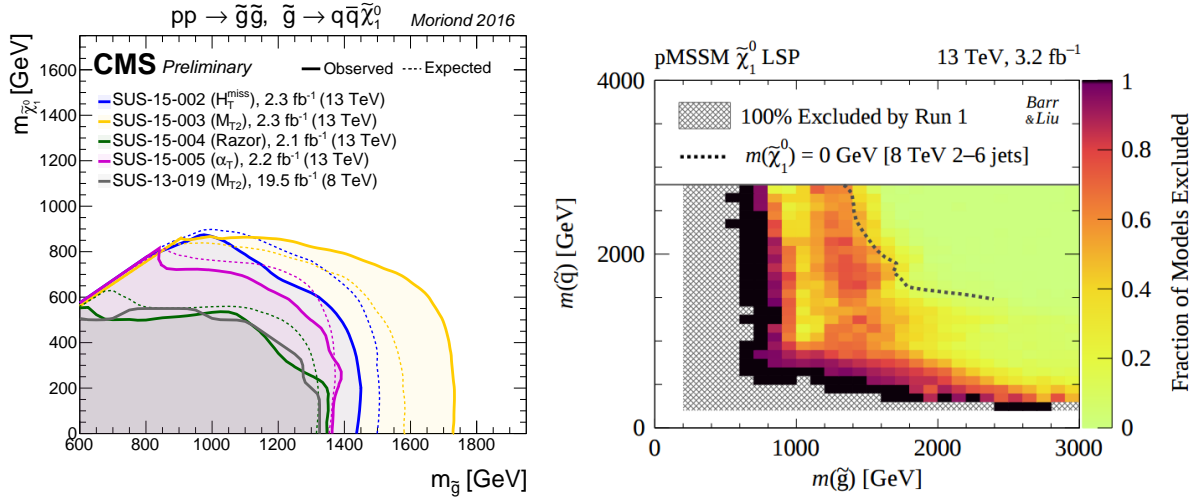
**Fig. 15:** Examples of interpretation of search limits: the left-hand panel shows simplified model exclusions from $2.3\text{fb}^{-1}$ of 13 TeV LHC analyses for di-gluino production and is from Ref. [6]. On the right-hand side, we see a more complete description in terms of the MSSM interpreting 3.2 $\text{fb}^{-1}$ of integrated luminosity and is from Ref. [9]. The simplified model exclusion on the left hand panel is that gluinos up to 1750 GeV are excluded (for zero neutralino mass) whereas on the right-hand panel, we see that models exist where gluinos of 800 GeV are allowed: these pass a list of negative searches for SUSY at 13 TeV in ATLAS. The 'fraction of models excluded' starts from a list of models in MSSM parameter space that had good dark matter properties, and otherwise passed the constraints from Run I. If there are *any* points with the masses listed on the axis, the fraction of models excluded is less than 1.

parameters, simplified searches are very convenient for searches, being less model dependent. However, exclusion limits from simplified models are not easy to interpret in more realistic models, and tend to be far too restrictive unless one interprets them with care. In Fig. 15, we see this in action: for massless neutralinos, gluinos up to 1750 GeV are ruled out in the simplified model, whereas in a (more realistic) phenomenological MSSM approximation, we see that gluinos of 800 GeV are still allowed for some points.

## 3 Extra Dimensions

For a review of extra dimensions and their phenomenology, see Ref. [10]. As mentioned above, extra dimensions correspond to an expansion of the Poincaré symmetry: there are additional generators associated with translation invariance in each extra spatial dimension. Superstring theory also requires them in addition to supersymmetry for internal consistency, but any theory incorporating them must explain why we only observe 3+1 (i.e.three space-like and one time-like). There are a couple of possibilities to 'hide' the extra dimensions from our perception:

- *We are stuck on a brane*: meaning that the bulk of space-time has more than 3+1 dimensions, but SM fields are stuck on a 3+1 dimensional hypersurface: a 'brane'. Gravity travels wherever space-time is, so that it *must* feel the effect of the additional dimensions. That's because gravity is a described by a quantum fluctuation of the metric, and the bulk metric is defined in the bulk space-time.

- *The extra dimensions are curled up on themselves*: each point in our 3+1 dimensional space time has a circle, or some other compact manifold, where one can travel – albeit periodically – in the extra dimensions, which are in an orthogonal direction to all of the other dimensions. If such
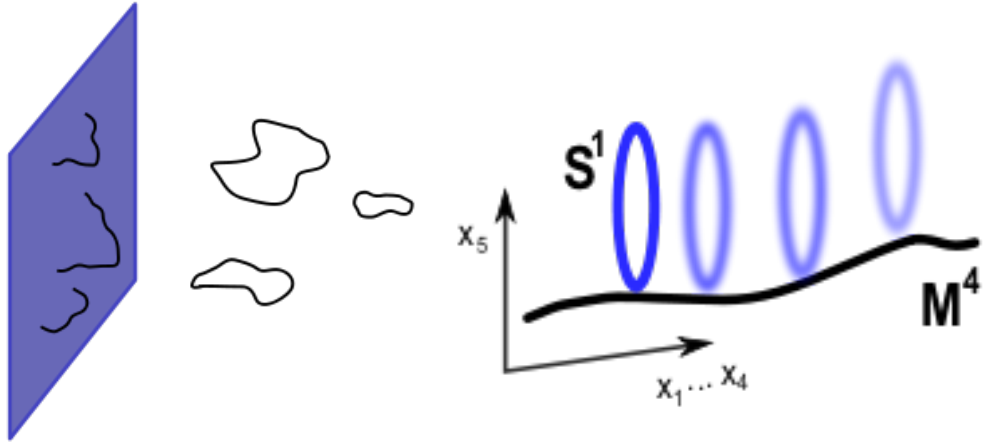
**Fig. 16:** Picture of different extra-dimensional set-ups: the brane (on the left), where in string theory SM states appear as open strings whose ends end upon the brane but gravitons appear as close string states in the bulk, or compactification (on the right), in this example we have taken the example of a circle $S^1$ times ordinary 4-dimensional Minkowski space $M^4$.

> manifolds are not too large (less than a millimeter, certainly), then current experimental bounds upon gravitational forces acting at relatively small distances may still not rule the model out.

We illustrate the two cases in Fig. 16. In the figure, we have taken the example of string theory to illustrate the brane case, but it is essentially valid in the field theory limit as well: SM fields may be confined to a hypersurface of the bulk space-time, whereas gravity travels everywhere.

### 3.1 Compactification and a Scalar Field in 5 Dimensions

Taking compactified extra dimensions as an example, consider a massless five dimensional (5D) scalar field (i.e. a scalar field living in a 5-dimensional bulk space-time) $\varphi(x^M)$, $M = 0, 1, \ldots, 4$ with action

$$\mathcal{S}_{5D} = \int d^5 x \partial^M \varphi \, \partial_M \varphi. \tag{43}$$

We single the extra dimension out by calling it $x^4 = y$. $y$ defines a circle of radius $r$ with $y \equiv y + 2\pi r$. Our space time is now $\mathbb{M}_4 \times S^1$. Periodicity in the $y$ direction implies that we may perform a discrete Fourier expansion

$$\varphi(x^\mu, y) = \sum_{n=-\infty}^{\infty} \varphi_n(x^\mu) \exp\left(\frac{iny}{r}\right). \tag{44}$$

Notice that the Fourier coefficients are functions of the standard 4D coordinates and therefore are (an infinite number of) 4D scalar fields. The equations of motion for the Fourier modes are the (in general massive) Klein-Gordon wave equations

$$\partial^M \partial_M \varphi = 0 \Rightarrow \sum_{n=-\infty}^{\infty} \left(\partial^\mu \partial_\mu - \frac{n^2}{r^2}\right) \varphi_n(x^\mu) \exp\left(\frac{iny}{r}\right) = 0$$

$$\Longrightarrow \boxed{\partial^\mu \partial_\mu \varphi_n(x^\mu) - \frac{n^2}{r^2} \varphi_n(x^\mu) = 0.} \tag{45}$$

These are then an infinite number of Klein Gordon equations for massive 4D fields. This means that each Fourier mode $\varphi_n$ is a 4D particle with mass $m_n^2 = \frac{n^2}{r^2}$. Only the zero mode ($n = 0$) is massless. One can visualise the states as an infinite tower of massive states (with increasing mass proportional to $n$). This is called a *Kaluza Klein tower* and the massive states ($n \neq 0$) are called *Kaluza Klein-states* or *momentum states*, since they come from the momentum in the extra dimension:
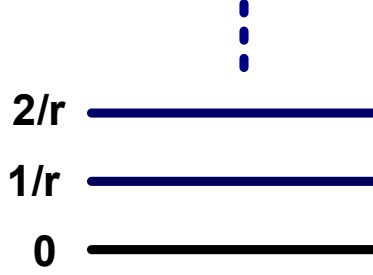


**Fig. 17:** The Kaluza Klein tower of massive states due to an extra $S^1$ dimension. Masses $m_n = |n|/r$ grow linearly with the fifth dimension's wave number $n \in \mathbb{Z}$.

In order to obtain the effective action in 4D for all these particles, let us plug the mode expansion of $\varphi$ Eq. 44 into the original 5D action Eq. 43:

$$
\begin{aligned}
\mathcal{S}_{5D} &= \int d^4x \int dy \sum_{n=-\infty}^{\infty} \left( \partial^\mu \varphi_n(x^\mu)\, \partial_\mu \varphi_n(x^\mu)^* - \frac{n^2}{r^2}|\varphi_n|^2 \right) \\
&= 2\pi r \int d^4x \left( \partial^\mu \varphi_0(x^\mu)\, \partial_\mu \varphi_0(x^\mu)^* + \ldots \right) = 2\pi r \mathcal{S}_{4D} + \ldots
\end{aligned}
$$

This means that the 5D action reduces to one 4D action for a massless scalar field plus an infinite sum of massive scalar actions in 4D. If we are only interested in energies smaller than the $\frac{1}{r}$ scale, we may concentrate only on the action of the massless mode.

### 3.2 Compactification of a Vector Field in 5 Dimensions

Vector fields are decomposed in a completely analogous way: $\{A_M\} = \{A_\mu, A_4 = \phi\}$. Consider the action

$$
\mathcal{S}_{5D} = \int d^5x \frac{1}{g_{5D}^2} F_{MN} F^{MN} \tag{46}
$$

with a field strength

$$
F_{MN} = \partial_M A_N - \partial_N A_M \tag{47}
$$

implying

$$
\partial^M \partial_M A_N - \partial^M \partial_N A_M = 0. \tag{48}
$$

If we now choose a gauge, e.g. the transverse gauge:

$$
\partial^M A_M = 0, A_0 = 0 \Rightarrow \partial^M \partial_M A_N = 0, \tag{49}
$$

then this obviously becomes equivalent to the scalar field case (for each component $A_M$) indicating an infinite tower of massive states for each massless state in 5D. In order to find the 4D effective action we

once again plug this into the 5D action:

$$
\begin{aligned}
\mathcal{S}_{5\mathrm{D}} &\mapsto \mathcal{S}_{4\mathrm{D}} \\
&= \int d^4x \left( \frac{2\pi r}{g_{5\mathrm{D}}^2} F_{(0)}{}^{\mu\nu} F_{(0)\mu\nu} + \frac{2\pi r}{g_{5\mathrm{D}}^2} \partial_\mu \rho_0 \partial^\mu \rho_0 + \dots \right).
\end{aligned}
$$

Therefore we end up with a 4D theory of a massless gauge particle $F_{(0)}^{\mu\nu}$, a massless scalar $\rho_0$ from the massless Kaluza-Klein state of $\phi$ and infinite towers of massive vector and scalar fields. Notice that the gauge couplings of 4- and 5 dimensional actions (coefficients of $F_{MN}F^{MN}$ and $F_{\mu\nu}F^{\mu\nu}$) are related by

$$
\frac{1}{g_4^2} = \frac{2\pi r}{g_5^2}. \tag{50}
$$

In $D$ space time dimensions, this generalises to

$$
\frac{1}{g_4^2} = \frac{V_{D-4}}{g_D^2} \tag{51}
$$

where $V_n$ is the volume of the $n$ dimensional compact space (e.g. an $n$ sphere of radius $r$).

### 3.2.1 The electric (and gravitational) potential

We apply Gauss' law for the electric field $\vec{E}$ and the potential $\Phi$ of a point charge $Q$:

$$
\oint_{S^2} \vec{E} \cdot d\vec{S} = Q \Rightarrow \|\vec{E}\| \propto \frac{1}{R^2}, \Phi \propto \frac{1}{R} \; : \; 4D
$$

$$
\oint_{S^3} \vec{E} \cdot d\vec{S} = Q \Rightarrow \|\vec{E}\| \propto \frac{1}{R^3}, \Phi \propto \frac{1}{R^2} \; : \; 5D
$$

Thus, the apparent behaviour of the force depends upon whether we are sensitive to the extra dimension or not: if we test the force at distances smaller than its size (i.e. at energies high enough to probe such small distance scales), it falls off as $1/R^3$: the field lines have an extra dimension to travel in. If we test the force at larger distances than the size of the extra dimension, we obtain the usual $1/R^2$ law.

In $D$ space time dimensions

$$
\|\vec{E}\| \propto \frac{1}{R^{D-2}}, \Phi \propto \frac{1}{R^{D-3}}. \tag{52}
$$

If one dimension is compactified (radius $r$) like in $\mathbb{M}_4 \times S^1$, then we have two limits

$$
\|\vec{E}\| \propto \begin{cases} \dfrac{1}{R^3} & : R < r \\ \dfrac{1}{R^2} & : R \gg r \end{cases}. \tag{53}
$$

Analogous arguments hold for gravitational fields and their potentials, but we shall not detail them here, preferring instead to sketch the resulting field content.

### 3.2.2 Sketch of Compactified Gravitation

The spin $2\hbar$ graviton $G_{MN}$ becomes the 4D graviton $g_{\mu\nu}$, some gravivectors $G_{\mu n}$ and some graviscalars $G_{mn}$ (where $m, n = 4, \dots, D-1$), along with their infinite Kaluza-Klein towers. The Planck mass squared $M_{Pl}^2 = M_D^{D-2} V_{D-4} \sim M_D^{D-2} r^{D-4}$ is a derived quantity. Fixing $D$, we can fix $M_D$ and $r$ to get the correct result for $M_{Pl} \sim 10^{19}$ GeV. So far, we require $M_D > 1$ TeV and $r < 10^{-16}$cm from Standard Model measurements since no significant confirmed signature of extra dimensions has been seen at the time of writing.
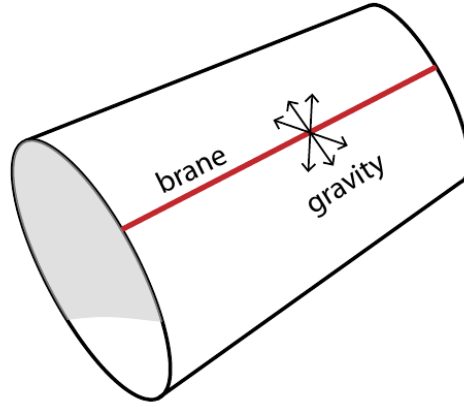
**Fig. 18:** Force field lines feel the effect of the extra dimensions. Here we show a 3+1 dimensional brane, where gravity spreads into the extra dimension and feels its effect.

### 3.3 Brane Worlds

In the brane world scenario, we are trapped on a 3+1 surface in a $D+1$ dimensional bulk space-time (see Fig. 18). There are two cases here: *large extra dimensions* and warped space-times. Since gravity itself is so weak, the constraints on brane world scenarios are quite weak: the extra dimension is constrained to be of a size $r < 0.1$ mm or so, potentially much larger than the $10^{-16}$ cm of the Standard Model, hence the name *large* extra dimensions.

#### 3.3.1 *Large extra dimensions*

There is the possibility to try to solve the hierarchy problem with the large extra dimensions scenario if we put $M_D \sim 1$ TeV. The idea is that this *is* the fundamental scale: there is no high scale associated with $M_{Pl}$ fundamentally - it is an illusion caused by the presence of the extra dimensions. In 5D for example, $M_{Pl^2} = M_D^{D-2} V_{D-4} \Rightarrow r \sim 10^8$ km, clearly ruled out by observations. Already in 6D though, $r = 0.1$ mm - consistent with experiments that measure the gravitational force on small distance scales. This rephrases the hierarchy problem to the question "why are the extra dimensions so large compared with $10^{-16}$ cm?"

Graviton phenomenology: each Kaluza-Klein mode couples weakly $\propto 1/M_{Pl}$, but there are so many modes that after summing over them, you end up with $1/M_D$ suppression only! One can approximate them by a *continuum* of modes with a cut-off. The graviton tower propagates into the bulk and takes away missing momentum leading to a $pp \to j + \vec{p}_T^{\text{miss}}$ signature (for example) by the process shown in Fig. 19.

#### 3.3.2 *Warped (or 'Randall-Sundrum' space-times*

Warped space-times are where the metric exponentially warps along the extra dimension $y$:

$$ds^2 = e^{-|ky|}\eta_{\mu\nu}dx^\mu dx^\nu + dy^2. \tag{54}$$

The metric changes from $y = 0$ to $y = \pi r$ via $\eta_{\mu\nu} \mapsto e^{-k\pi r}\eta_{\mu\nu}$. Here, we set $M_D = M_{Pl}$, but this gets warped down to the weak brane:

$$\Lambda_\pi \sim M_{Pl}e^{-k\pi r} \sim \mathcal{O}(\text{TeV}), \tag{55}$$

if $r \sim 10/k$. Here, $k$ is of order $M_{Pl}$ and so we have a *small extra dimension*, but the warping explains the smallness of the weak scale. Note that we still have to stabilise the separation between the branes, which can involve extra tuning unless extra structure is added to the model.
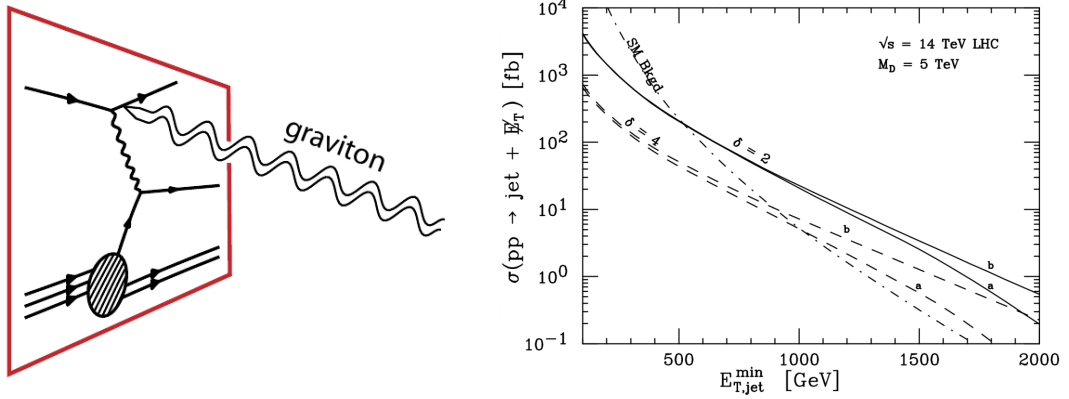
**Fig. 19:** Example production of $\vec{p}_T^{\text{miss}}$ at a 14 TeV LHC through production of a Kaluza-Klein tower of graviton states that propagate off into the bulk. The left-hand panel shows a heuristic picture: the red hypersurface representing the brane, and the graviton tower being emitted into the bulk. In the right-hand panel, it is shown how the cross-section varies with the transverse momentum of the jet ($E_{T,jet}^{min}$) for the SM background, and the case of $d$ extra dimensions. The lines a (b) are constructed by integrating the cross-section over $\hat{s} < M_D^2$ (all $\hat{s}$), respectively. Both the diagram and the plot are from Ref. [11].
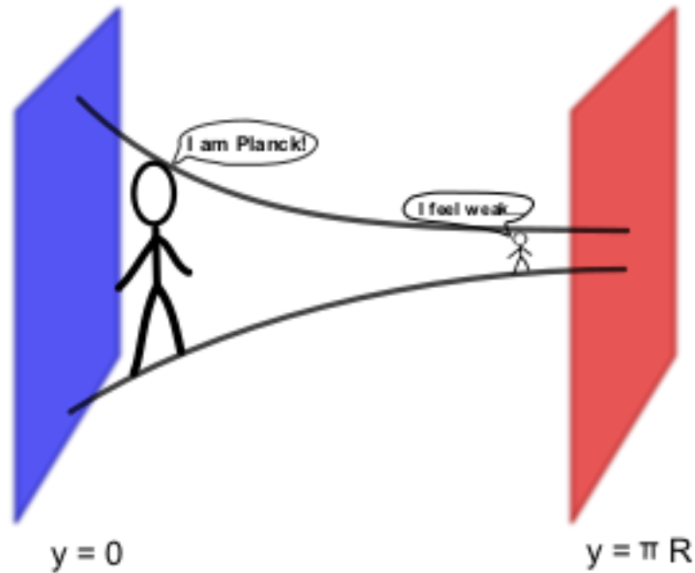


**Fig. 20:** Picture of the Randall-Sundrum I set-up. On the left-hand side at $y = 0$ we have the Planck brane, which is warped down to the weak brane at the right hand side ($y = \pi R$). The idea is that the Higgs boson (and some other fields) are localised on the weak brane.
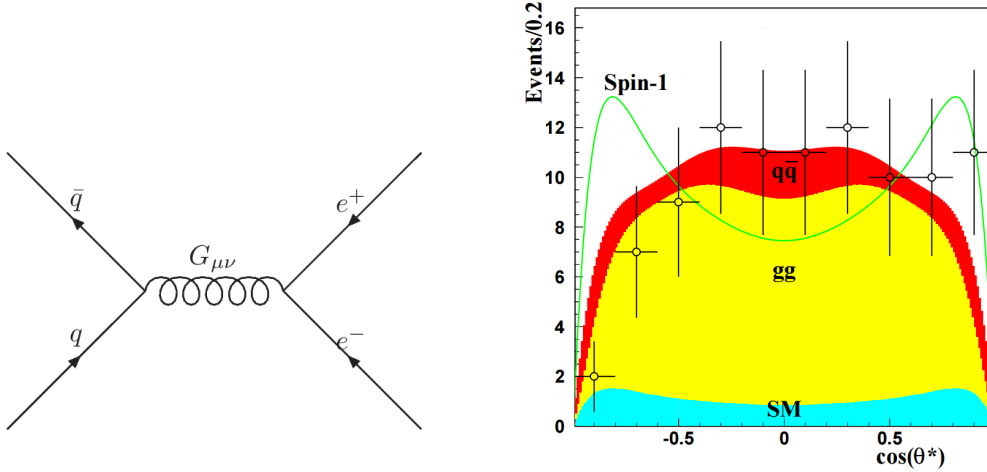
**Fig. 21:** Production of an RS graviton and subsequent decay into $e^+e^-$: both figures are taken from Ref. [12]. On the right-hand side, the angular distribution of the electron with respect to the beam line in the rest-frame ($\theta^*$) of $G_{\mu\nu}$ is shown for a 1.5 TeV graviton. The different shaded colours show the contribution from $q\bar{q}$ collisions, from $gg$ and from other SM processes. This is contrasted against a straw-man spin $1\hbar$ distribution in the green line. The typical expected size of statistical uncertainties resulting from 100 fb$^{-1}$ of integrated luminosity at a 14 TeV LHC is shown on the points. This would be enough to discriminate against the spin $1\hbar$ hypothesis, which is much more forward than the spin $2\hbar$ hypothesis.

The interaction Lagrangian is

$$\mathcal{L}_I = -G^{\mu\nu}T_{\mu\nu}/\Lambda_\pi, \tag{56}$$

where $T_{\mu\nu}$ is the *stress energy tensor*, containing products of the other Standard Model fields. $\Lambda_\pi \sim \mathcal{O}(\text{TeV})$, so the interaction leads to electroweak-strength cross sections, not gravitationally suppressed ones. Thus, the LHC can produce the resonance: one will tend to produce the lightest one most often, as it is less suppressed by parton distribution functions. The ratios of masses of higher modes are given by zeros of Bessell functions, so they are not as regular as they are in large extra dimensions.

Randall-Sundrum phenomenology: one looks for the TeV scale first resonances, which are weakly coupled to Standard Model states. If only gravity travels in the extra dimensions, then the resonance is the 'Randall-Sundrum graviton': it has universal coupling to all particles via Eq. 56 and so it can decay into $q\bar{q}, WW, ZZ, \gamma\gamma, gg, l^+l^-$ or $h^0h^0$ with branching ratios that are of a similar order of magnitude to each other. Flavour considerations imply that this isn't the end of the story: one requires additional flavour structure, otherwise the model violates flavour bounds from experiment. One common way of adding flavour structure is to allow the other particles into the bulk, but have different profiles of fermions in the bulk, leading to different overlaps with the weak brane, where the Higgs field is localised (the overlap would be proportional to the particle in question's Yukawa coupling). In this case, one could look for the first Kaluza Klein modes of gauge bosons and fermions, too.

Kaluza Klein modes that have masses that are heavier than the centre of mass energy of the beams may also be looked for via their virtual effects. Searching for particles that mediate interactions that are occurring at collisions with less energy than their mass has been historically very important (particularly in terms of the weak interactions which were indirectly observed before the discovery of the $W^\pm$ and $Z^0$ bosons). Such a kinematic situation can be approximated by *effective field theories*, which in turn reduces model dependence. We now sketch effective field theories, along with caveats pertinent to their use.

## 4 Effective Field Theories

At low momenta $p^\mu$, we can model the effects of particles with a much heavier mass $M^2 \gg p^2$ and a small width $\Gamma \ll M$ with *effective field theory*. This squeezes a propagator down to a point:

$$\lim_{p^2/M^2 \to 0, \Gamma/M \to 0} \frac{1}{p^2 - M^2 + iM\Gamma} \approx -\frac{1}{M^2}, \tag{57}$$

in a fairly model independent way. Thus, for example $W$ boson couplings like

$$\mathcal{L} = -\frac{g}{2\sqrt{2}}\bar{e}\gamma^\rho(1-\gamma_5)W_\rho\nu_e - \frac{g}{2\sqrt{2}}\bar{\nu}_\mu\gamma^\rho(1-\gamma_5)W_\rho\mu \tag{58}$$

become

$$\mathcal{L} \approx -\frac{G_F}{\sqrt{2}}\left(\bar{e}\gamma^\rho(1-\gamma_5)\nu_e\right)\left(\bar{\nu}_\mu\gamma^\rho(1-\gamma_5)\mu\right), \tag{59}$$

where $G_F = \sqrt{2}g^2/(8M_W^2)$. One has to be careful at the LHC with the range of validity of the effective field theory, however, because the LHC has a large centre of mass energy. If some of the collisions have $p^2 \geq M^2$, then for those collisions the effective field theory is a bad approximation: there, one becomes sensitive to the full structure of the propagator. Effective field theory methods can be useful for parameterising searches for new physics at low momentum: these four-fermion operators are often called *contact operators*, e.g. for some fermionic dark matter particle $\chi$,

$$\mathcal{L} = \frac{\lambda^2}{M^2}(\bar{q}\gamma^\mu q)(\chi\gamma_\mu\chi) \tag{60}$$

for some coupling strength $\lambda$ [18]. However, for dark matter production at the LHC (e.g. in the monojet channel), the energies are often higher than the messenger mass and so a more precise (simplified?) model is needed [19]. Such a move to more specified models increases model dependence, but may be necessary if one requires a large régime of validity for one's description of high energy collisions.

## 5 Conclusion

At the time of writing, 13 TeV collisions at the LHC have yet to yield direct, unambiguous and confirmed discoveries of new physics. In some channels, around 36 fb$^{-1}$ of integrated luminosity has been analysed in each general purpose experiment. However, there is plenty of room for new physics to be hiding: in more data or in other analyses. I personally and perhaps naively expect some signal to show up in the first 100 fb$^{-1}$ of Run II data. Certainly it seems unlikely that if there are no excesses at all in that amount of data (in some channel), there is unlikely to be a $5\sigma$ discovery at Run II in the same channel. If CERN increases the beam energy, for example from 13 TeV to 14 TeV, the search sensitivity gains a sudden boost, and indeed this will be interesting in Run III or beyond. There is a plan (the 'high-energy' LHC, or HE-LHC) to increase the beam energy to around 27 TeV with new magnets. This would lead to a large increase in sensitivity.

On the other hand, there are several interesting excesses in $B$ physics measurements as compared to SM measurements, which we have not explicitly discussed in these Beyond the Standard Model lectures. Probably the theoretically cleanest of these are those of $R_{K(*)}$ as shown in Table 1. from the LHCb experiment [13, 14]. Large theoretical uncertainties associated with mesonic physics cancel well in such a ratio, particularly when one is probing final states involving leptons. In the SM, $R_K$ is a firm prediction from diagrams like Fig. 22, and so the measurements in Table 1 indicate non-SM lepton flavour non-universality at the $4\sigma$ level. In fact, a fit to this and other data indicates that a new physics effective field theory operator on top of the SM

$$\mathcal{L} = C_9^{(\mu)}(\bar{s}_L\gamma^\mu b_L)(\bar{\mu}_L\gamma_\mu\mu_L) + \dots \tag{61}$$
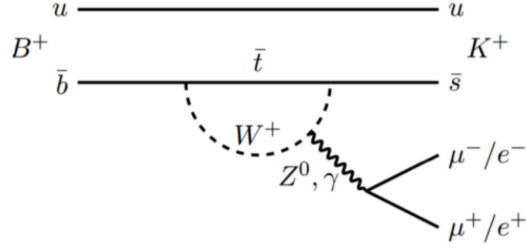
**Fig. 22:** Example Feynman diagram contributing to $B$ meson decays that form the variable $R_K$.

|  | $q^2/\text{GeV}^2$ | SM | LHCb 3 fb | $\sigma$ |
|---|---|---|---|---|
| $R_K$ | $[1, 6]$ | $1.00 \pm 0.01$ | $\mathbf{0.745^{+0.090}_{-0.074}}$ | 2.6 |
| $R_{K^*}$ | $[0.045, 1.1]$ | $0.91 \pm 0.03$ | $\mathbf{0.66^{+0.11}_{-0.07}}$ | 2.2 |
| $R_{K^*}$ | $[1.1, 6]$ | $1.00 \pm 0.01$ | $\mathbf{0.69^{+0.11}_{-0.07}}$ | 2.5 |

**Table 1:** Predictions and measurements of $R_{K^{(*)}} \equiv \frac{BR(B^\pm \to K^{(*)\pm}\mu^+\mu^-)}{BR(B^\pm \to K^{(*)\pm}e^+e^-)}$ [13] in different bins of momentum transfer squared. The uncertainty on the SM prediction includes estimated theoretical uncertainties.
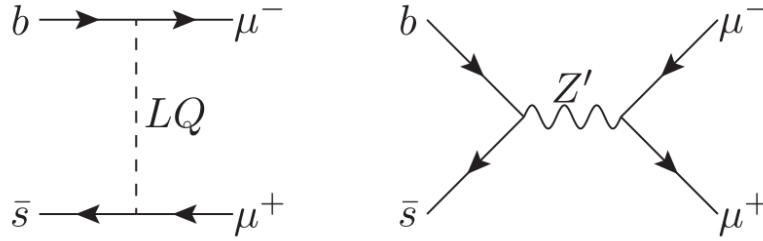


**Fig. 23:** Feynman diagrams showing different possibilities for BSM particles of mass $M \gg m_B$ which lead to the effective operator in Eq. 61 at tree-level. The additional particles lead to a change in $C_9^{(\mu)}$ proportional to $-\lambda_1 \lambda_2/M^2$, where $\lambda_{1,2}$ are dimensionless couplings of the respective particles. This combination of couplings and masses are then fixed to predict the central value of the experimental measurements of $B-$data.

is preferred to be non-zero at the $4.3\sigma$ level [15–17]. A BSM operator proportional to $(\bar{s}_L \gamma^\mu b_L)(\bar{\mu}\gamma_\mu\mu)$ (i.e. a vector-like coupling to muons, rather than a left-handed coupling to them) also works approximately as well. At the tree-level, these operators can be caused by a couple of different BSM particles: leptoquarks or flavourful $Z'$s, as depicted in Fig. 23.

The leptoquark can either be a scalar triplet $S_3$ of $SU(2)_L$ or a vector particle: either an $SU(2)_L$ triplet, or singlet. Leptoquarks couple (by definition) to a lepton and a quark: in order to preserve QCD they must therefore be coloured. Hadron collider and other searches then focus on pair production of them, e.g. by the process $gg \to S_3\bar{S}_3 \to (\mu^+ b)(\mu^- \bar{s})$, where the bracketed particles should form a resonance and have a bump in their invariant mass spectra. In the case of $Z'$ particles, the diagram in Fig. 23 leads to resonant production of $Z'$, since the initial $b$ quark can be obtained from an initial proton from a gluon splitting into a $b\bar{b}$ pair.

We show the projected sensitivity of a future 100 TeV $pp$ collider to such particles that explain the errant $B$ decays in Fig. 24. In the left-hand plot, we see that a 27 TeV energy upgraded LHC option covers a small portion of the $Z'$ parameter space, whereas the 100 TeV option can essentially cover all
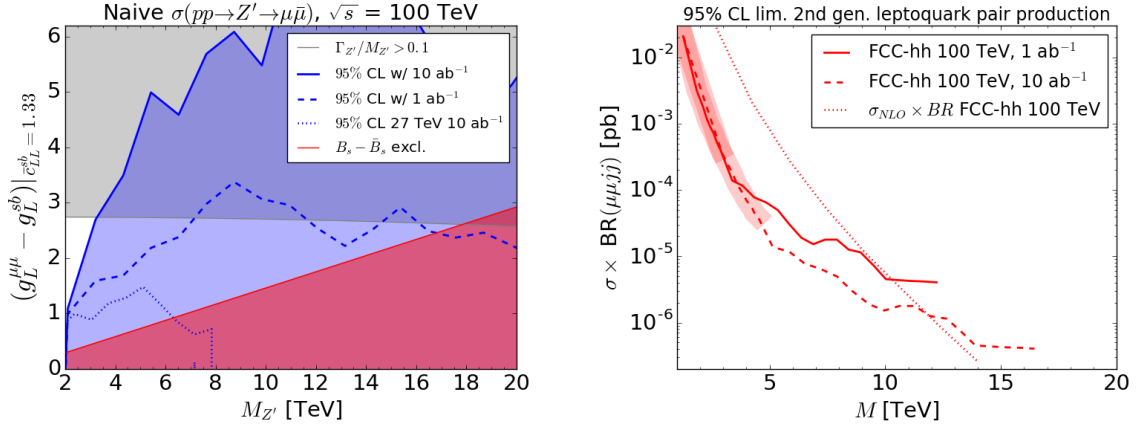
**Fig. 24:** 100 TeV Future Circular Collider (FCC) reach for flavourful $Z'$s (left panel) and leptoquarks that explain the $R_{K^{(*)}}$ data (right panel). The red area in the left-hand plot is excluded by measurements of $B_s - \bar{B}_s$ mixing, whereas the blue area is the 95% confidence level (CL) projected sensitivity. The grey area shows wide $Z'$ particles, where perturbativity is being lost. In the right-hand panel, the region above each curve is covered at 95%, whereas the production cross-section is shown by the dotted curve. Figures from Ref. [20].

of the allowed perturbative parameter space. In the right-hand plot, we deduce that leptoquarks with a mass up to 12 TeV can be covered by a FCC pair-production search. Leptoquarks up to 40 TeV in mass can explain the $B-$data whilst still satisfying other constraints[12]. However, since it is the particular combination $\lambda_1\lambda_2/M^2$ is fixed by the $B-$data, $M$ and $\lambda_{1,2}$ can all be much smaller. Searches at the LHC, HL-LHC and HE-LHC are therefore of high priority (this also goes for $Z'$s). LHCb is expected to announce further measurements of the quantities in Table 1 in 2019, with a roughly similar-size and independent data set.

We close with a quote from William Blake[13] from *The Marriage of Heaven and Hell*:

> "The road of excess leads to the palace of wisdom".

## Acknowledgements

## References

[1] F. Quevedo, S. Krippendorf and O. Schlotterer, arXiv:1011.1491 [hep-th].

[2] K.A.Olive *et al* (Particle Data Group), Chin. Phys. **C38** (2014) 090001.

[3] S. P. Martin, Adv. Ser. Direct. High Energy Phys. **21** (2010) 1 [Adv. Ser. Direct. High Energy Phys. **18** (1998) 1] doi:10.1142/9789812839657_0001, 10.1142/9789814307505_0001 [hep-ph/9709356].

[4] B. C. Allanach, Comput. Phys. Commun. **143** (2002) 305 doi:10.1016/S0010-4655(01)00460-X [hep-ph/0104145].

---

[12]One can also search for single leptoquark production. This depends upon the leptoquark couplings, and looks promising when they are large: see Ref. [20].

[13]this is slightly ironic, since the interesting '$B-$data' are really a *deficit* in the muonic channel.

[5] B. C. Allanach, S. Grab and H. E. Haber, JHEP **1101** (2011) 138 Erratum: [JHEP **1107** (2011) 087] Erratum: [JHEP **1109** (2011) 027] doi:10.1007/JHEP07(2011)087, 10.1007/JHEP09(2011)027, 10.1007/JHEP01(2011)138 [arXiv:1010.4261 [hep-ph]].

[6] C. Young (on behalf of ATLAS and CMS Collaborations), talk at Electroweak Session of Moriond 2016 conference, La Thuile, Italy.

[7] D. Alves *et al.* [LHC New Physics Working Group Collaboration], J. Phys. G **39** (2012) 105005 doi:10.1088/0954-3899/39/10/105005 [arXiv:1105.2838 [hep-ph]].

[8] G. Aad *et al.* [ATLAS Collaboration], JHEP **1510** (2015) 134 doi:10.1007/JHEP10(2015)134 [arXiv:1508.06608 [hep-ex]].

[9] A. Barr and J. Liu, arXiv:1605.09502 [hep-ph].

[10] S. Raychaydhuri and K. Sridhar, Cambridge Monographs on Mathematical Physics, Cambridge University Press (2016), ISBN 9780521768566.

[11] G. F. Giudice, R. Rattazzi and J. D. Wells, Nucl. Phys. B **544** (1999) 3 doi:10.1016/S0550-3213(99)00044-9 [hep-ph/9811291].

[12] B. C. Allanach, K. Odagiri, M. A. Parker and B. R. Webber, JHEP **0009** (2000) 019 doi:10.1088/1126-6708/2000/09/019 [hep-ph/0006114].

[13] R. Aaij *et al.* [LHCb Collaboration], Phys. Rev. Lett. **113** (2014) 151601 doi:10.1103/PhysRevLett.113.151601 [arXiv:1406.6482 [hep-ex]].

[14] R. Aaij *et al.* [LHCb Collaboration], JHEP **08** (2017) 055 [arXiv:1705.05802 [hep-ex]]

[15] S. Descotes-Genon, J. Matias and J. Virto, Phys. Rev. D **88** (2013) 074002 doi:10.1103/PhysRevD.88.074002 [arXiv:1307.5683 [hep-ph]].

[16] W. Altmannshofer and D. M. Straub, Eur. Phys. J. C **75** (2015) no.8, 382 doi:10.1140/epjc/s10052-015-3602-7 [arXiv:1411.3161 [hep-ph]].

[17] S. Descotes-Genon, L. Hofer, J. Matias and J. Virto, JHEP **1606** (2016) 092 doi:10.1007/JHEP06(2016)092 [arXiv:1510.04239 [hep-ph]].

[18] J. Goodman, M. Ibe, A. Rajaraman, W. Shepherd, T. M. P. Tait and H. B. Yu, Phys. Rev. D **82** (2010) 116010 doi:10.1103/PhysRevD.82.116010 [arXiv:1008.1783 [hep-ph]].

[19] J. Abdallah *et al.*, Phys. Dark Univ. **9-10** (2015) 8 doi:10.1016/j.dark.2015.08.001 [arXiv:1506.03116 [hep-ph]].

[20] B.C. Allanach, B. Gripaios and T. You, JHEP **1803** (2018) 021, [arXiv:1710.06363 [hep-ph]].

# Introduction to cosmology and dark matter

*A. De Simone*
SISSA and INFN Sezione di Trieste, Trieste, Italy

### Abstract
These are the lectures notes for the course on Introduction to Cosmology and Dark Matter given at the CERN European School for High Energy Physics (ESHEP) 2018. The audience consists of graduate students with particle physics background.

### Keywords
Cosmology; universe dynamics; expansion; dark matter; lectures.

## 1 Introduction

### 1.1 Motivations and Outline

These lectures are addressed to an audience of graduate students in experimental particle physics. So the first question usually is:

"Why should a particle physicist care about Cosmology"?

There are at least three main reasons to attend an introductory course on cosmology such as this one.

1. Cosmology provides insights on particle physics at energy scales which are impossible to probe on Earth. In the very early moments after the Big Bang the universe had a temperature (or energy) which would never be reachable again. So the by-products of the early universe dynamics we can measure today give us information about the physics at incredibly high energies.
2. Cosmology provides alternative (sometimes competitive) constraints on particle physics properties (e.g. neutrino physics, dark matter, etc.)
3. Cosmology provides motivations for (or completions of) particle physics models beyond the Standard Model. The need to solve cosmological issues like inflation, baryogenesis, dark matter calls for new particle physics which is able to model them and make predictions.

In these lectures I will give an overview of the Standard Model of Cosmology, its main successes and its drawbacks, with particular focus on the particle physics side.

In Lecture 1, I will describe the universe around us, its dynamics, the energy budget (Section 2), and provide introductory information about the 3 pillars constituting the Standard Model of Cosmology: Expansion, Big Bang Nucleosynthesis, Cosmic Microwave Background (Sections 2, 3, 4).

In Lecture 2, I will discuss the problem of Dark Matter (Section 5).

In Lecture 3, the main pitfalls of Standard Big Bang Cosmology and their possible resolution with the inflationary paradigm are described in Section 6, and then I conclude the course by mentioning the problem of the Baryon Asymmetry of the universe and some models of Baryogenesis in Section 7.

### 1.2 Warm-up

Throughout the course we will adopt the so-called "natural units", where the dimensions of basic physical quantities are related as

$$[\text{Energy}] = [\text{Mass}] = [\text{Temperature}] = [\text{Length}]^{-1} = [\text{Time}]^{-1} , \qquad (1)$$
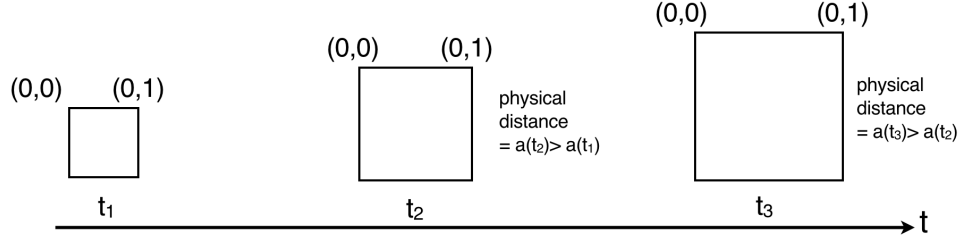
**Fig. 1:** *Expansion of physical distances.*

and the main physical constants are set as

$$\hbar = c = k_B = 1\,,\tag{2}$$

from which it follows that

$$1 = \hbar \cdot c \simeq 197.33\,\text{MeV} \cdot \text{fm} = 1.9733 \times 10^{-14}\,\text{GeV} \cdot \text{cm}\,,\tag{3}$$

so

$$1\,\text{GeV}^{-1} = 1.9733 \times 10^{-14}\,\text{cm} = 1.9733 \times 10^{-14}\,\frac{\text{cm}}{c} = 6.5822 \times 10^{-25}\,\text{s}\tag{4}$$

The fundamental mass scale of gravitational interactions is the Planck mass $M_P = 1.22 \times 10^{19}\,\text{GeV}$, while astronomical distances often appear in units of $1\,\text{pc} = 3.08 \times 10^{18}\,\text{cm}$.

## 2 The universe around us

### 2.1 Kinematics of the universe

#### 2.1.1 Expansion

From observations of the universe around us we can draw the conclusion that the universe is expanding. If you look at any two points in space, their relative distance was smaller in the past. In fact, the observation of red-shifted spectra of distant galaxies firmly supports the idea that the universe is expanding.

In an expanding universe, the physical distances between two points get larger and larger. They are proportional to a factor measuring the expansion of the universe: the *scale factor* $a(t)$ (Fig. 1). The velocity $v$ at which a galaxy at distance $d$ is going away from us is governed by the Hubble law

$$v = H_0 d\,,\tag{5}$$

where the velocity $v$ is related to the wavelengths of the photon emitted and observed

$$v = \frac{\lambda_{\text{observation}} - \lambda_{\text{emission}}}{\lambda_{\text{emission}}} \equiv z\,,\tag{6}$$

with $z$ being the *redshift* of the emission time $t_e$ with respect to the present time $t_0$

$$1 + z \equiv \frac{\lambda_{\text{observation}}}{\lambda_{\text{emission}}} = \frac{a(t_0)}{a(t_e)}\,,\tag{7}$$

measuring how much the universe has expanded since the emission of that photon. By Taylor-expanding the ratio of scale factors around the present time $t_0$

$$\frac{a(t)}{a(t_0)} = 1 + H_0(t - t_0) - \frac{1}{2} q_0 H_0^2 (t - t_0)^2 + \cdots\tag{8}$$

where the local expansion rate today ('Hubble constant') is

$$H_0 \equiv \left. \frac{\dot{a}}{a} \right|_{t_0}.$$ (9)

The linear term in Eq. (8) gives the Hubble law Eq. (5), while the quadratic term depends on the deceleration parameter $q$

$$q_0 \equiv -\left. \frac{\ddot{a}}{aH_0^2} \right|_{t_0} = -\left. \frac{\ddot{a}a}{\dot{a}^2} \right|_{t_0},$$ (10)

and it encodes the deviations from the Hubble law. The latest measurement [1] give

$$1/H_0 \simeq 1.4 \times 10^{10} \text{ yrs} \simeq 4.3 \, \text{Gpc} \simeq 1.3 \times 10^{26} \, \text{m},$$ (11)

for the 'Hubble time', or 'Hubble length', and it is customary to define

$$h \equiv \frac{H_0}{100 \, \text{km s}^{-1} \, \text{Mpc}^{-1}} \simeq 0.67.$$ (12)

When observed on very large scales ($> 100 \, \text{Mpc}$), the universe around us appears to be

- **homogenous**: the distribution of matter in the universe has a roughly constant density, or in other words the 2-point function of galaxies and galaxy clusters is much smaller than the Hubble length $1/H_0$;
- **isotropic**: if the expansion of the universe were not isotropic, we would observe large temperature anisotropies in the Cosmic Microwave Background.

The invariance under rotations is around any point of the Unverse, so it is isotropic.

These observations lead us to consider that no observer is special and there are no preferred directions, so the universe is homogenous and isotropic (Cosmological Principle).

### 2.1.2 Friedmann-Robertson-Walker metric

We now want to build a metric describing a homogeneous and isotropic universe. The Friedmann-Robertson Walker metric is

$$ds^2 = dt^2 - a(t)^2 \left[ \frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \right],$$ (13)

where the parameter $k$ can take three values

$$k = \begin{cases} +1 & \text{positive spatial curvature} \\ 0 & \text{zero spatial curvature} \\ -1 & \text{negative spatial curvature} \end{cases}$$ (14)

The scalar curvature of 3-dimensional spatial slices is

$$|{}^3\mathcal{R}| = \frac{6|k|}{a^2} \equiv \frac{6}{R_{\text{curv}}^2},$$ (15)

where $R_{\text{curv}}$ is a sort of curvature radius of the universe.

## 2.2 Dynamics of the universe

### 2.2.1 Einstein equations

The laws of gravity are the Einstein Equations, where the spacetime metric $g_{\mu\nu}$ and its corresponding Ricci tensor $R_{\mu\nu}$ and Ricci scalar $R$ are related to energy content expressed through the energy-momentum tensor $T_{\mu\nu}$

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G_N T_{\mu\nu} + \Lambda g_{\mu\nu} \,, \tag{16}$$

where $G_N$ is the Newton constant and we also included the cosmological constant term $\Lambda$.

The energy-momentum tensor for a perfect fluid with pressure $p$, energy density $\rho$ and 4-velocity $u^\mu$ is

$$T_{\mu\nu} = (p + \rho)u_\mu u_\nu - pg_{\mu\nu} \,, \tag{17}$$

which assumes a diagonal form in the reference frame of the fluid (comoving frame) where $u^\mu = (1, 0, 0, 0)$

$$T_{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & -p & 0 & 0 \\ 0 & 0 & -p & 0 \\ 0 & 0 & 0 & -p \end{pmatrix} \,. \tag{18}$$

The energy-momentum conservation law is simply expressed as the vanishing of the covariant derivative of the energy-momentum tensor:

$$\nabla_\mu T^{\mu\nu} = 0 \,, \tag{19}$$

whose $\nu = 0$ component in an expanding universe reads

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0 \,, \tag{20}$$

which is also known as "continuity equation". Alternatively, from the first law of thermodynamics $dU + pdV = TdS$, together with entropy conservation $dS = 0$, one gets

$$dU + pdV = 0 \implies d(a^3\rho) + pd(a^3) = 0 \,, \tag{21}$$

which is the same as Eq. (20).

### 2.2.2 Fluids

The perfect fluids are characterized by a proportionality relation between energy density and pressure

$$p = w\rho \tag{22}$$

where $w$ is constant in time. In this case, the continuity equation in Eq. (20) becomes

$$\frac{\dot{\rho}}{\rho} = -3(1 + w)\frac{\dot{a}}{a} \implies \rho \propto a^{-3(1+w)} \,. \tag{23}$$

This means that the energy density of different kinds of fluids scales down with the expansion of the universe with different powers of the scale factor:

$$\begin{aligned} \text{radiation:} \quad & w = 1/3 \quad \implies \quad \rho \propto a^{-4} \\ \text{dust:} \quad & w = 0 \quad \implies \quad \rho \propto a^{-3} \\ \text{vacuum energy:} \quad & w = -1 \quad \implies \quad \rho \propto \text{const.} \end{aligned}$$

The phases of the universe where the radiation/matter/vacuum energy are the dominant components are called radiation domination (RD), matter domination (MD) and vacuum energy domination, respectively.

Another argument to reach the same conclusions is to consider that the rest-mass energy must be a constant quantity unaffected by the expansion; the volume scales like $a^3$ in the expanding universe. So the energy density (energy per unit volume) of matter should scale like $a^{-3}$. For radiation, the energy density has a further $1/a$ factor due to the redshift, so $a^{-4}$.

### 2.2.3 Friedmann Equation

The $(00)$ component of the Einstein equations for the FRW metric with parameter $k$ gives the Friedmann Equation

$$H^2 = \frac{8\pi G_N}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}\,, \tag{24}$$

where the Hubble parameter $H \equiv \dot{a}/a$ is not a constant. The $(ii)$ component of the Einstein equations

$$\frac{\ddot{a}}{a} = -\frac{4\pi G_N}{3}(\rho + 3p) + \frac{\Lambda}{3} \tag{25}$$

does not add anything new with respect to the combination of the Friedmann Equation (24) and the conservation law (20).

So, the system of equations

$$H^2 = \frac{8\pi G_N}{3}\rho_{\text{tot}} - \frac{k}{a^2}\,, \tag{26}$$

$$\dot{\rho} + 3H(\rho + p) = 0\,, \tag{27}$$

where

$$\rho_\Lambda \equiv \frac{\Lambda}{8\pi G_N}\,, \tag{28}$$

is the cosmological constant energy density and

$$\rho_{\text{tot}} \equiv \rho + \rho_\Lambda\,, \tag{29}$$

is the total energy density, encodes the evolution of the universe and its constituents. The so-called "curvature energy density", encoded in the term proportional to $k$ is also sometimes indicated as $\rho_k = -3k/(8\pi G_N a^2)$.

### 2.2.4 Cosmological Dynamics

Let us first introduce a notation which is often used in cosmology. The "critical" energy density is defined as

$$\rho_c \equiv \frac{3H^2}{8\pi G_N}\,, \tag{30}$$

which today is $\rho_c \simeq 1.88 \times 10^{-29} h^2 \text{g cm}^{-3}$ or $\rho_c \simeq 1.05 \times 10^{-5} h^2 \text{ GeV cm}^{-3}$. The energy density today of each component is normalized to the critical density to provide the corresponding "Omega parameter" for matter, radiation, curvature and cosmological constant

$$\Omega_m \equiv \frac{\rho_m}{\rho_c}\,, \tag{31}$$

$$\Omega_r \equiv \frac{\rho_r}{\rho_c}\,, \tag{32}$$

$$\Omega_k \equiv \frac{\rho_k}{\rho_c} = -\frac{k}{a^2 H^2}\,, \tag{33}$$

$$\Omega_\Lambda \equiv \frac{\rho_\Lambda}{\rho_c} = \frac{\Lambda}{3H^2}\,. \tag{34}$$

In terms of the Omega parameters, the Friedmann equation (26) can be simply written as a sum rule

$$\Omega_m + \Omega_r + \Omega_\Lambda + \Omega_k = 1\,. \tag{35}$$

Now we have all the tools to study the time evolution of the scale factor of the universe. Let us carry out the case where the universe is filled by a single fluid, either matter dust ($w = 0$), radiation ($w = 1/3$) or vacuum energy ($w = -1$). The Friedmann Equation can be written as

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2 \left(\frac{\dot{a}}{a_0}\right)^{-3(1+w)}, \tag{36}$$

where $_0$ subscripts indicate present-time quantities. By introducing the new variable $y \equiv a/a_0$, a simple manipulation gives

$$\dot{y} = H_0 y^{1-\frac{3(1+w)}{2}} \implies y^{\frac{1}{2}+\frac{3}{2}w} dy = H_0 dt, \tag{37}$$

which in turns leads to

$$a(t) \propto t^{\frac{2}{3(1+w)}} \qquad (w \neq -1), \tag{38}$$
$$a(t) \propto e^{H_0 t} \qquad (w = -1), \tag{39}$$

So in a vacuum-dominated universe the scale factor expands exponentially, while in a radiation-dominated (RD) or matter-dominated (MD) phase the expansion is power-law, with exponents

$$a(t) \propto t^{2/3} \quad (w = 0, \text{MD}), \tag{40}$$
$$a(t) \propto t^{1/2} \quad (w = 1/3, \text{RD}). \tag{41}$$

From the Friedmann equation it follows that the total energy density of the universe equals the critical energy density if and only if the FRW parameter $k = 0$, which means the universe is flat

$$\rho_{\text{tot}} = \rho_c \iff k = 0 \iff \text{Flat universe}. \tag{42}$$

### 2.2.5 Energy Budget

The picture emerging from Cosmic Microwave Background (CMB) measurements performed by PLANCK in 2018 [1] is

$$h = 0.6736 \pm 0.0054 \tag{43}$$

$$\Omega_m h^2 = 0.1430 \pm 0.0011 \quad \begin{cases} \Omega_b h^2 = 0.02237 \pm 0.00015 \\ \Omega_{\text{CDM}} h^2 = 0.1200 \pm 0.0012 \end{cases} \tag{44}$$

$$\Omega_k h^2 = 0.0007 \pm 0.0019 \tag{45}$$
$$\Omega_\Lambda = 0.6847 \pm 0.0073 \tag{46}$$

So the curvature term is consistent with 0% of the energy budget (our Unvierse is flat!), while non-relativistic matter contributes to about 32% of the budget (split into 5% of ordinary baryonic matter and 27% of unknown dark matter), while the remaining 68% of the energy density of the present universe is in the form of unknown vacuum energy. In summary, we only know the nature of the 5% of what surrounds us.

### 2.2.6 Age of the universe

Very early (uncertain) times give an almost irrelevant contribution to the age of the universe, so we can compute the age of the universe form when it started RD or MD eras. Start with the definition of the Hubble parameter

$$\frac{\mathrm{d}a}{\mathrm{d}t} = Ha, \tag{47}$$

and the Friedmann equations written in terms of $\Omega$'s

$$H(a)^2 = H_0^2 \left[ \Omega_r \left( \frac{a_0}{a} \right)^4 + \Omega_m \left( \frac{a_0}{a} \right)^3 + \Omega_k \left( \frac{a_0}{a} \right)^2 + \Omega_\Lambda \right] . \tag{48}$$

From the two equations above it follows that

$$\mathrm{d}t = \frac{\mathrm{d}a}{aH_0} \frac{1}{\left[ \Omega_r \left( \frac{a_0}{a} \right)^4 + \Omega_m \left( \frac{a_0}{a} \right)^3 + \Omega_k \left( \frac{a_0}{a} \right)^2 + \Omega_\Lambda \right]^{1/2}} . \tag{49}$$

The estimate of the age of the universe is $t \simeq 13 \text{ Gyrs} = 1.3 \times 10^{10}$ yrs.

In an MD universe ($\Omega_r = \Omega_k = \Omega_\Lambda = 0, \Omega_m = 1$) Eq. (49) gives

$$t_0 = \frac{2}{3} \frac{1}{H_0} \simeq 9 \times 10^9 \text{ yrs}, \tag{50}$$

which is too young. By allowing a 70% contribution from vacuum energy, as suggested by Eq. (46): $\Omega_r = \Omega_k = 0, \Omega_m = 1 - \Omega_\Lambda = 0.3$, Eq. (49) can be integrated as

$$t = \frac{2}{3} \frac{1}{H_0 \sqrt{\Omega_\Lambda}} \sinh^{-1} \sqrt{\frac{\Omega_\Lambda}{1 - \Omega_\Lambda} \left( \frac{a}{a_0} \right)^3}, \tag{51}$$

so the present age of the universe would be

$$t_0 = \frac{2}{3} \frac{1}{H_0 \sqrt{\Omega_\Lambda}} \sinh^{-1} \sqrt{\frac{\Omega_\Lambda}{1 - \Omega_\Lambda}} \simeq 1.3 \times 10^{10} \text{ yrs}, \tag{52}$$

in perfect agreement with the estimate. The contribution of $\Lambda$ makes the universe older.

### 2.2.7 Distance-Redshift Relation

The light rays travel along geodesics defined by $\mathrm{d}s = 0$, so in the FRW metric (13) with $k = 0$ the trajectory of light rays is $\theta =$const., $\phi =$const. and $\mathrm{d}r = \mathrm{d}t/a$. Using $\mathrm{d}t$ from the definition of $H = (1/a)\mathrm{d}a/\mathrm{d}t$, we get

$$\mathrm{d}r = \frac{\mathrm{d}a}{a^2 H} , \tag{53}$$

which combined with Eq. (48) gives

$$r(a) = \frac{1}{H_0} \int_a^{a_0} \frac{\mathrm{d}a'}{a'^2 \left[ \Omega_r \left( \frac{a_0}{a'} \right)^4 + \Omega_m \left( \frac{a_0}{a'} \right)^3 + \Omega_k \left( \frac{a_0}{a'} \right)^2 + \Omega_\Lambda \right]^{1/2}} . \tag{54}$$

This equation is immediately rewritten in terms of the redshift $1 + z = a_0/a$, to get the distance-redshift relation

$$r(z) = \frac{1}{H_0} \int_0^z \frac{\mathrm{d}z'}{\left[ \Omega_r \left( 1 + z' \right)^4 + \Omega_m \left( 1 + z' \right)^3 + \Omega_k \left( 1 + z' \right)^2 + \Omega_\Lambda \right]^{1/2}} , \tag{55}$$

which allows to infer the distance of an object of known redshift $z$, depending on the energy content of the universe.

It is convenient also to introduce the *luminosity distance* $d_L$ of an object of given luminosity, from the definition of the flux of photons received from the object

$$\text{Flux} = \frac{\text{Luminosity}}{4\pi r(z)^2 (1+z)^2} \equiv \frac{\text{Luminosity}}{4\pi d_L^2} \tag{56}$$

so $d_L \equiv (1 + z)r(z)$, which again depends on the universe content. The two powers of $(1 + z)$ in the denominator are originated from the redshift of the energy and the relativistic dilation of time.
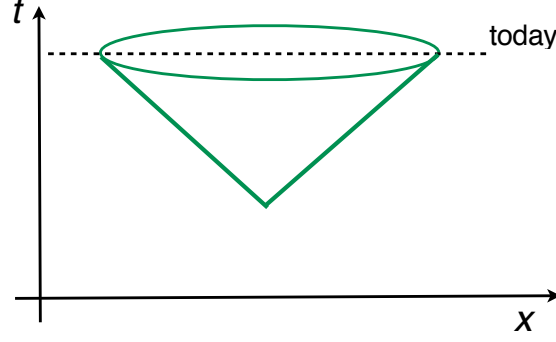
**Fig. 2:** *Particle horizon.*

### 2.2.8 Particle Horizon

A very important concept in cosmology is the notion of particle horizon. It is defined as the boundary between the visible universe and the part of the universe from which light signals have not reached us

$$d_H(t) = a(t) \int_0^t \frac{\mathrm{d}t'}{a(t')} \, . \tag{57}$$

So the particle horizon measures the portion of the universe in causal contact with us (see Fig. 2). Eq. (57) can be re-written using $\mathrm{d}t = \mathrm{d}a/(aH)$ and the expression for $H$ as in Eq. (48), so

$$d_H(t) = \frac{a(t)}{H_0} \int_0^{a(t)} \frac{\mathrm{d}a'}{a'^2 \left[ \Omega_r \left( \frac{a_0}{a'} \right)^4 + \Omega_m \left( \frac{a_0}{a'} \right)^3 + \Omega_k \left( \frac{a_0}{a'} \right)^2 + \Omega_\Lambda \right]^{1/2}} \, , \tag{58}$$

The special cases of matter-, radiation- and vacuum-domination are particularly interesting

$$d_H(t) = 3t = \frac{2}{H(t)} \propto a^{3/2} \qquad (\mathrm{MD}) \, , \tag{59}$$

$$d_H(t) = 2t = \frac{1}{H(t)} \propto a^2 \qquad (\mathrm{RD}) \, , \tag{60}$$

$$d_H(t) = \infty \, , \tag{61}$$

so particle horizon is growing with powers of the scale factors in MD and RD universes while there is no horizon in a universe dominated by vacuum energy.

### 2.2.9 Equilibrium Thermodynamics

It is useful for future developments to collect some formulae related to the equilibrium thermodynamics of the universe. Let us consider a particle species $A$ with $g_A$ degrees of freedom and chemical potential $\mu_A$, and characterized by a phase space distribution function $f_a(\mathbf{p})$ in momentum space, and energy $E(\mathbf{p})$. The distribution function of a species $A$ takes the form

$$f_A(\mathbf{p}) = \frac{1}{e^{(E(\mathbf{p}) - \mu_A)/T} \mp 1} \, , \quad E(\mathbf{p}) = \sqrt{|\mathbf{p}|^2 + m_A^2} \tag{62}$$

with $-(+)$ for Bose-Einstein (Fermi-Dirac) statistics, respectively. The equilibrium number density and energy density are given by

$$n_A^{\mathrm{eq}} = \frac{g_A}{(2\pi)^3} \int f(\mathbf{p}) \mathrm{d}^3 p \, , \tag{63}$$

$$\rho_A^{\text{eq}} \;=\; \frac{g_A}{(2\pi)^3} \int E(\mathbf{p}) f(\mathbf{p}) \mathrm{d}^3 p \,. \tag{64}$$

In extreme cases these integrals can be solved analytically.

For non-relativistic species (whose mass is much greater than the temperature $T \ll m_A$), the above expressions simplify to

$$n_A^{\text{eq}} \;=\; g_A \left( \frac{m_A T}{2\pi} \right)^{3/2} e^{-(m_A - \mu_A)/T} \tag{65}$$

$$\rho_A^{\text{eq}} \;=\; n_A^{\text{eq}} \cdot m_A \,. \tag{66}$$

Notice that these quantities are exponentially suppressed by the large mass of the species.

In the opposite regime of ultra-relativistic species ($T \gg m_A, \mu_A$), the expressions depend on the statistics. The number density (at equilibrium) is

$$n_A^{\text{eq}} = \frac{\zeta(3)}{\pi^2} g_A T^3 \begin{cases} 1 & (\text{bosons}) \\[2mm] \frac{3}{4} & (\text{fermions}) \end{cases}. \tag{67}$$

The Riemann zeta function of 3 is $\zeta(3) \equiv \sum_{n=1}^{\infty}(1/n^3) \simeq 1.20206$. The energy density (at equililbrium) is

$$\rho_A^{\text{eq}} = \frac{\pi^2}{30} g_A T^4 \begin{cases} 1 & (\text{bosons}) \\[2mm] \frac{7}{8} & (\text{fermions}) \end{cases}. \tag{68}$$

Because of the suppression in the non-relativistic regime, the energy density at given temperatures is exponentially dominated by the degrees of freedom which are ultra-relativistic at that temperature. For a collection of several particle species in equilibrium where the species $i$ has thermal distribution with temperature $T_i$, to a very good approximation the total energy density is

$$\rho_{\text{tot}} = \frac{\pi^2}{30} g_*(T) T^4 \,, \tag{69}$$

where $g_*(T)$ is the total number of relativistic (massless) degrees of freedom at temperature $T$ given by

$$g_*(T) = \sum_{b \in \text{bosons}} g_b \left( \frac{T_b}{T} \right)^4 + \frac{7}{8} \sum_{f \in \text{fermions}} g_f \left( \frac{T_f}{T} \right)^4 \,. \tag{70}$$

The Hubble rate in the RD era (where $a(t) \propto \sqrt{t}$ so $H = 1/(2t)$) can thus be written as

$$H^2 = \frac{8\pi G_N}{3} \rho_{\text{tot}} = \frac{8\pi G_N}{3} \frac{\pi^2}{30} g_*(T) T^4 \simeq 1.66 \sqrt{g_*(T)} \frac{T^2}{M_P} \,, \tag{71}$$

hence, we obtain the time-temperature relation

$$t \simeq \frac{0.30}{\sqrt{g_*(T)}} \frac{M_P}{T^2} \simeq \frac{2.41}{\sqrt{g_*(T)}} \left( \frac{\text{MeV}}{T} \right)^2 \text{s} \,. \tag{72}$$

### 2.2.10 Temperature-Expansion Relation

The 1st law of thermodynamics relates the change in the energy $\mathrm{d}U$ to the change of entropy $\mathrm{d}S$ as

$$\mathrm{d}U + p\,\mathrm{d}V = T\,\mathrm{d}S \tag{73}$$

The entropy density $s$ is therefore

$$s(T) \equiv \frac{S(V,T)}{V} = \frac{\rho(T) + p(T)}{T} = \frac{4}{3}\frac{\rho(T)}{T} = \frac{2\pi^2}{45}g_{*,s}(T)T^3 \qquad (74)$$

where we have used that $p = 1/3\rho$ for RD and defined the quantity

$$g_{*,s}(T) = \sum_{b \in \text{bosons}} g_b \left(\frac{T_b}{T}\right)^3 + \frac{7}{8}\sum_{f \in \text{fermions}} g_f \left(\frac{T_f}{T}\right)^3 , \qquad (75)$$

which is similar to $g_*(T)$ in Eq. (70) but with the different temperature dependence. Since the energy density scales like $\rho(T) \propto T^4$, the entropy density scales like

$$s(T) \propto g_{*,s}T^3 , \qquad (76)$$

and therefore the conservation of the entropy within a comoving volume $V$ in thermal equilibrium gives

$$S(V,T) = s(T)V = \text{const.} \Longrightarrow g_{*,s}T^3a^3 = \text{const.} \Longrightarrow T \propto \frac{1}{g_{*,s}^{1/3}a} . \qquad (77)$$

In periods where $g_{*,s}$ is also a constant, the temperature simply scales as the inverse of the scale fatctor $T \propto 1/a$.

## 3 Big Bang Nucleosynthesis

Big Bang Nucleosynthesis (BBN) occurs at times $1\,\text{s} \lesssim t \lesssim 10^3\,\text{s}$ or equivalently at temperatures of the universe $1\,\text{MeV} \gtrsim T \gtrsim 10\,\text{keV}$. Before BBN, the photons have sufficiently high energy to prevent the formation of nuclei by dissociating them. As the universe cools down, the nuclei of light elements H, D, $^3$He, $^4$He, $^7$Li, get produced with predicted abundances in a remarkably good agreement with the observed ones. All elements heavier than $^7$Li are produced later in the history of the universe by nuclear reactions in stars or by other astrophysical processes like supernovae.

BBN is the earliest probe of the universe. Before BBN, we do not know anything about the universe. We are not even sure that the universe existed with temperatures above the MeV.

BBN is one of the main successes of standard cosmology. This success has 3 important consequences:

1. it confirms the theory of the early universe;
2. it provides a determination of the baryon-to-photon ratio $\eta$;
3. to avoid spoiling its success, particle physics theories beyond the Standard Model are constrained.

The predictions for abundances of light elements span 9 orders of magnitude and are all well fitted by a single parameter: the baryon-to-photon ratio $\eta \equiv n_B/n_\gamma$ (see Figure 3). This is one of the greatest successes of Standard Cosmology.

The measurement of light element abundances implies a measurement of $\eta$ and hence a measurement for $\Omega_b$ today. In fact, the energy density in baryons (non-relativistic particles with mass equal to the nucleon mass $m_N$) can be written as

$$\Omega_b h^2 = \frac{m_N n_B}{\rho_c/h^2} = \eta \frac{m_N n_\gamma}{\rho_c/h^2} = \frac{\eta}{2.74 \times 10^{-8}} . \qquad (78)$$

From BBN $0.019 \leq \Omega_b h^2 \leq 0.024$, in good agreement with the independent measurement from CMB (see Eq. (44)). These values are consistent with $\eta \simeq 6 \times 10^{-10}$.

Therefore, together with the measurement of $\Omega_m$ from CMB, BBN predicts that $\Omega_b < \Omega_m$, thus providing a compelling argument for the existence of a non-baryonic matter component of the universe, called Dark Matter (DM).
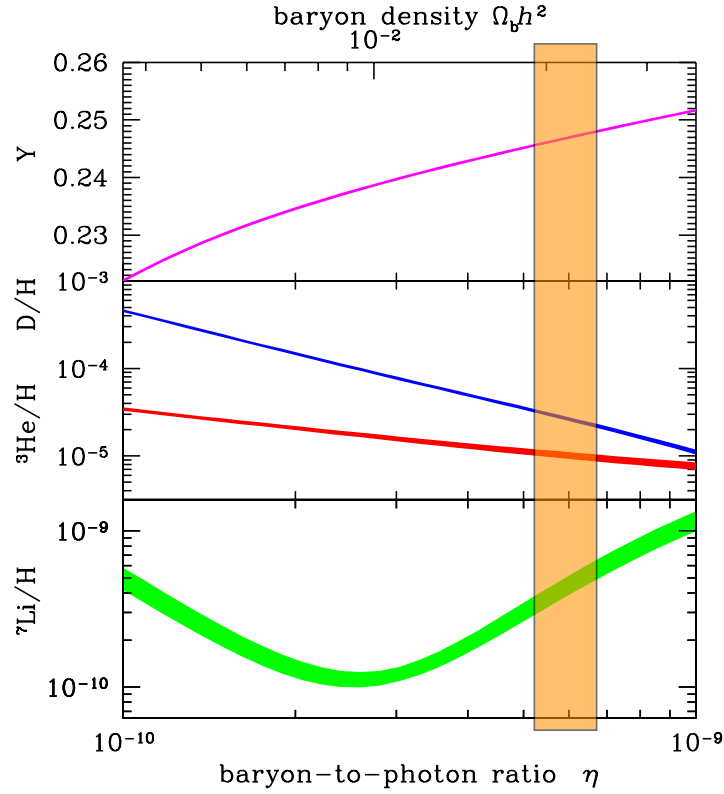
**Fig. 3:** *BBN predictions for light elements abundances.* (adapted from Ref. [2])

## 3.1 Helium fraction

As an example, let us see how much Helium-4 is produced during BBN. We will be deliberately oversimplifying the discussion, for simplicity. For a more general and complete treatment of the BBN dynamics see e.g. the excellent textbooks in Refs. [3–5].

Observations of metal-poor stars and gas clouds provide an experimental determination of the abundance of $^4$He nuclei with respect to the total number of baryons (protons + neutrons) as

$$Y \equiv \frac{4n_{He}}{n_n + n_p} \simeq 0.24 \,. \tag{79}$$

So, in order to compute $Y$ we need to compute $n_n/n_p$ and $n_{He}/n_p$.

At very early times ($T \gg 1$ MeV, $t \ll 1$ s), there are only protons and neutrons which are kept in equilibrium by the reactions

$$n + \nu_e \quad \leftrightarrow \quad p + e^- \tag{80}$$

$$n + e^+ \quad \leftrightarrow \quad p + \bar{\nu}_e \tag{81}$$

$$n \quad \leftrightarrow \quad p + e^- + \bar{\nu}_e \tag{82}$$

so $n_n = n_p$. When these reactions are in equilibrium, they enforce the balance of chemical potentials as

$$\mu_n + \mu_{\nu_e} = \mu_p + \mu_e \,. \tag{83}$$

Since the chemical potentials of electrons and neutrinos are negligibly small, we can conclude that $\mu_n = \mu_p$. At MeV temperatures neutrons and protons are non-relativistic, and their mass difference starts to be

important. So, recalling the equilibrium number densities for non-relativistic species Eq. (65) we obtain the neutron-to-proton ratio (in equilibrium) as

$$\frac{n_n^{\rm eq}}{n_p^{\rm eq}} = \frac{g_n}{g_p} \left(\frac{m_n}{m_p}\right)^{3/2} e^{-Q/T} e^{(\mu_n - \mu_p)/T} \simeq e^{-Q/T}\,, \tag{84}$$

where the neutron-proton mass difference is

$$Q = m_n - m_p \simeq 1.29\,{\rm MeV}\,. \tag{85}$$

and we have used that $m_n \simeq m_p$, $\mu_n = \mu_p$ and $g_n = g_p = 2$. The equilibrium is broken by expansion as temperature goes down. The total scattering rate of the reactions involving neutrons and protons is

$$\Gamma = \Gamma(n + \nu_e \leftrightarrow p + e^-) + \Gamma(n + e^+ \leftrightarrow p + \bar{\nu}_e) \simeq 0.96 \left(\frac{T}{\rm MeV}\right)^5 {\rm s}^{-1}\,, \tag{86}$$

while the Hubble parameter during radiation domination at $T \lesssim 1\,{\rm MeV}$ is

$$H \simeq 1.66\sqrt{10.75}\frac{T^2}{M_P} \simeq 0.68 \left(\frac{T}{\rm MeV}\right)^2 {\rm s}^{-1}\,. \tag{87}$$

Therefore, the comparison of the scattering rate with the expansion rate gives

$$\frac{\Gamma}{H} \simeq \left(\frac{T}{0.8\,{\rm MeV}}\right)^3\,. \tag{88}$$

So at $T \gtrsim 0.8$ MeV, the neutron-to-proton ratio follows its equilibrium value, while at $T \lesssim 0.8$ MeV it decouples and freezes out at a value

$$\frac{n_n}{n_p}(T \lesssim 0.8\,{\rm MeV}) = \frac{n_n^{\rm eq}}{n_p^{\rm eq}}(T = 0.8\,{\rm MeV}) = e^{-1.29/0.8} \simeq 0.2 \tag{89}$$

and correspondingly the neutron fraction is

$$X_n(T \lesssim 0.8\,{\rm MeV})|_{\rm before\,decay} \equiv \frac{n_n}{n_n + n_p} = \frac{\frac{n_n}{n_p}}{1 + \frac{n_n}{n_p}} \simeq \frac{e^{-1.29/0.8}}{1 + e^{-1.29/0.8}} \simeq 0.17\,. \tag{90}$$

After that time, some neutrons decay ($\tau_n \simeq 886\,{\rm s}$) and at later time, at the onset of BBN, $T_D \simeq 70\,{\rm keV}$ (the temperature at which deuterium production becomes thermodynamically favoured), there are slightly fewer neutrons. The time-temperature relation in the temperature regime $T < T_D$ is obtained from Eq. (72), with $g_* = 3.36$ (after $e^\pm$ annihilations),

$$t(T) \simeq 1.32 \left(\frac{1\,{\rm MeV}}{T}\right)^2 {\rm s}\,, \tag{91}$$

and therefore the neutron fraction at late times is

$$X_n(T < T_D) = X_n|_{\rm before\,decay} \times e^{-t(T_D)/\tau_n} \simeq 0.12\,, \tag{92}$$

leading to the neutron-to-proton ratio

$$\frac{n_n}{n_p}(T < T_D) \simeq 0.14\,. \tag{93}$$

Next, we need the number density of Helium nuclei $n_{He}$. Helium-4 is not produced by direct synthesis of $2n$ and $2p$, as the corresponding reaction rates are highly suppressed in the dilute (high entropy) limit,

but rather by burning of deterium $^2H = D$. So, in order to produce $^4He$ one needs to "wait" until the $D$ is produced, which occurs relatively late, because of the small binding energy of $D$ (the so-called "deuterium bottleneck"). Only after $D$ is formed, it can be burnt by the reactions

$$D + D \quad \rightarrow \quad ^3He + n \,, \tag{94}$$
$$D + D \quad \rightarrow \quad ^3H + p \,, \tag{95}$$

and then provide the fuel for Helium-4 production

$$^3He + D \quad \rightarrow \quad ^4He + p \,, \tag{96}$$
$$^3H + D \quad \rightarrow \quad ^4He + n \,. \tag{97}$$
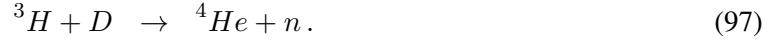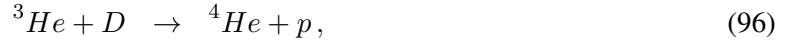
Deuterium is formed by the direct synthesis of one neutron and one proton in the reaction

$$p + n \rightarrow D + \gamma \,. \tag{98}$$

The D production becomes efficient at temperatures $T \lesssim T_D \simeq 70$ keV, and nearly all free neutrons get bound into Helium-4 nuclei, so

$$n_{^4He} \simeq n_n/2 \,, \tag{99}$$

since each Helium-4 nucleus contains two neutrons. Now we have all the ingredients to estimate the $^4$He abundance produced by BBN, by combining Eqs. (93) and (99)

$$Y = \frac{4 n_{^4He}}{n_n + n_p} = \frac{4 n_n/2}{n_n + n_p} = \frac{2 n_n/n_p}{1 + n_n/n_p} \simeq 0.24 \,, \tag{100}$$

in very good agreement with the observed value.

## 4 Cosmic Microwave Background

In 1965 Arno Penzias and Robert Wilson published a paper where they admitted to have failed to eliminate a background noise coming from all directions, corresponding to a residual photon background with temperature of about 3 K. Ten years later they shared the Nobel prize in physics for the discovery of the Cosmic Microwave Background (CMB) radiation!

What was that noise?

At temperatures above the electron mass $m_e$ $(T > m_e)$ the electrons/positrons and radiation were in thermal equilibrium. When the temperature of the universe goes down to a fraction of the electron mass $(T \lesssim m_e)$, electrons and positrons become non-relativistic and their equilibrium number densities become exponentially suppressed compared to the number density of photons, so the reaction

$$e^+ e^- \longleftrightarrow \gamma \gamma \tag{101}$$

goes out of equilibrium and chemical equilibrium is broken. However, matter (residual electrons) and radiation are still in kinetic equilibrium, through the elastic reaction (Compton scattering)

$$e^- \gamma \longleftrightarrow e^- \gamma \,, \tag{102}$$

whose cross section for non-relativistic electrons reduces to the Thomson scattering cross section (in classical electrodynamics)

$$\sigma_T \simeq \frac{8\pi}{3} \left( \frac{\alpha}{m_e} \right)^2 \simeq 6.7 \times 10^{-25}\,\mathrm{cm}^2 \,. \tag{103}$$

Notice that the analogous Compton scattering of photons off protons $p\gamma \leftrightarrow p\gamma$ is irrelevant since the corresponding cross section is suppressed by $(m_e/m_p)^2 \sim 10^{-6}$. The Compton scattering keeps the

photons coupled to matter until a much later time (photon decoupling) than when $T \simeq m_e$. Until then, kinetic equilibrium is attained and photons are continuously scattering off electrons. The universe was filled with an almost perfect black-body radiation.

When Compton scatterings start becoming ineffective with respect to the expansion rate of the universe, the photons "decouple" from matter and then propagate freely until today. They just cooled with expansion down to a temperature $T_0 = 2.7$ K $= 2.3 \times 10^{-4}$ eV. This radiation is the residual electromagnetic radiation from the Big Bang, observed as a highly isotropic "noise". The CMB can therefore be interpreted as a snapshot of the universe when it was very young (about 300,000 yrs old, as we will see later). The CMB provides a huge deal of information about our universe, and it is the most powerful cosmological probe available today.

### 4.1 Photon energy density

The energy distribution of thermal photons follows the Planck distribution (let us restore the units of $c, \hbar, k_B$ in this subsection)

$$n(\omega, T)\mathrm{d}\omega = \frac{1}{c^3}\frac{1}{e^{\hbar\omega/(k_B T)} - 1}\frac{2\mathrm{d}^3\omega}{(2\pi)^3} = \frac{1}{\pi^2 c^3}\frac{\omega^2 \mathrm{d}\omega}{e^{\hbar\omega/(k_B T)} - 1} \tag{104}$$

the differential energy spectrum

$$u(\omega, T)\mathrm{d}\omega = (\hbar\omega)n(\omega, T)\mathrm{d}\omega = \frac{\hbar}{\pi^2 c^3}\frac{\omega^3 \mathrm{d}\omega}{e^{\hbar\omega/(k_B T)} - 1} \tag{105}$$

is the usual one for a black body. Therefore the total energy density in radiation is given by the integral

$$\rho_\gamma = \int_0^\infty u(\omega, T)\mathrm{d}\omega = \frac{\hbar}{\pi^2 c^3}\left(\frac{k_B T}{\hbar}\right)^4\int_0^\infty\frac{\xi^3 \mathrm{d}\xi}{e^\xi - 1} = \frac{\pi^2 k_B^4}{15\hbar^3 c^3}T^4 \equiv \sigma T^4 \tag{106}$$

with $\sigma = 4.72 \times 10^{-3}$ eV cm$^{-3}$K$^{-4}$ being the Stefan-Boltzmann constant. This energy density today (T=2.7 K) is $\rho_\gamma \simeq 0.26$ eV cm$^{-3}$, which translates into

$$\Omega_r h^2 = \frac{\rho_\gamma}{\rho_c/h^2} \simeq 4 \times 10^{-5} \,, \tag{107}$$

so the present radiation energy density is neglible.

### 4.2 Photon decoupling

Let us compute the time when the CMB formed (or equivalently the redshift of photon decoupling $z_{\mathrm{dec}}$). We will work under the simplifying approximation that the plasma is in chemical and thermal equilibrium among all its components, aiming at providing the reader with the basic elements. For more exhaustive discussions please refer e.g. to Refs. [4, 5].

The goal is to estimate the time of *photon decoupling*, namely when the photons stop interacting with matter and propagate through the universe along geodesics. We first need to find the number density of free electons, or equivalently the free electron fraction (or ionization fraction)

$$X_e \equiv \frac{n_e}{n_p + n_H} \,. \tag{108}$$

Free electrons get bound to protons to form neutral H atoms through the capture reaction

$$p + e^- \longleftrightarrow H + \gamma \,, \tag{109}$$

and the binding energy of H is

$$B_H \equiv m_e + m_p - m_H = 13.6\,\text{eV}\,. \tag{110}$$

When $T \ll m_e$, electrons, protons and Hydrogen atoms are non-relativistic and their equilibrium number densities are given by Eq. (65)

$$n_i^{\text{eq}} = g_i \left(\frac{m_i T}{2\pi}\right)^{3/2} e^{-(m_i - \mu_i)/T}, \quad (i = e, p, H), \tag{111}$$

and recall that $g_p = g_e = 2$. The first type of Hydrogen involved in the electron capture in Eq. (109) is $H_{1s}$, i.e. the ground state (with 2 hyperfine states, one with spin 0 and one with spin 1, so $g_{H_{1s}} = 1 + 3 = 4$). By adding to the 3 conditions for $n_i^{\text{eq}}$ the 3 relations following from the equilirium reaction in Eq. (109)

$$\begin{cases} \mu_p + \mu_e &= \mu_H \qquad &\text{(chemical equilibrium)} \\ n_e &= n_p \qquad &\text{(charge neutrality)} \\ n_p + n_H &= 0.76\,n_B = 0.76\,\eta n_\gamma \qquad &\text{(tot. number of baryons without } He\text{)} \end{cases} \tag{112}$$

(recall that after BBN about 24% of the baryons consists of $^4$He) we have 6 equations for the 6 unknowns $n_{p,e,H}, \mu_{p,e,H}$. We can then compute, in the equilibrium approximation, the quantity

$$\begin{aligned} \left.\frac{X_e^2}{1 - X_e}\right|_{\text{eq}} &= \frac{n_e^{\text{eq}} + n_H^{\text{eq}}}{n_H^{\text{eq}}} \frac{n_e^{\text{eq}} n_p^{\text{eq}}}{(n_e^{\text{eq}} + n_H^{\text{eq}})^2} = \frac{n_e^{\text{eq}} n_p^{\text{eq}}}{n_H^{\text{eq}}(n_e^{\text{eq}} + n_H^{\text{eq}})} = \frac{1}{0.76 \cdot \eta n_\gamma} \frac{n_e^{\text{eq}} n_p^{\text{eq}}}{n_H^{\text{eq}}} \\ &= \frac{1}{0.76 \cdot \eta n_\gamma} \frac{g_e g_p}{g_H} \left(\frac{m_e T}{2\pi}\right)^{3/2} e^{(\mu_e + \mu_p - \mu_H)/T} e^{-(m_e + m_p - m_H)/T} \\ &= \frac{1}{0.76 \cdot \eta n_\gamma} \left(\frac{m_e T}{2\pi}\right)^{3/2} e^{-B_H/T}\,. \end{aligned} \tag{113}$$

Inserting $n_\gamma^{\text{eq}} = (2/\pi^2)\zeta(3)T^3$, we get the *Saha equation* for the equilibrium ionization fraction of electrons

$$\left.\frac{X_e^2}{1 - X_e}\right|_{\text{eq}} = \frac{\sqrt{\pi}}{0.76 \cdot 4\sqrt{2}\zeta(3)} \frac{1}{\eta} \left(\frac{m_e}{T}\right)^{3/2} e^{-B_H/T}\,. \tag{114}$$

The latter equation can be solved in the two temperature regimes

$$T \gtrsim B_H \implies \left.\frac{X_e^2}{1 - X_e}\right|_{\text{eq}} \simeq 10^9 \left(\frac{m_e}{T}\right)^{3/2} \simeq 10^5 \implies X_e \simeq 1 \quad \text{(all H ionized)} \tag{115}$$

$$T < B_H \implies X_e^{\text{eq}} \ll 1 \implies X_e^{\text{eq}}(T) \simeq \left[\frac{\sqrt{\pi}}{0.76 \cdot 4\sqrt{2}\zeta(3)} \frac{1}{\eta} \left(\frac{m_e}{T}\right)^{3/2} e^{-B_H/T}\right]^{1/2}\,. \tag{116}$$

Now that we have an expression for the free electron fraction at late times, we can proceed to compute the time of photon decoupling.

Photon decoupling occurs when the rate of photon-electron (Compton) scattering is less than the expansion rate: $\Gamma_e \simeq n_e \sigma_T \lesssim H$. Assume for simplicity a matter-dominated universe with $\Omega_m = 1$ (but generalizations are straightforward), so

$$H(T) \simeq H_0 a^{-3/2} = H_0 \left(\frac{T}{T_0}\right)^{3/2}\,. \tag{117}$$

Then

$$n_e = X_e n_B \,, \tag{118}$$

$$n_B = \frac{\Omega_b \rho_c}{m_p}\left(\frac{a_0}{a}\right)^3 = \frac{\Omega_b \rho_c}{m_p}\left(\frac{T}{T_0}\right)^3 \simeq 2.2 \times 10^{-7}\,\mathrm{cm}^{-3}\left(\frac{T}{T_0}\right)^3 \,, \tag{119}$$

from which it follows that the rate for electron Compton scatterings is

$$\Gamma_e(T) = n_e \sigma_T \simeq X_e\, 1.5 \times 10^{-31}\,\mathrm{cm}^{-1}\left(\frac{T}{T_0}\right)^3 \,. \tag{120}$$

Then compare $\Gamma_e$ with $H$ from Eq. (117), to get the temperature $T_{\mathrm{dec}}$ at which they are equal (recall $H_0^{-1} = 9.3 \times 10^{27} \cdot h^{-1}\,\mathrm{cm}$, $h \simeq 0.7$)

$$\Gamma_e(T_{\mathrm{dec}}) = H(T_{\mathrm{dec}}) \Longrightarrow \left(\frac{T_{\mathrm{dec}}}{T_0}\right)^{3/2} \simeq \frac{1}{2 \times 10^{-3} X_e(T_{\mathrm{dec}})} \,, \tag{121}$$

and then solve numerically the implicit equation for $T_{\mathrm{dec}}$, where $X_e(T)$ is given by the Saha equation in Eq. (116), arriving at

$$T_{\mathrm{dec}} \simeq 1000\, T_0 \simeq 0.2\,\mathrm{eV} \Longrightarrow 1 + z_{\mathrm{dec}} = \frac{T_{\mathrm{dec}}}{T_0} \simeq 1000$$

$$\Longrightarrow t_{\mathrm{dec}} = \frac{2}{3 H_0 (1 + z_{\mathrm{dec}})^{3/2}} \simeq 300\,000\,\mathrm{yrs.} \tag{122}$$

This is the time of photon decoupling (last scattering), when the CMB is formed. Before photon decoupling the plasma is opaque, because of photons scattering off free electrons. After decoupling, photons do not scatter anymore and the universe becomes transparent to radiation.

If instead of the Hubble parameter given by Eq. (117) for $\Omega_m = 1$ one considers a more realistic $\Lambda CDM$ model where $\Omega_m = 0.27, \Omega_\Lambda = 1 - \Omega_m$, one gets $z_{\mathrm{dec}} \simeq 1089$, so $T_{\mathrm{dec}} \simeq 0.26$ eV.

## 4.3 Concluding remarks

The CMB is actually not perfectly isotropic. There are temperature anisotropies of the order of

$$\frac{\Delta T}{T} \sim 10^{-5} \,. \tag{123}$$

Indeed, these anisotropies carry a great deal of cosmological information. For example, the two-point correlation functions of the temperature maps crucially depend on the cosmological paramters like $H_0, \Omega_b, \Omega_{\mathrm{tot}}$ etc. By a careful analysis of these anisotropies, satellite experiments like COBE, WMAP and lately PLANCK were able to determine the cosmological parameters with greater and greater accuracy. The CMB anisotropies in the CMB are well described by acoustic oscillations in the photon-baryon plasma. Both ordinary baryonic and dark matter interact gravitationally with radiation, but only ordinary matter interacts also electromagnetically. So baryonic and dark matter affect the CMB differently. From the peaks of the CMB it is possible to determine the density of baryonic and dark matter. The resulting best-fit 'concordance' cosmological model is known as $\Lambda$CDM (cosmological constant plus cold dark matter), where roughly

$$\Omega_{\mathrm{tot}} \sim 1.0\,, \quad \Omega_{\mathrm{matter}} \sim 0.3\,, \quad \Omega_{\mathrm{radiation}} \sim 0.0\,, \quad \Omega_\Lambda \sim 0.7 \tag{124}$$

The accurate determination of the energy content of the universe was another great triumph of standard cosmology!

So, although standard cosmology is very successful at providing a picture of the universe from BBN to today, there are several questions still lacking an answer, for instance: what is the dark matter made of? why there is a matter-antimatter asymmetry? what happened in the first three minutes of the universe (before BBN)? We will discuss some possible answers to these (and other) questions in the remainder of the course.

## 5 Dark Matter

### 5.1 Evidences for dark matter

We already discussed that both BBN and CMB observations (see Section 2.2.5) provide compelling arguments in favour of the existence of an unknown component of the universe consisting of non-baryonic matter, dubbed Dark Matter (DM). The existence of DM is by now firmly established also by other types of observational evidences.

#### 5.1.1 Galaxy clusters

– **Coma cluster.** In 1933, F. Zwicky measured the proper motion of galaxies in the Coma galaxy cluster (a group of $\sim 1000$ galaxies, within a radius of $\sim 1$ Mpc).

The mass $M$ and the size $R$ of the cluster of $N$ galaxies can be related to the velocity dispersion of galaxies (the velocities are projected along the line of sight) according to the virial theorem:

$$
\begin{aligned}
\langle V \rangle + 2\langle K \rangle &= 0 \,, \\
\langle V \rangle &= -\frac{N^2}{2} G_N \frac{\langle m^2 \rangle}{R} \quad \text{(average pot energy due to } N^2/2 \text{ pairs of galaxies)} \,, \\
\langle K \rangle &= N \frac{\langle m v^2 \rangle}{2} \quad \text{(average kin energy due to } N \text{ galaxies)} \,.
\end{aligned}
\tag{125}
$$

The total mass $M$ is

$$
M = N\langle m \rangle \sim \frac{2R\langle v^2 \rangle}{G_N} \,,
\tag{126}
$$

from which it was computed the mass-to-luminosity ratio to be much larger than the one for an average star like the Sun

$$
\frac{M}{L} \sim 300 h \frac{M_\odot}{L_\odot} \,.
\tag{127}
$$

So the value obtained is about 300 times greater than expected from their luminosity, which means that most of the matter is not luminous, so it is dark.

– **X-ray observations**. The gravitational potential (and hence the total mass) of galaxy clusters can also be measured by X-ray observations. In fact, most of the ordinary mass in cluster is in the form of hot gas, emitting X-ray frequencies. The X-rays are produced by electrons.

It is possible to measure the spatial distributions of the electron number density $n_e(r)$ and of the electron temperature $T_e(r)$. The number density of baryons $n_b(r)$ will be proportional to $n_e$ up to a factor $\mu$ depending on the chemical composition: $n_b(r) = \mu n_e(r)$. The pressure is mostly due to electrons, so $P(r) = n_e(r)T_e(r)$.

The hydrostatic equilibrium relates the pressure $P$ to the radius $R$ through the mass $m$ which in turn depends on the energy density in baryons $\rho_b$

$$
\mathrm{d}P = -\mathrm{d}m \frac{\text{acceleration}}{\text{Area}} = -\rho_b(R)\frac{\mathrm{d}V}{\text{Area}}\frac{G_N M(R)}{R^2} = -\rho_b(R)\frac{G_N M(R)}{R^2}\mathrm{d}R \,,
\tag{128}
$$

where the total mass enclosed in a sphere of radius $R$ is

$$
M(R) = 4\pi \int_0^R \rho(r) r^2 \mathrm{d}r \,.
\tag{129}
$$

This leads to

$$
\frac{\mathrm{d}P}{\mathrm{d}R} = -n_b(r) m_b \frac{G_N M(R)}{R^2} \,.
\tag{130}
$$

In this equation, the left-hand side is measured from temperature maps from X-ray spectra, the term $n_b(r)$ is obtained from X-ray luminosity and spatial distributions of electrons, so only $M(R)$
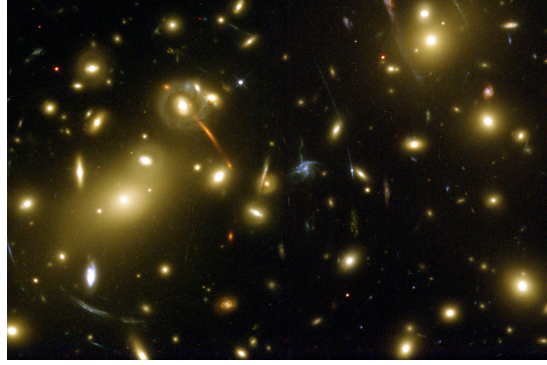
**Fig. 4:** *The gravitational lensing from Abell NGC2218.*
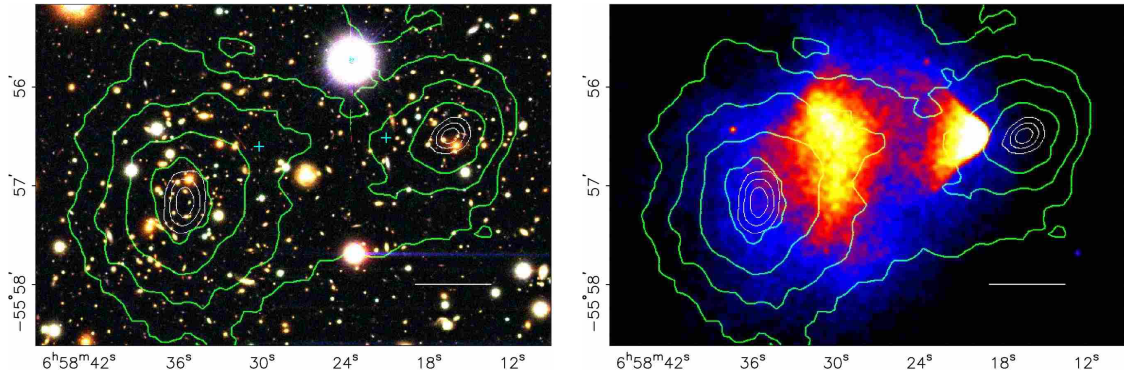


**Fig. 5:** *The "Bullet cluster" 1E0657-558. The image shows two colliding clusters of galaxies. The green lines show the gravitational equipotential surfaces, measured by gravitational lensing. Brighter regions are the hot baryonic gas, observed in X-ray by Chandra. Figure taken from Ref. [6].*

is uknown and can be determined. The result for $M$ is again that $M$ should be more than the contribution of just visible (baryonic) matter $M_b$.

– **Gravitational lensing**. Gravitational lensing techniques use the gravitational distortion of images of distant galaxies due to a gravitational mass (e.g. a cluster) along the line of sight (see Fig. 4). This way, it is possible to reconstruct the gravitational potential, and hence the total mass distribution of the cluster. The result is that more matter than the visible one is required, and also differently distributed.

– **Bullet cluster**. The so-called "bullet cluster" (see Fig. 5) is a recent merging of galaxy clusters. The gravitational potential is not produced by baryons, but by DM. In the collision, the hot gas is collisional and loses energy, so it slows down and lags behind DM; the DM clusters are collisionless and passed through each other.

### 5.1.2 Galaxies

The dependence of the velocity $v(R)$ of stars in a galaxy, as a function of the distance $R$ from the galactic center (**rotation curves**), is given by Newton's law (assuming circular motion)

$$v(R) = \sqrt{\frac{G_N M(R)}{R}}, \qquad M(R) = 4\pi \int_0^R \rho(r) r^2 \mathrm{d}r, \tag{131}$$

wnere $\rho(r)$ is the mass density. The contribution to $\rho$ from luminous matter would lead to $v(r) \propto R^{-1/2}$ at large $R$. But observationally, one has $v(R) \simeq$ constant, see Fig. 6. Explaining the observed rotation
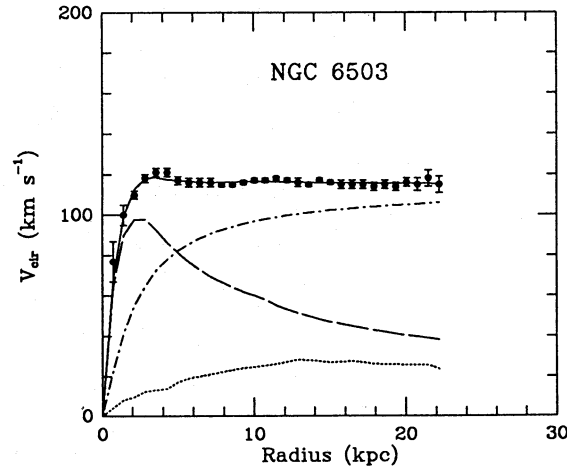
**Fig. 6:** *The rotation curve of galaxy NGC 6503. The different curves show the contribution of the three major components of matter to the gravitational potential (from top to bottom): halo, disk, gas. Figure taken from Ref. [7].*

curves requires more matter abundance, and differently distributed, than the visible one: a constant rotation curve requires $M_{\mathrm{DM}} \propto R$, which is attained for a DM density distribution $\rho_{\mathrm{DM}}(r) \propto 1/r^2$.

There exists several other dynamical constraints from studying the motion of stars in the Milky Way. They are then compared to a mass model for the galaxy and allow a determination of the local DM density (at the location of the Sun)

$$\rho(r_{\odot}) \simeq 0.01 M_{\odot}\,\mathrm{pc}^{-3} \simeq 0.4\,\mathrm{GeVcm}^{-3} \tag{132}$$

(recall 1 pc = $3.08 \times 10^{18}$ cm, and $1 M_{\odot} = 1.12 \times 10^{57}$ GeV).

### 5.1.3 *Large-scale structures*

Without DM, density perturbations would start to grow only after recombination, so today there would not be structures yet. Different DM types lead to different scenarios for the formation of structures: in the so-called Hot Dark Matter scenario large structures are formed first and then fragment into smaller pieces ("top-down"), while in the Cold Dark Matter scenario smaller objects merge into bigger structures hierarchically ("bottom-up"). Cosmological observations and numerical simulations exclude the Hot DM case.

### 5.2 Key Properties of Particle DM

A particle candidate for DM must satisfy at least the following fundamental properties:

1. stable, or at least with a lifetime longer than the age of the universe;
2. no electric charge, no color charge;
3. non-collisional, or at least much less collisional than baryons: self-annihilation cross sections must be smaller than QCD $\sigma_{\mathrm{DM\,DM}} \ll 1/m_p^2$, and weak cross sections $\sigma_{\mathrm{DM\,DM}} \ll 1/m_Z^2$;
4. not "hot", as it would be excluded by large-scale structure formation;
5. in the fluid limit, not in the form of a collection of discrete compact objects. We have not seen any discreteness effects in DM halos. Granularities would affect the stability of astrophysical systems. MAssive Compact Halo Objects (MACHOs) are astrophysical objects with macroscopic mass, such as large planets or small dead stars. Searches for MACHOs (EROS+MACHO results) exclude

the range

$$10^{-7} M_\odot \lesssim M \lesssim 10 M_\odot \,, \tag{133}$$

using gravitational microlensing. Several other constraints also apply, due to e.g. non-observation of lensing effects in the direction of Gamma-Ray Bursts (GRBs) or towards compact radio sources. However, it is natural to expect these objects to be baryonic and created in the late universe. So this would upset predictions for BBN and CMB and they are excluded. A small window for primordial black holes (PBH) is actually still open;

6. non-relativistic (classical) today, in order to be confined on galactic scales (1 kpc or so), for densities $\sim \text{GeV cm}^{-3}$, and velocities $\sim 100 \text{ km s}^{-1}$. This leads to lower limits on its mass, depending on whether DM is made out of bosons or fermions.

   – For *bosons*, the De Broglie wavelength $\lambda = h/p$ ($h$ is the Planck's constant, restored in this section) must be less than 1 kpc, so

   $$m \gtrsim \frac{h}{1\text{kpc} \cdot v} \simeq 10^{-22}\text{eV} \,, \tag{134}$$

   where $v \simeq 100$ km/s has been used.

   – For *fermions*, because of Pauli exclusion principle, the DM quantum occupation number must be smaller than one, so

   $$\rho(r_\odot) \lesssim \frac{m}{\lambda^3} \implies m \gtrsim \left[ \frac{h^3 \rho(r_\odot)}{v^3} \right]^{1/4} \simeq 1\,\text{keV} \,, \tag{135}$$

   where $\rho(r_\odot) = 0.4 \text{ GeV cm}^{-3} \simeq (0.04\,\text{eV})^4$ has been used (Gunn-Tremaine bound)

None of the SM particles satisfies the above requirements. Therefore the quest for a viable particle candidate for DM needs to be carried out in the realm of physics beyond the SM.

### 5.3 Weakly Interacting Massive Particles

There is a really wide landscape of DM models, where the DM mass spans several orders of magnitude, from ultra-light scalars at about $10^{-22}$ eV to primordial black holes at $10^{20}$ kg. There is no *a priori* preferred mass scale, so we are not sure where to look for DM.

Among the many possible categorizations of the DM models, one that is particularly useful is to divide the DM candidates into whether or not they are Weakly Interacting Massive Particles (WIMPs). The advantage is that all WIMPs share pretty much the same production mechanism in the early universe, through the so-called thermal *freeze-out*, which we will describe later, while each of the other non-WIMP DM candidates are produced in peculiar ways to be studied case-by-case.

Just to mention a few out of the many realizations of each category, the WIMPs can be the supersymmetric neutralino, minimal DM, Higgs-portal scalar, heavy neutrino, inert Higgs doublet, lightest Kaluza-Klein particle, etc. Some notable non-WIMP candidates are axions, sterile neutrinos, gravitinos, asymmetric DM, techni-baryons, Q-balls, primordial black holes, dark photons, topogical defects, etc.

For simplificy, from now on we will only focus on WIMPs. The basic ingredients for a WIMP model are:

   – a massive particle in the $\sim 1 \text{ GeV} - \sim 100$ TeV range;
   – weak interactions with the SM;
   – production via thermal freeze-out in the early universe.

### 5.4 Freeze-out of Thermal Relics

In this section we want to compute the thermal relic abundance of a particle whose interactions "freeze-out" in the early universe. We will first do a simple estimate, in order to highlight the important quantities into play, then we will describe a more formal calculation using the Boltzmann equations.

Let us start by assuming the Standard Model of particle physics is augmented with the inclusion of a particle $\chi$ (the DM) of mass $m_\chi$ such that

- $\chi$ is a stable;
- $\chi$ is coupled to lighter SM species;
- $\chi$ is in thermal equilibrium in the early universe at temperature $T \gg m_\chi$.

The DM particle $\chi$ is kept in equilibrium by number-changing annihilation processes of the kind

$$\chi\,\chi \leftrightarrow \text{SM SM}\,, \tag{136}$$

where SM is any SM particle. This follows from the assumptions 1. and 2. above.

At temperatures much bigger than $m_\chi$, these processes are fast and the DM is in equilibrium with the rest of the plasma. But as the unvierse expands, the rate for the processes (136) becomes slower than the expansion rate of the universe, and such reactions go out of equilibrium. This happens when the annihilation rate $\Gamma \lesssim H$. From this point on, the DM decouples from the plasma and its number density does not change anymore, it "freezes out".

Let us find the freeze-out temperature $T_f$, defined by the condition

$$n_\chi^{\text{eq}}(T_f)\sigma = H(T_f)\,. \tag{137}$$

Let us assume for simplicity here that the annihilation cross section of the process (136) does not depend on the relative velocity: $\sigma = \sigma_0$ (the so-called $s$-wave annihilation). Now, during radiation domination: $H(T_f) = \sqrt{(4\pi^3/45)g_*(T_f)}T_f^2/M_P$. The equilibrium number density, for $m_\chi \gg T_f$ is

$$n_\chi^{\text{eq}}(T_f) = g_\chi \left(\frac{m_\chi T_f}{2\pi}\right)^{3/2} e^{-m_\chi/T_f}\,. \tag{138}$$

Eq. (137) can be then manipulated to arrive at an implicit equation for $T_f$ which does not admit closed-form solution, but it can be solved iteratively, giving at leading order

$$T_f \simeq \frac{m_\chi}{\ln K}\,, \tag{139}$$

where

$$K \equiv \frac{3\sqrt{5}}{4\sqrt{2}\pi^3}\frac{g_\chi}{\sqrt{g_*}}\sigma_0 m_X M_P\,. \tag{140}$$

For reference values $m_X = 100\,\text{GeV}$, $g_X = 2$, $g_* = 100$ and $\sigma_0 = 1\,\text{pb} = 10^{-36}\,\text{cm}^2 \simeq 2.6 \times 10^{-9}\,\text{GeV}^{-2}$, we get

$$K = 2.4 \times 10^{10} \left(\frac{g_\chi}{2}\right)\left(\frac{100}{g_*}\right)^{1/2}\left(\frac{\sigma_0}{1\,\text{pb}}\right)\left(\frac{m_X}{100\,\text{GeV}}\right) \tag{141}$$

so $\ln K \simeq 24$, and therefore typically $T_f \sim m_\chi/20 \div m_\chi/30$.

Then it follows that the number density of $\chi$ at freeze-out is

$$n_\chi(T_f) = \frac{H(T_f)}{\sigma_0} = \sqrt{\frac{4\pi^3}{45}g_*(T_f)}\frac{T_f^2}{M_P} \tag{142}$$

which is roughly constant until today, up to a redhift dilution of non-relativistic matter

$$n_\chi(T_0) = \left(\frac{T_0}{T_f}\right)^3 n_\chi(T_f) \propto \frac{1}{T_f} \propto \frac{1}{m_\chi}\,. \tag{143}$$

So the energy density today of $\chi$ particles is $\rho_\chi(T_0) = n_\chi(T_0)m_\chi$ does not depend on $m_\chi$! Actually, there is still a mild (logarithmic) residual dependence on $m_\chi$ in $K$.

$$\Omega_\chi h^2 = \frac{\rho_\chi(T_0)}{\rho_c/h^2} = \frac{n_\chi(T_0)m_\chi}{\rho_c/h^2} \simeq 0.1 \frac{3 \times 10^{-26}\mathrm{cm}^3/\sec}{\sigma_0} \simeq 0.1 \frac{1\,\mathrm{pb}}{\sigma_0} \tag{144}$$

The present relic abundance of $\chi$ is mostly driven by its cross section. Notice that a pb cross section is the typical cross section of weak interactions.

Recall from Eq. (44) that the present energy density of DM is observed to be $\Omega_{\mathrm{DM}}h^2 \simeq 0.12$. Therefore, the relic density today of a cold relic produced by the freeze-out mechanism in the early universe can explain the observed value of the DM energy density, as shown by Eq. (144). If $\sigma_0$ is bigger than about $10^{-26}$ cm$^3$/s, the relic abundance is too big and would overclose the universe.

A cold relic with weak-scale interactions is a DM candidate. A typical annihilation cross section for a particle with couplings $g$ is $\sigma \sim g^4/M^2$, so a pb cross section is realized by

$$M/g^2 \sim \mathrm{TeV} \tag{145}$$

so the weak scale! A particle of weak-scale mass and couplings gives rise to a relic abundance in the right ballpark of the observed DM abundance. This remarkable coincidence is also known as the "WIMP miracle".

So there are several reasons why the WIMPs are so appealing as as DM candidates:

- the WIMP "miracle" (which may just be a numerical coincidence);
- a common production mechanism (freeze-out);
- the link with beyond-the-SM physics at the weak scale, possible related to the solution of the hierarchy problem (e.g. Supersymmetry)
- the possibility to perform multi-sided searches: the three pillars of WIMP searches are the so-called direct detection, indirect detection and collider searches; they may be interpreted as the searches for signatures due to three different realizations of the same WIMP-quark interactions: WIMP-quark scattering, WIMP self-annihilations and WIMP pair production from quarks. In the next subsections we will discuss each of them.

## 5.5 Direct Detection

Direct Detection (DD) of DM consists of looking for the scatterings of galactic halo DM on heavy nuclei in underground laboratories. Suppose a halo particle $\chi$ with mass $m_\chi$ and velocity $v$ scatters from a target nucleus at rest of atomic mass number $A$ and mass $M_A$ with an angle $\theta$ (in the c.o.m. frame). The c.o.m. recoil momentum, or momentum transfer, is

$$|\vec{q}|^2 = 2\mu_{\chi A}^2 v^2(1 - \cos\theta)\,, \qquad \mu_{\chi A} = m_\chi M_A/(m_\chi + M_A)\,. \tag{146}$$

The recoil energy imprinted on the nucleus is then

$$E_R = \frac{|\vec{q}|^2}{2M_A}\,. \tag{147}$$

which is maximum for $\theta = \pi$ ($|\vec{q}_{\max}| = 2\mu v$), so

$$E_R^{\max} = 2v^2 \frac{m_\chi^2 M_A}{(m_\chi + M_A)^2} = 2\frac{\mu_{\chi A}^2 v^2}{M_A} . \tag{148}$$

Such a recoil energy of the scattered nucelus can be measured and can signal the occurrence of a DM particle scattering by.

As an examples to get an idea of the orders of magnitude involved: for a DM particle with mass $m_\chi = 100$ GeV scattering off a $^{131}$Xe nucleus, we get

$$E_R^{\max} = 2 \left(\frac{v}{200 \text{ km/s}}\right)^2 \left(\frac{2}{3}10^{-3}\right)^2 \frac{100^2 \cdot 131}{231^2} 10^6 \text{ keV} \simeq 22 \text{ keV} \left(\frac{v}{200 \text{ km/s}}\right)^2 \tag{149}$$

so the recoil energies are typically in the $\mathcal{O}(1 \div 10)$ keV range. The experiments are able to tag the event and measure $E_R$ by directly observing one or two of the following 3 end-products: 1) heat; 2) ionization; 3) scintillation.

Let us make a back-of-the-envelope estimate of the expected number of events per unit of time. Consider a detector consisting of $N_T$ nuclei with mass number $A$ and mass $M_A \simeq A \cdot m_p \simeq A$ GeV. The total target mass of the detector is $M_T = N_T M_A$ (alternatively, the number density of target nuclei is $N_T = N_{\text{Avogadro}}/A$ ). Let $\sigma_{\chi A}$ be the nucleus-DM cross section, so

$$
\begin{aligned}
\frac{\# \text{ events}}{\text{time}} &= (\# \text{ targets}) \times (\text{WIMP flux on Earth}) \times (\text{cross section}) = N_T \left(\frac{\rho_\oplus}{M_\chi} v\right) \sigma_{\chi A} \\
&\simeq \frac{1 \text{ event}}{\text{yr}} \times \frac{M_T/A}{\text{kg}} \times \frac{\sigma}{10^{-39} \text{cm}^2} \times \frac{\rho_\oplus}{0.3 \text{ GeVcm}^{-3}} \times \frac{v}{200 \text{ km/s}} \times \frac{100 \text{ GeV}}{m_\chi} .
\end{aligned}
$$

More precisely, the spectrum of events per recoil energies is given by

$$\frac{\mathrm{d}R}{\mathrm{d}E_R} = N_T \int_{|\vec{v}| > v_{\min}} |\vec{v}| \frac{\mathrm{d}\sigma_{\chi A}}{\mathrm{d}E_R} \mathrm{d}n_{\text{DM}} = N_T \frac{\rho_\oplus}{m_\chi} \int_{|\vec{v}| > v_{\min}} \mathrm{d}^3 v |\vec{v}| f(\vec{v}, t) \frac{\mathrm{d}\sigma_{\chi A}}{\mathrm{d}E_R} \tag{150}$$

where we inserted the differential particle density

$$\mathrm{d}n_{\text{DM}} = \frac{\rho_\oplus}{m_\chi} f(v) \mathrm{d}^3 v \tag{151}$$

with $f(v)$ being the velocity distribution and $v_{\min} = \sqrt{M_A E_R^{\text{th}}/(2\mu_{\chi A}^2)}$ is the minimal DM velocity needed to transfer a threshold kinetic energy $E_R^{\text{th}}$ to the nucleus.

The most recent results for SI cross sections are from Xenon1T experiment [8] (about 2 tons of liquid Xe) are shown in Fig. 7. The SD cross section is much less constrained, a few orders of magnitude weaker bound than SI.

The characteristic shapes of the bounds can be understood as follows. The total event rate turns out to be proportional to $R \propto \sigma \mu_{\chi A}^2/m_\chi < R_{\text{observed}}$. Therefore a bound on the total number of observed events translates into a limit on the coupling

$$\sigma < \sigma_{\text{bound}} \propto \frac{m_\chi}{\mu_{\chi A}^2} \sim \begin{cases} m_\chi^{-1} & (m_\chi \ll m_A) \\ m_\chi & (m_\chi \gg m_A) \end{cases} \tag{152}$$

This dependence explains the typical exclusion curves shown by the experimental collaborations, and have a dip (maximal exclusion) around $m_\chi \simeq m_A$ where the reduced mass is maximal.

The vector interactions mediated by $Z$ exchange would typically lead to a spin-independent cross section $\sigma \sim \alpha_W^2 m_p^2/M_Z^4 \approx 10^{-39}$ cm$^2$, which is already excluded by orders of magnitude.
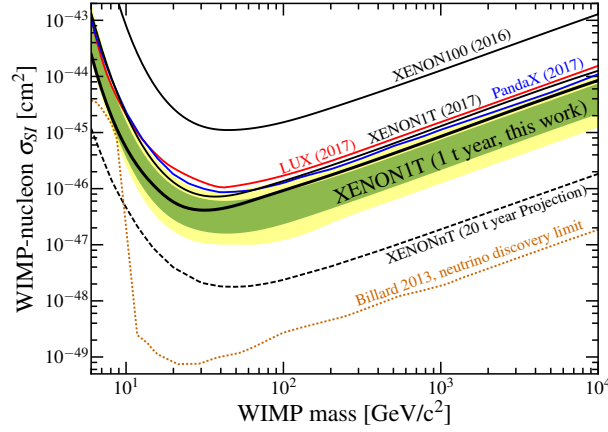
**Fig. 7:** *90% confidence level upper limit on spin-indepdenent WIMP-nucleon cross section from Xenon1T. A comparison with previous Xenon results, LUX and PandaX experiments is also shown. Figure from Ref. [8]*

## 5.6 Indirect Detection

The indirect searches for DM are based on identifying excesses in fluxes of gamma rays/cosmic rays with respect to their presumed astrophysical backgrounds. These stable Standard Model particles may be the end product of the annihilation (or decay) of DM in the galactic halo or in the Sun. The schematic chain of processes leading from DM self-annihilations to observable fluxes at Earth is

$$\chi\chi \to \text{SM SM} \xrightarrow{\text{hadron./decay}} \text{stable species} \xrightarrow{\text{astrophys. prop.}} \text{fluxes at Earth} . \tag{153}$$

Promising sources of DM annihilations are generically the regions where DM is expected to be the densest, such as the galactic center, the inner halo of our Galaxy, nearby galaxies dominated by DM, the center of the Sun, the center of the Earth. However, in some of these regions it is usually very complicated to understand the underlying astrophysics. So the best detection opportunities might come from selecting targets which are not necessarily the richest in DM but with well-identified backgrounds (favourable signal/background ratio). This also depends on which species of cosmic ray one is looking for.

The first step of the chain (153) DM annihilations into primary channels (like $q\bar{q}, \ell^+\ell^-, W^+W^-$, etc.) is controlled by the DM model lagrangian describing the elementary interactions of the DM particle with the SM. Once the primary products of annihilations are produced, they will undergo standard SM evolution, like decay, radiation, hadronization, controlled by QED, EW and QCD interactions. The end-product of this step is to have stable particle species (e.g. $e^\pm, \gamma, \nu, p$ etc.). Such stable particles are then travelling through the galaxy from their production point to the Earth, and they are subject to a number of astrophysical processes. Finally, the result of the astrophysical propagation of stable particles is the fluxes at detection (Earth) which is what can be measured.

This chain has to be reversed in order to extract information on the original DM model from observations of the fluxes. As it is clear, in this inversion process a lot of uncertainties come into play, especially those from the astrophysical propagation mechanisms.

The SM particles giving best information are photons, neutrinos and stable anti-particles: e.g. positrons and anti-protons (also, maybe, anti-deuteron, anti-helium). Why anti-matter? Because there is little anti-matter from early universe and (possibly) little anti-matter in primary cosmic rays. Observations provide a positron fraction of the order $e^+/(e^+ + e^-) \sim 0.1$ and a antiproton-to-proton ratio of the order $\bar{p}/p \sim 10^{-4}$. Each stable species has advantages and disadvantages to be used as a DM indirect detection probe:

- **Photons.** They freely propagate, in the galactic environment. However DM is electrically neutral,
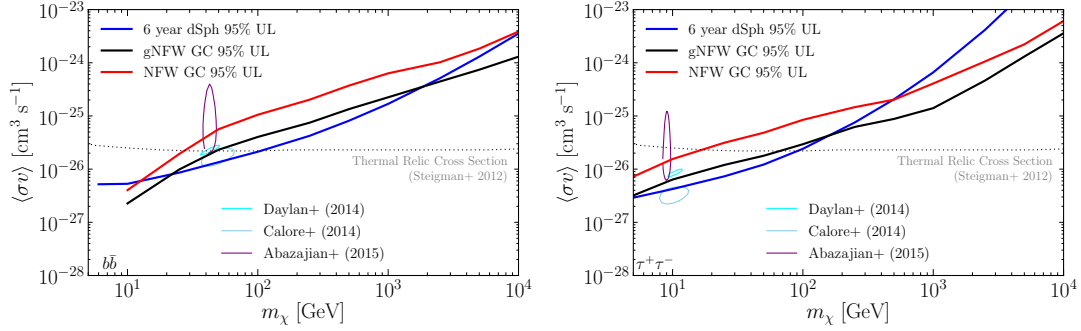
**Fig. 8:** *DM upper limits on DM self-annihilation cross section from Fermi-LAT observations of the Galactic Center, as a function of the DM mass, for annihilations into $b\bar{b}$ (left panel) and $\tau^+\tau^-$ (right panel). The results are shown at the 95% confidence level assuming the generalized NFW (black) and NFW (red) DM profiles. The upper limits from the recent analysis of 15 dwarf spheroidal galaxies using 6 years of Fermi-LAT data are shown in blue. The dotted line indicates the thermal relic cross section. Figures taken from Ref. [9].*

so that photons can be produced only via some subdominant mechanism (e.g. loops) or as secondary radiation (synchrotron, bremsstrahlung). The spectrum is suppressed, and the astrophysical background difficult.

– **Positrons.** They diffuse in the galactic magnetic fields with energy losses due to: synchrotron emission, Coulomb scattering, ionization, bremsstrahlung and inverse Compton. The DM contribution is dominated by the nearby regions of the galaxy. Below a few GeV, the effect of solar activity is important.

– **Anti-protons.** They diffuse in the galactic magnetic fields with negligible energy losses, until they scatter on matter. Therefore even far-away regions of the Galaxy can contribute to the flux collected on Earth and, as a consequence, its normalization has significant astrophysical uncertainties. Below a few GeV, the effect of solar activity is important.

– **Neutrinos.** TeV-scale neutrinos propagate freely in the Galaxy and can also propagate through the dense matter of the Sun and the Earth. Neutrinos are difficult to detect, they are measured indirectly via the detection of charged particles (e.g. muons) produced by a neutrino interaction in the rock or water surrounding a neutrino telescope. The incoming neutrino energy can only be partially reconstructed.

As an example, in Fig. 8 we show the upper limits on the self-annihilation cross section of DM from gamma-ray observations, in two different annihilation channels: $b\bar{b}$ and $\tau^+\tau^-$. The thermal relic cross section sets the reference to exclude models giving lower cross section (that would lead to too much DM abundance today), so one can exclude (e.g. using 6-year data on dwarf spheroidal galaxies) DM annihilating into $b\bar{b}$ or $\tau^+\tau^-$ with masses $m_\chi \lesssim 100$ GeV.

## 5.7 Collider Searches

How does DM (a WIMP) show up in a collider, such as the LHC? A WIMP must be stable (over collider scales) and very weakly interacting. So, even if a WIMP is produced in a high-energy collision, it escapes the detectors with no interaction, thus leaving no visible tracks. The DM behaves exacly like a neutrino, for collider purposes, so its unavoidable signal is just "missing energy". This implies that the irreducible background of the DM searches (and very often the dominant background) is due to $Z \to \nu\bar{\nu}$ (e.g. with $Z$ produced via Drell-Yan process).

If the DM is stabilized by an exact $Z_2$ symmetry under which it is odd, while the SM is even, then DM must be produced in pairs.

Since the missing energy alone is a rather poor signal, one needs something else as a handle to select events involving DM production. The identification of the most suitable extra handle, is a model-dependent issue. It may be jets or other objects from initial state radiation, accompanying particles, displaced vertices, etc.

At this point, an important *caveat* is in order: the LHC cannot discover the DM. It may only discover a weakly interacting particle with lifetime larger than the size of the detecting, but there is not way to test the stability of the escaping particles on cosmological scales.

So far, there has been no signal for DM at LHC. There may be three reasons for that:

1. DM may not interact with ordinary matter: indeed, we are only sure that DM has gravitational interactions;

2. DM physics may not be accessible by LHC: e.g. DM may be too light/heavy or interacting too weakly with ordinary matter;

3. we may not have explored all the possibilities: DM may be buried under large backgrounds or hiding behind unusual/unexplored signatures.

The simplest handle to correlate with missing transverse energy (MET) is to consider the Initial State Radiation (ISR) of some SM particle $X$, where $X$ may be a quark/gluon (producing a jet in the final state), a photon, a W/Z, or even a Higgs. This class of signatures are called mono-X searches. However, the mono-jet is what provides the strongest limits in most situations.

Mono-X+MET searches have the virtue of being rather general, the backgrounds are relatively well-known and they provide complementary/competitive results with direct detection. The main drawbacks are that some background is irreducible, there is a small signal-to-background ratio and the searches are limited by systematics.

Whether or not one chooses to explore DM at LHC using the mono-X signal, one very important question to ask is: which DM model to test? The interaction between quarks and DM can be modelled in many different ways. One can nonetheless divide the infinite-dimensional space of DM and Beyond-the-Standard-Model (BSM) theories into three broad categories:

1. Complete models. These are models of BSM physics like Supersymmetry, Composite Higgs, etc. which provide a valid description of elementary particles up to very high energies (ultraviolet complete), typically including also a solution of the hierarchy problem.
   Pros: they are ultraviolet (UV) complete, motivated by BSM issues (like the hierarchy problem). Cons: they have many parameters, and typically include sources of fine tuning.

2. Effective operators. They are coming from integrating out whatever heavy physics is responsible for mediating the SM-DM interactions, e.g. the heavy mediator. This approach has been often considered as "model-independent", but it is not, since one needs to specify up the energy cutoff up to which the Effective Field Theory (EFT) is valid. This depends on the UV completion.
   Pros: it is an economical approach (no need to specify mediators) and provide a common language to compare results from different experiments, e.g. direct/indirect detection. Cons: they are less complete than complete models; EFT is not always applicable, especially at very high energies involved in LHC processes.

3. Simplified models. They are a sort of mid-way between the two extremes described above. A heavy mediator particle, mediating the interactions between SM and DM, is exchanged in the $s$- or $t$-channel (for a review, see e.g. Ref [10]).
   If the DM sector is more complicated than just an extra particle coupled to the SM, the heavy mediator approach is however a simple and good enough representation of what is going on. Simplified models are therefore simple versions of more complicated theories but with only the minimum amount of degrees of freedom which are necessary to model the physical process of interest.
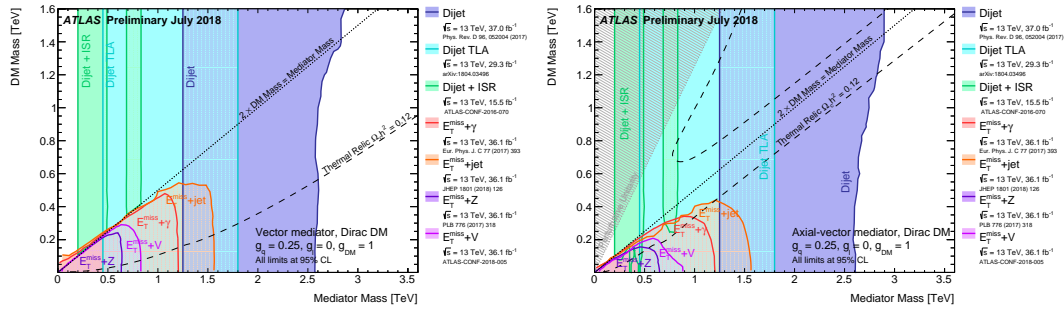
**Fig. 9:** *Regions in a DM mass-mediator mass plane excluded at 95% CL by a selection of ATLAS dark matter searches, for vector mediator (left panel) and an axial-vector mediator (right panel) mediating the interactions between the SM and DM. Figures taken from ATLAS summary plots.*

Pros: they provide a good representation of more complicated situations with minimum number of degrees of freedom; they are theoretically consistent. Cons: they require extra propagating degrees of freedom, beyond just the DM particle, so more parameters than EFTs; with a single simplified model it is hard to catch all phenomenological possibilites of complete models.

The great advantage of simplified models is that one can combine the search for the DM with the search for the mediator itself, e.g. looking for it as a di-jet resonance (see Fig. 9). As it is clear from Fig. 9, the di-jet searches for the mediators are actually setting stronger limits than the mono-X + MET searches.

The limits in the DM mass-mediator mass plane can be recast into constraints on WIMP-nucleon scattering cross sections, to be compared with those from direct detection experiments. The main result is that for spin-independent couplings, the direct detection experiments set the most stringent limits (for $m_\chi \gtrsim 5$ GeV), but for spin-dependent couplings (for which direct detection is weaker) the LHC bounds are actually stronger.

Is this the whole story? The negative results of DM searches are calling for new efforts towards developing other tools and methodologies to increase the power of the searches. A couple of topics along this direction which are worth mentioning are: exploring less conventional, unexplored phenomenological signatures for DM; use data-driven approaches (e.g. machine learning) to get new and deeper views into the available and upcoming data.

## 6 Inflation

Inflation [11–13] is one of the basic ideas of modern cosmology and has become a paradigm for the physics of the early universe. In addition to solving the shortcomings of the standard Big Bang theory, inflation has received a great deal of experimental support, for example it provided successful predictions for observables like the mass density of the universe and the fluctuations of the cosmic microwave background radiation. Before discussing inflation in more detail, let us first review some background material about standard cosmology, which serves also to introduce the notation, and outline its major shortcomings.

### 6.1 Shortcomings of Big Bang Cosmology

**Flatness problem.** Recall from Eq. (45) that curvature parameter $\Omega_k$ is of the order of 0.1% ($10^{-3}$) today, which means that to a very good approximation we live in a flat universe. Let us define the total Omega parameter as

$$\Omega \equiv \frac{\rho_{\text{tot}}}{\rho_c} = \Omega_m + \Omega_r + \Omega_\Lambda \,. \tag{154}$$

From the sum rule of Omega parameters in Eq. (35) we get that

$$\Omega - 1 = \frac{k}{a^2 H^2} \, . \tag{155}$$

The parameter $|\Omega - 1|$ grows with time during radiation- and matter-dominated eras. In particular, during radiation domination $H^2 \propto \rho_{\text{radiation}} \propto a^{-4}$, so

$$|\Omega - 1| \propto \frac{1}{a^2 H^2} \propto \frac{1}{a^2 a^{-4}} \propto a^2 \, . \tag{156}$$

Let us extrapolate this value back in time until the Planck time $t_P \sim 10^{-43}$ s

$$\frac{|\Omega - 1|_{T=T_P}}{|\Omega - 1|_{T=T_0}} \approx \left( \frac{a(t_P)}{a(t_0)} \right)^2 \approx \left( \frac{T_0}{T_P} \right)^2 \approx 10^{-64} \, . \tag{157}$$

Since we observe today that the energy density of the universe is very close to the critical density (i.e. a very small $\Omega_k$) the $\Omega$ parameter must have been close to unity to an extremely high accuracy (of about one part in $10^{64}$ if we start the radiation-dominated era at the Planck time. Therefore, an extreme degree of fine tuning is necessary to arrange such a precise initial value of the density parameter of the universe. This is the flatness (or fine-tuning) problem.

**Entropy problem.** The flatness problem is also connected to the entropy problem, which is understanding why the total entropy of the visible universe is incredibly large. In fact, recall that the entropy in a comoving volume of radius $a$ and temperature $T$ is $S \simeq (aT)^3 =$ which is constant, and today the entropy within the horizon is

$$S_0 \sim H_0^{-3} s_0 \sim H_0^{-3} T_0^3 \sim 10^{90} \, , \tag{158}$$

which is huge with respect to that in the early universe. During radiation domination the Hubble parameter is $H \sim T^2 / M_P$, where the Planck mass is $M_P \equiv G^{-1/2} = 1.22 \times 10^{19}$ GeV, so Eq. (155) can be re-written as

$$|\Omega - 1| \propto 1 a^2 H^2 \propto \frac{1}{a^2 T^4} \propto \frac{1}{T^2 S^{2/3}} \, . \tag{159}$$

This relation tells us that $\Omega$ at early times is so close to 1 because the total entropy of the universe is enormous. For example, at the Planck scale, the entropy of $10^{90}$ corresponds to $\Omega - 1 \sim 10^{-60}$.

**Horizon problem.** As already mentioned in Section 4.3, the CMB has an amazingly high degree of homogeneity, about one part in $10^5$. But this poses a serious problem for cosmology. Recall from Section 2.2.8 that the particle horizon is the distance travelled by photons. Let us consider our current particle horizon $d_0$ and track it back in time to the time of photon decoupling (last scattering), when CMB formed $T_{\text{dec}} \sim 0.3$ eV. The CMB temperature today is $T_0 \simeq 2.3 \times 10^{-4}$ eV, so particle horizon at dec

$$\lambda_H|_{\text{dec}} = d_0 \frac{a_{\text{dec}}}{a_0} = d_0 \frac{T_0}{T_{\text{dec}}} \, . \tag{160}$$

From the Friedmann equation we know that during matter domination (from photon decoupling to today) the Hubble radius, i.e. the size of the observable universe, redshifts as $a^{-3/2} \sim T^{3/2}$. Therefore

$$\left( \frac{\lambda_H|_{\text{dec}}}{H_{\text{dec}}^{-1}} \right)^3 = \left( \frac{T_{\text{dec}}}{T_0} \right)^{3/2} \approx 10^5 \, . \tag{161}$$

So this result is telling us that when CMB formed, the lenght $\lambda_H$ corresponding to our unviverse today was much larger (by a factor $10^5$) than the size of the causally connected universe at that time ($H_{\text{dec}}^{-1}$). So at photon decoupling there were $10^5$ causally disconnected regions that now correspond to our horizon! In other words, the photons received today were emitted from regions that were causally disconnected

at the time of photon decoupling, because they were out of the particle horizon. Why regions that were not in causal contact share the same temperature to a very high precision? This is the so-called horizon problem.

**Monopole problem.** Lastly, another issue that was plaguing the Standard Big Bang Cosmology in the context of Grand Unified Theories (GUTs) is the overproduction of magnetic monopoles. Indeed, magnetic monopoles are a generic prediction of GUTs and they are produced at a phase transition at $T = T_c$, after which they behave as non-relativistic matter. To estimate the number density of monopoles, we consider the simple argument of expecting approximately 1 monopole per correlation volume $\ell_{\rm cor}^3$, where the correlation length is bounded by the horizon at the critical temperature $\ell_{\rm cor} \lesssim H(T_c)^{-1}$. Therefore, the number density of monopoles is roughly given by

$$n_M \simeq \ell_{\rm cor}^{-3} \gtrsim H(T_c)^3 \simeq \left(1.66 \sqrt{g_*(T_c)}\right)^3 \frac{T_c^6}{M_P^3} \quad \Longrightarrow \quad \frac{n_M}{s} \sim \sqrt{g_*(T_c)} \left(\frac{T_c}{M_P}\right)^3 \tag{162}$$

their number density behaves like $n_M(t) \propto a^{-3}(t) \propto s(t)$, and therefore using (162) we can estimate the abundance of magnetic monopoles today as

$$\begin{aligned}
\rho_M(T_0) &= m_M n_M(T_0) = m_M \frac{n_M(T_c)}{s(T_c)} s(T_0) \sim m_M \sqrt{g_*(T_c)} \left(\frac{T_c}{M_P}\right)^3 g_*(T_0) T_0^3 \\
&\sim 10^{12} \left(\frac{m_M}{10^{16}\,{\rm GeV}}\right) \left(\frac{T_c}{10^{16}\,{\rm GeV}}\right)^3 \sqrt{\frac{g_*(T_c)}{10^2}}\,{\rm GeV\,cm}^{-3}
\end{aligned} \tag{163}$$

while the critical density is $\rho_c \simeq 10^{-5}\,{\rm GeV\,cm}^{-3}$, so

$$\frac{\rho_M}{\rho_c} \sim 10^{17}\,, \tag{164}$$

for monopoles with GUT-scale mass ($\sim 10^{16}$ GeV). This overabundance of magnetic monopoles is the so-called monopole problem. Therefore, one should suppose either that the universe was never at temperatures as high as $T_c \sim 10^{16}$ GeV, or that Grand Unification is not there. The monopole problem was the primary motivation behind the idea of inflation.

## 6.2 The inflationary solution

Inflation elegantly solves at once the problems associated with the standard Big Bang cosmology. The inflationary era is defined as the epoch in the early history of the universe when it underwent a period of accelerated expansion

$$\ddot{a} > 0\,. \tag{165}$$

According to Eq. (25), this condition is equivalent to $\rho + 3p < 0$ (for negligible cosmological constant). For the sake of simplicity, we shall only consider here a more stringent condition for inflation, $p = -\rho$ (negative pressure!). This condition is also known as de Sitter phase, and corresponds to constant energy density and Hubble parameter $H_I$, and thus the scale factor grows exponentially in time

$$a(t) \propto e^{H_I t}\,. \tag{166}$$

Inflation delivers a flat universe, thus providing an explanation for the initial condition that $\Omega$ is close to 1 to a high precision. In fact, during inflation, the Hubble rate is nearly constant and the curvature parameter $\Omega - 1$ is proportional to $1/a^2$ (see Eq. (155)), thus its final value at the end of inflation $t = t_f$ is related to the primordial initial value at $t = t_i$ by

$$\frac{|\Omega - 1|_{\rm final}}{|\Omega - 1|_{\rm initial}} = \left(\frac{a(t_i)}{a(t_f)}\right)^2 = e^{-H_I(t_f - t_i)}\,. \tag{167}$$

If inflation lasts for long enough, the $\Omega$ parameter will be exponentially driven to unity. Therefore, the universe emerging at the end of inflation is spatially flat to a very high accuracy.

Furthermore, the large amount of entropy produced during the non-adiabatic phase transition from the end of inflation and the beginning of the radiation-dominated era also produces a huge entropy

$$\frac{S_f}{S_i} \sim \left(\frac{a(t_f)}{a(t_i)}\right)^3 \left(\frac{T_f}{T_i}\right)^3 \sim e^{3H_I(t_f - t_i)} \left(\frac{T_f}{T_i}\right)^3 . \tag{168}$$

Therefore, a period of exponential expantion can easily account for a large amount of entropy and it can greatly dilute all magnetic monopoles down to an unobservable level.

If the universe underwent a period when the physical scales evolve faster than the horizon scale, it is possible to make the CMB photons in causal contact at some primordial time before the photon decoupling. The physical size of a perturbation grows as the scale factor: $\lambda \sim a$, while the horizon scale is $H^{-1} = a/\dot{a}$. If a period exists in the early history of the universe when

$$\frac{d}{dt}\frac{\lambda}{H^{-1}} = \ddot{a} > 0 , \tag{169}$$

the CMB photons may have been in causal contact at that time, thus explaining the high level of homogeneity and isotropy observed in the CMB today. Such an epoch of accelerated expansion is precisely the inflationary stage.

The mechanism of inflation can be simply realized by means of a scalar field $\phi$, called the inflaton, whose energy is dominant in the universe and with potential energy $V(\phi)$ much larger than the kinetic energy. The generic lagrangian for the inflaton is

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi) , \tag{170}$$

while the energy-momentum tensor is

$$T^{\mu\nu} = \partial^\mu\phi\partial^\nu\phi - g^{\mu\nu}\mathcal{L} . \tag{171}$$

Neglecting the spatial gradients, the 00 and $ii$ components of the energy-momentum tensor, corresponding to the energy density and the pressure of the inflaton respectively, are given by

$$T^{00} \quad = \quad \rho_\phi = \frac{\dot{\phi}^2}{2} + V(\phi) , \tag{172}$$

$$T^{ii} \quad = \quad p_\phi = \frac{\dot{\phi}^2}{2} - V(\phi) . \tag{173}$$

If the kinetic energy is negligible with respect to the potential energy

$$V(\phi) \gg \dot{\phi}^2 , \tag{174}$$

and if the energy density of the inflaton dominates over other forms of energy density (such as matter or radiation), then we would have a de Sitter stage $p_\phi = -\rho_\phi$ and the Friedmann equation would read

$$H^2 \simeq \frac{8\pi G_N}{3}V(\phi) . \tag{175}$$

Thus, inflation is driven by the vacuum energy of the inflaton field.

The equation of motion of the inflaton field in an expanding universe is

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0 , \tag{176}$$

where the prime refers to the derivative with respect to $\phi$. When $V(\phi) \gg \dot{\phi}^2$ and $\ddot{\phi} \ll 3H\dot{\phi}$, the scalar field "slowly rolls" down its potential. Under the slow-roll conditions, the equation of motion reduces to

$$3H\dot{\phi} \simeq -V'(\phi)\,. \tag{177}$$

It is straightforward to derive some important relations from the two equations in Eqs. (175) and (177) and the slow-roll condition in Eq. (174). From Eqs. (175) and (177) one can show that

$$\dot{H} = -4\pi G_N \dot{\phi}\,, \tag{178}$$

while using Eqs. (174) and (177) one obtains

$$\frac{(V')^2}{V} \ll H^2\,, \tag{179}$$

and finally using Eqs. (174), (175), (177) and (178) we arrive at

$$V'' \ll H^2\,. \tag{180}$$

It is customary to define the "slow-roll parameters" $\epsilon, \eta$ as

$$\epsilon \;\equiv\; \frac{1}{16\pi G_N}\left(\frac{V'}{V}\right)^2\,, \tag{181}$$

$$\eta \;\equiv\; \frac{1}{8\pi G_N}\left(\frac{V''}{V}\right)^2\,. \tag{182}$$

in such a way that the conditions Eqs. (179)-(180) derived by the slow-roll regime can be simply recast into $\epsilon \ll 1, |\eta| \ll 1$. Furthermore, from the Friedmann equation (175) and Eq. (178), one can rewrite the $\epsilon$ parameter as

$$\epsilon = -\frac{\dot{H}}{H^2} \tag{183}$$

which allows one to express the second derivative of the scale factor in terms of $\epsilon$

$$\frac{\ddot{a}}{a} = \dot{H} + H^2 = (1-\epsilon)H^2 > 0 \Longleftrightarrow \epsilon < 1\,. \tag{184}$$

So the condition defining inflation $\ddot{a} > 0$ is equivalent to $\epsilon < 1$, and inflation ends when $\epsilon \simeq 1$.

## 6.3 Consequences of inflation

**Spectral Parameters.** As the inflaton rolls down its potential energy, it undergoes two kind of fluctuations: a classical one and a quantum one. During a Hubble time $H^{-1}$, these fluctuations behave as

$$(\delta\phi)_{\text{cl}} \;\sim\; \dot{\phi} H^{-1}\,, \tag{185}$$

$$(\delta\phi)_{\text{qu}} \;\sim\; H/(2\pi)\,. \tag{186}$$

The so-called power spectrum of scalar pertubations is given by the ratio of these two kinds of fluctuations at a momentum scale $k$ equal to the horizon scale $aH$

$$\mathcal{P}(k) = \left[\frac{(\delta\phi)_{\text{qu}}}{(\delta\phi)_{\text{cl}}}\right]^2 = \left(\frac{H}{\dot{\phi}}\right)^2 \left(\frac{H}{2\pi}\right)^2 \bigg|_{k=aH}\,, \tag{187}$$

and after some manipulations we arrive at the expression in terms of the slow-roll parameter $\epsilon$

$$\mathcal{P}(k) = \frac{8G_N^2}{3}\frac{V}{\epsilon}\bigg|_{k=aH} \tag{188}$$

The spectral index $n_s$ is defined as

$$n_s - 1 \equiv \frac{\mathrm{d} \ln \mathcal{P}(k)}{\mathrm{d} \ln k} \tag{189}$$

which can be interpreted as the exponent of the $k$-dependence of the power spectrum $\mathcal{P}(k) \propto k^{n_s - 1}$. It is easy to show that

$$\frac{\mathrm{d}}{\mathrm{d} \ln k} = -\frac{1}{8\pi G_N} \frac{V'}{V} \frac{\mathrm{d}}{\mathrm{d} \phi} \tag{190}$$

from which it follows that

$$\frac{\mathrm{d}\epsilon}{\mathrm{d} \ln k} = -2\epsilon\eta + 4\epsilon^2 \tag{191}$$

and finally the spectral index in terms of the slow-roll parameters is

$$n_s = 1 - 6\epsilon + 4\eta \,. \tag{192}$$

So, in slow-roll inflation where $\epsilon, |\eta| \ll 1$, the spectral index is very close to 1, meaning that the spectrum of scalar perturbations is nearly scale-independent.

Other kinds of perturbations are the so-scalled tensor perturbations (or gravity waves), whose power spectrum turns out to be

$$\mathcal{P}_g = \left. \frac{128\pi G_N^2}{3} V \right|_{k=aH} \,, \tag{193}$$

from which one can derive the important tensor-to-scalar ratio $r$

$$r = \frac{\mathcal{P}_g}{\mathcal{P}} = 16\epsilon \ll 1 \,, \tag{194}$$

which is also predicted as very small in slow-roll inflation.

**Evolution of Perturbations.** The Fourier expansion of inflaton field fluctuations in $k$-modes can be written as

$$\delta\phi(\mathbf{x}, t) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}} \delta\phi_k(t) \,, \tag{195}$$

and the $k$-modes obey the equation of motion

$$\ddot{\delta\phi}_k + 3H\dot{\delta\phi}_k + \frac{k^2}{a^2}\delta\phi_k = 0 \,. \tag{196}$$

This can be studied more easily in two extreme regimes, according to whether the modes are inside or outside the horizong. The modes inside the horizon are characterized by a length scale $\lambda \propto (a/k) \ll H^{-1}$, which is equivalent to the condition $k \gg aH$, so the equation of motion reads

$$\ddot{\delta\phi}_k + \frac{k^2}{a^2}\delta\phi_k = 0 \,. \tag{197}$$

This is a simple harmonic oscillator with $\delta\phi_k \propto \lambda^{-1}$, so fluctuations are stretched during inflation.

The modes outside the horizon are characterized by a length scale $\lambda \propto (a/k) \gg H^{-1}$, which is equivalent to $k \ll aH$, so the equation of motion reads

$$\ddot{\delta\phi}_k + 3H\dot{\delta\phi}_k = 0 \,. \tag{198}$$

This is an oscillator with friction, and the corresponding fluctuations are constant ("frozen").

So the fluctuations of the inflaton field grow exponentially during inflation, until their wavelength exits the horizon; then fluctuations get frozen outside the horizon; after inflation ends, fluctuations re-enter the horizon (see Figure 10 for a pictorial representation).
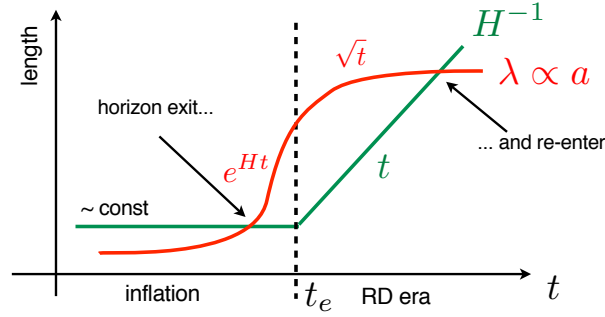
**Fig. 10:** *Evolution of the length scales ($\lambda \propto a$) during and after inflation, in red. For comparison, in green it is shown the evolution of the Hubble scale $H^{-1}$, during inflation and after inflation (in the radiation-dominated epoch).*

**CMB and Large-scale structures.** Inflation can also be responsible for the physical processes giving rise to the CMB anisotropies and the matter structures we observe in the universe today. In fact, primordial small quantum fluctuations of the energy density are excited during inflation and stretched to cosmological scales; then they exit the horizon and get frozen; when they re-enter the horizon at some matter- or radiation-dominated epoch, these fluctuations will start growing giving rise to the formation of all the structures we observe.

Physically, the mechanism works because the fluctuations are connected to the metric perturbations (gravity) via Einstein's equations and gravity acts as a messenger: once a given wavelength re-enters the horizon, gravity communicates the perturbations to baryons and photons. Therefore, the primordial quantum fluctuations of the inflaton field during inflation provide the seeds of the CMB temperature fluctuations and the large-scale strucutres observed today.

## 7 Baryogenesis

Our universe has a matter-antimatter asymmetry. We observe our universe to consist of matter, and not antimatter in appreciable quantities. More precisely, the difference between the number density of baryons and that of anti-baryons is expressed in terms of the baryon-to-photon ratio today

$$\eta \equiv \left. \frac{n_B - n_{\bar{B}}}{n_\gamma} \right|_0 , \tag{199}$$

(recall that the photon number density is $n_\gamma = 2\zeta(3)T^3/\pi^2$). The accurate measurement of the matter-antimatter asymmetry has been mainly provided by two independent and solid types of experiments.

– Big Bang Nucleosynthesis. We have already discussed in Section 3 that the simultaneous fit to primordial element abundances in terms of the single free parameter $\eta$ is a remarkable success of standard cosmology and provides

$$5.2 \times 10^{-10} < \eta < 6.6 \times 10^{-10} \qquad (95\% \text{ CL}) . \tag{200}$$

– Cosmic Microwave Background. The position and height of acoustic peaks in the power spectrum of CMB temperature anisotropies, probing the baryon/photon fluid at the last scattering surface, allow us to constrain the baryon energy density and therefore $\eta$ [1]:

$$\eta = (6.13 \pm 0.04) \times 10^{-10} , \tag{201}$$

The agreement of these two independent measurements is evident. Within the standard cosmological model $\eta$ is not predicted, it is a free parameter whose value is fixed by observations. Explaining this number is challenging, and a definitive answer is still missing.

If there was an era of cosmological inflation, any initial asymmetry would have been diluted by the enormous entropy increase during such epoch; hence, at the end of inflation the universe looks perfectly symmetric. Therefore, explaining the origin of the tiny (but non-zero) asymmetry we observe today requires that some post-inflationary mechanism is at work. The mechanism by which a baryon asymmetry is dynamically produced in the early universe is generically called baryogenesis.

In 1967 Sakharov pointed out three necessary conditions for a baryon asymmetry to be produced in the early universe and observed today:

I. Baryon number violation.
   This condition is quite obvious. Let us suppose to start from a baryon symmetric universe $B(t_0) = 0$, at a certain $t_0$. The quantum mechanical evolution of the operator $B$ is $B(t) \propto \int_{t_0}^{t} [B, H] dt'$, where $H$ is the hamiltonian of the system. If $B$ is conserved, $[B, H] = 0$ and then $B(t) = 0$ at all times.

II. $C$ and $CP$ violation.
   If $C$ were an exact symmetry, the probability of the process $i \to f$ would be equal to the one of the conjugated process $\bar{i} \to \bar{f}$. Therefore the same amount of $f$ and $\bar{f}$ would be present in the final state. But $B$ is odd under $C$, so $B(\bar{f}) = -B(f)$ and so the net baryon number $B$ would vanish.
   Due to the $CPT$ theorem, $CP$ invariance is equivalent to $T$ invariance and this implies that the probability of the process $i(\mathbf{r}_i, \mathbf{p}_i, \mathbf{s}_i) \to f(\mathbf{r}_j, \mathbf{p}_j, \mathbf{s}_j)$ is equal to that of the time-reversed process $f(\mathbf{r}_j, -\mathbf{p}_j, -\mathbf{s}_j) \to i(\mathbf{r}_i, -\mathbf{p}_i, -\mathbf{s}_i)$, where $\mathbf{r}_i, \mathbf{p}_i, \mathbf{s}_i$ denote coordinate, momentum and spin of the $i$-th particle, respectively. After performing an integration over all momenta and summation over all spins, the total baryon asymmetry vanishes.

III. Departure from thermal equilibrium.
   Let us consider a species $\psi$ carrying baryon number and being in thermal equilibrium, and distinguish the situations when it does or does not have a chemical potential.
   If $\psi$ has zero chemical potential, the CPT invariance implies that particles and anti-particles have the same mass and therefore $n_\psi = n_{\bar{\psi}}$, which implies $B \propto n_\psi - n_{\bar{\psi}} = 0$.
   If $\psi$ has chemical potential $\mu_\psi$ and is in chemical and thermal equilibrium and takes part in the $B$-violating process $\psi\psi \to \bar{\psi}\bar{\psi}$ (first Sakharov condition), then the relation $\mu_\psi = \mu_{\bar{\psi}}$ must hold. But on the other hand it must be that $\mu_{\bar{\psi}} = -\mu_\psi$, implying that $\mu_\psi$ must vanish and the previous argument applies.

Are these conditions met in the Standard Model (SM)? No.

1. In the SM the baryon number symmetry is anomalous so $B$-violation is present at a quantum level. The baryon ($B$) and lepton ($L$) numbers are exactly conserved at a classical level. But at a quantum level, these symmetries fail to be exact, they are anomalous.

2. The only source of $CP$-violation within the SM is provided by the complex phase of the CKM matrix. But it is too small to explain the observed baryon asymmetry because it is suppressed by small quark masses.

3. The departure from thermal equilibrium could be attained during the electroweak phase transition. Unfortunately, for the phase transition being strong enough to assure departure from equilibrium the Higgs mass should be $m_h \lesssim 60$ GeV, excluded by experimental data.

Therefore the baryon asymmetry is somehow linked to new physics beyond the SM, which is why it is so interesting. Any successful model of baryogenesis needs some new ingredient to be added to the SM. Because of our ignorance about what there is at energy scales well above TeV, one has to postulate some physics at those scales, check that the three Sakharov conditions are fulfilled and compute the generated baryon asymmetry.

Many models of baryogenesis have been proposed so far. Some of the most interesting and most popular ones are

– **Out-of-equilibrium decay**.

The out-of-equilibrium decay of a heavy boson provides a viable mechanism for successful baryo-gensis. Such a heavy boson may be embedded in a Grand Unified Theory (GUT). Let us suppose that a heavy scalar particle $X$ of mass $M_X$ couples to the SM fermions $f$ and has $B$-violating decay modes.

At high temperatures $T \gg M_X$, all particles are in thermal equilibrium and follow their equilibrium number densities. The reactions $X(\bar{X}) \leftrightarrow f\bar{f}$ are in equilibrium. The number density of $X, \bar{X}$ track the equilibrium number density, $n_{X,\bar{X}} = n_X^{\text{eq}}$, so $B = 0$.

When the temperature lowers to $T \lesssim M_X$, the lifetime of $X$ is of the order of the age of the universe $\Gamma_X^{-1} \sim H^{-1}$, and the interactions mantaining the number densities of $X, \bar{X}$ at their equilibrium value are not so effective anymore, provided that the $X$ is sufficiently heavy. So the decays and inverse decays of $X(\bar{X}) \leftrightarrow f\bar{f}$ slow down and the $X$ particles become overabundant with respect to their equilibrium distribution; this is the departure from thermal equlibrium needed for baryogenesis.

If the $X$ decay violates baryon number $B$, a net baryon number is produced for each decay, which would be erased by the opposite baryon asymmetry generated by the decay of $\bar{X}$. So we need the condition that C, CP are violated in the decays, i.e. $\text{BR}(X \to f\bar{f}) \neq \text{BR}(\bar{X} \to f\bar{f})$. These conditions ensure that a net baryon asymmetry is produced for each decay of $X, \bar{X}$.

– **Baryogensis via Leptogenesis**.

A lepton asymmetry is produced in the early universe by out-of-equilibrium decay of heavy right-handed neutrinos. Such asymmetry is then reprocessed at the electroweak scale into a baryon asymmetry (by sphalerons). The mechanism of producing a net lepton asymmetry is similar to the one described for the out-of-equilibrium decay scenario. The appeal of leptogenesis is that it is built in see-saw models motivated by explaining the light neutrino masses.

– **Electroweak Baryogenesis**.

It is a rather complex mechanism aiming at realizing baryogenesis at the electroweak phase transition, by adding new physics at the electroweak scauple that would allow the phase transition to be "strong" enough to provide enough baryon asymmetry. The SM needs to be extended by new extra bosonic degrees of freedom. The generic prediction is the presence of new CP-violating phases in the theory, which may be probed by experiments looking for electron and neutron electric dipole moments.

In conclusion, the cosmology/particle physics interplay has been and currently is a very successful and fascinating *liaison*, which may hold for us even more exciting surprises in the near future.

## Acknowledgements

## References

[1] N. Aghanim *et al.* [Planck Collaboration], "Planck 2018 results. VI. Cosmological parameters," arXiv:1807.06209 [astro-ph.CO].

[2] R. H. Cyburt, B. D. Fields, K. A. Olive and T. H. Yeh, "Big Bang Nucleosynthesis: 2015," Rev. Mod. Phys. **88**, 015004 (2016) doi:10.1103/RevModPhys.88.015004 [arXiv:1505.01076 [astro-ph.CO]].

[3] E. W. Kolb and M. S. Turner, The Early universe, Addison-Wesley (New York, 1990).

[4] D. S. Gorbunov, V. A. Rubakov, "Introduction to the Theory of the Early universe - Hot Big Bang Theory", World Scientific (2011).

[5] V. Mukhanov, "Physical Foundations of Cosmology", Cambridge University Press (2005).

[6] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones and D. Zaritsky, "A direct empirical proof of the existence of dark matter," Astrophys. J. **648**, L109 (2006) doi:10.1086/508162 [astro-ph/0608407].

[7] K. G. Begeman, A. H. Broeils and R. H. Sanders, "Extended rotation curves of spiral galaxies: Dark haloes and modified dynamics," Mon. Not. Roy. Astron. Soc. **249**, 523 (1991).

[8] E. Aprile *et al.* [XENON Collaboration], "Dark Matter Search Results from a One Ton-Year Exposure of XENON1T," Phys. Rev. Lett. **121**, no. 11, 111302 (2018) doi:10.1103/PhysRevLett.121.111302 [arXiv:1805.12562 [astro-ph.CO]].

[9] M. Ackermann *et al.* [Fermi-LAT Collaboration], "The Fermi Galactic Center GeV Excess and Implications for Dark Matter," Astrophys. J. **840**, no. 1, 43 (2017) doi:10.3847/1538-4357/aa6cab [arXiv:1704.03910 [astro-ph.HE]].

[10] A. De Simone and T. Jacques, "Simplified models vs. effective field theory approaches in dark matter searches," Eur. Phys. J. C **76**, no. 7, 367 (2016) doi:10.1140/epjc/s10052-016-4208-4 [arXiv:1603.08002 [hep-ph]].

[11] A. H. Guth, "The Inflationary universe: A Possible Solution To The Horizon And Flatness Problems," Phys. Rev. D **23** (1981) 347.

[12] A. D. Linde, "A New Inflationary universe Scenario: A Possible Solution Of The Horizon, Flatness, Homogeneity, Isotropy And Primordial Monopole Problems," Phys. Lett. B **108**, 389 (1982).

[13] A. Albrecht and P. J. Steinhardt, "Cosmology For Grand Unified Theories With Radiatively Induced Symmetry Breaking," Phys. Rev. Lett. **48**, 1220 (1982).

# Introduction to flavour physics

*J. Zupan*
Department of Physics, University of Cincinnati, Cincinnati, Ohio, USA

### Abstract

We give a brief introduction to flavour physics. The first part covers the flavour structure of the Standard Model, how the Kobayashi-Maskawa mechanisem is tested and provides examples of searches for new physics using flavour observables, such as meson mixing and rare decays. In the second part we give a brief overview of the recent flavour anomalies and how the Higgs can act as a new flavour probe.

### Keywords

Flavour physics; heavy quarks; $B$ physics; meson mixing; new physics; Higgs; lectures.

## 1   Introduction

The term "flavour" was coined in 1971 by Murray Gell-Mann and his student at the time, Harald Fritzsch, while sitting at a Baskin-Robbins ice-cream store in Pasadena, CA [1]. Just as ice-cream has both colour and flavour so do quarks. "Flavour" is now used slightly more generally to denote the species of any Standard Model (SM) fermion, both quarks and leptons. "Flavour physics" thus has little to do with one's adventures in kitchen, but rather is a research area that deals with properties of quarks and leptons.

Grouped according to their QCD and QED quantum numbers, $SU(3) \times U(1)_{\mathrm{em}}$, the SM fermions are,

$$
\begin{array}{lll}
3_{2/3} : & \text{up type quarks;} & u, c, t, \\
3_{-1/3} : & \text{down type quarks;} & d, s, b, \\
1_{-1} : & \text{charged leptons;} & e, \mu, \tau, \\
1_0 : & \text{neutrinos;} & \nu_e, \nu_\mu, \nu_\tau.
\end{array}
\tag{1}
$$

Each fermion type comes in three copies, i.e., the SM fermions group into three generations.

In this brief introduction to flavour physics we will cover some of the classic topics on the subject: the flavour structure of the Standard Model (SM), how the Kobayashi-Maskawa mechanism is tested, as well as the constraints on the New Physics (NP) due to flavour observables such as the meson mixing and decays. We will also touch on the more recent developments: the $B$ physics anomalies and the Higgs as a new probe of flavour. Along the way we will address two major questions currently facing particle physics. The first question is why do the SM fermions exhibit such a hierarchical structure, shown in Fig. 1? This is commonly referred to as the SM flavour puzzle. The other question is what lies above the electroweak scale? Here flavour physics offers a way to probe well above the electroweak scale.

Other excellent introductions to flavour physics the reader may want to consult include Refs. [2–7]. Ref. [2] in particular is chock full of physics insights without too much burdensome formalism. Section 2 borrows liberally from [5], which, while slightly outdated, is still a masterful introduction to the basic topics in flavour physics. For a reader that is seeking much more depth a good starting point can be Refs. [8–11].

## 2   The flavour of the Standard Model

### 2.1   The SM symmetry structure

A renormalizable particle physics model is defined by specifying (i) the gauge group and (ii) the particle field content. The next step is to write down the most general renormalizable Lagrangian. The SM gauge
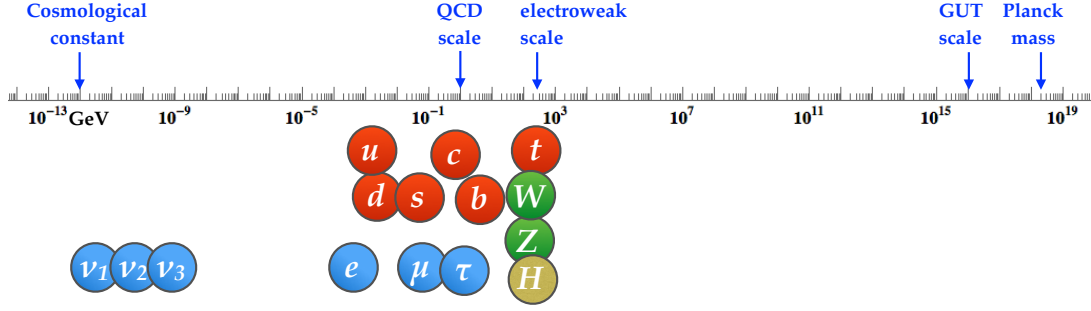
**Fig. 1:** The distribution of masses of the elementary particles, along with some of the relevant energy scales. The absolute values of neutrino masses are not known - their placement on the graph is indicative of the upper bound.

group is

$$\mathcal{G}_{\rm SM} = SU(3)_c \times SU(2)_L \times U(1)_Y. \tag{2}$$

Here $SU(3)_c$ is the gauge group of strong interactions, Quantum Chromodynamics (QCD), the $SU(2)_L$ is the gauge group of weak isospin, and $U(1)_Y$ the gauge group of hypercharge. The field content of the SM consists of a single scalar, EW doublet

$$H \sim (1,2)_{1/2}, \tag{3}$$

and a set of fermion fields,

$$\begin{aligned} Q_{Li} &\sim (3,2)_{+1/6}, & u_{Ri} &\sim (3,1)_{+2/3}, & d_{Ri} &\sim (3,1)_{-1/3}, \\ L_{Li} &\sim (1,2)_{-1/2}, & \ell_{Ri} &\sim (1,1)_{-1}. \end{aligned} \tag{4}$$

Each of the fields comes in three copies (three generations), $i = 1, 2, 3$. To simplify the discussion we will set neutrino masses to zero. The modifications due to nonzero neutrino masses are given in appendix A. The $\mathcal{G}_{\rm SM}$ is spontaneously broken by the Higgs vacuum expectation value, $\langle H \rangle = (0, v/\sqrt{2}), v = 246$ GeV, down to

$$\mathcal{G}_{\rm SM} \to SU(3) \times U(1)_{em}. \tag{5}$$

After the electroweak symmetry breaking the field content in (4) splits into up and down quarks, charged leptons and neutrinos as listed in Eq. (1).

## 2.2 The SM Lagrangian

The SM Lagrangian is the most general renormalizable Lagrangian that is consistent with the gauge group $\mathcal{G}_{\rm SM}$ and the field content (3), (4)

$$\mathcal{L}_{\rm SM} = \mathcal{L}_{\rm kin} + \mathcal{L}_{\rm Yukawa} + \mathcal{L}_{\rm Higgs}. \tag{6}$$

The kinetic terms in the Lagrangian are determined by the gauge structure through the covariant derivative

$$D_\mu \psi = (\partial_\mu + i g_s G_\mu^a t^a + i g W_\mu^i \tau^i + i g' B_\mu Y) \psi. \tag{7}$$

The strong interaction term is a product of the strong coupling, $g_s$, the eight gluon fields, $G_\mu^a$, and the generators $t^a$ of $SU(3)_c$. For color triplet $\psi$ these are $t^a = \lambda^a/2$, with $\lambda^a$ the eight $3 \times 3$ Gell-Mann matrices, while for color singlet $\psi$, $t^a = 0$. The $SU(2)_L$ term is a product of the weak coupling, $g$, the three weak gauge bosons, $W_\mu^i$, and the generators of $SU(2)_L$, $\tau^i$ (equal to $\tau^i = \sigma^i/2$ for $\psi$ that is a doublet, with $\sigma^i$ the Pauli matrices, while for singlets $\tau^i = 0$). The last term is due to the hypercharge $U(1)_Y$.

The covariant derivatives are flavour blind, i.e., generation independent. For instance, for $Q_L^i$ the kinetic term is

$$\mathcal{L}_{\text{kin}}\big|_{Q_L} = i\bar{Q}_L^i\big(\partial_\mu + ig_s G_\mu^a \tfrac{1}{2}\lambda^a + igW_\mu^i \tfrac{1}{2}\sigma^i + i\tfrac{1}{6}g'B_\mu\big)\delta^{ij}Q_L^j, \tag{8}$$

for up quarks it is

$$\mathcal{L}_{\text{kin}}\big|_{u_R} = i\bar{u}_R^i\big(\partial_\mu + ig_s G_\mu^a \tfrac{1}{2}\lambda^a + i\tfrac{2}{3}g'B_\mu\big)\delta^{ij}u_R^j, \tag{9}$$

and similarly for the other fields. Each of the kinetic terms is invariant under the global $U(3) = SU(3) \times U(1)$ transformations. Thus $\mathcal{L}_{\text{kin}}$ has a global flavour symmetry

$$\mathcal{G}_{\text{flavour}} = U(3)_{\text{q}}^3 \times U(3)_{\text{lep}}^2, \tag{10}$$

where

$$U(3)_{\text{q}}^3 = U(3)_Q \times U(3)_u \times U(3)_d, \tag{11}$$

$$U(3)_{\text{lep}}^2 = U(3)_L \times U(3)_\ell. \tag{12}$$

That is, each of the five different types of fermions in Eq. (4) can be separately rotated in flavour space, $\psi^i \to U_j^i \psi^j$, where $U_j^i$ is a unitary $3 \times 3$ matrix, without changing $\mathcal{L}_{\text{kin}}$.

However, $\mathcal{G}_{\text{flavour}}$ cannot be an exact symmetry of the whole Lagrangian. We know from observations that, e.g., the top quark differs from the up quark due to their differing masses. The part of the Lagrangian that breaks $\mathcal{G}_{\text{flavour}}$ is

$$\mathcal{L}_{\text{Yukawa}} = -Y_d^{ij}\bar{Q}_L^i H d_R^j - Y_u^{ij}\bar{Q}_L^i H^c u_R^j - Y_\ell^{ij}\bar{L}_L^i H \ell_R^j + \text{h.c.}. \tag{13}$$

The above Yukawa interactions break

$$\mathcal{G}_{\text{flavour}} \to U(1)_B \times U(1)_e \times U(1)_\mu \times U(1)_\tau \times U(1)_Y, \tag{14}$$

where $U(1)_B$ is the baryon number, and $U(1)_\ell$ are the separate lepton numbers. That $\mathcal{L}_{\text{Yukawa}}$ breaks the flavour symmetry is not surprising, since it is the origin of fermion masses, once the Higgs obtains the vacuum expectation value (vev), $\langle H \rangle = (0, v/\sqrt{2})$, with $v = 246$ GeV.

## 2.3 A Standard Model vs. the Standard Model

Before we proceed further in understanding the breaking pattern in Eq. (14), let us make a small detour and elaborate on the difference between *a Standard Model* and *the Standard Model*. *A Standard Model* denotes any model with the SM gauge group (2) and the SM field content (3), (4), but with some arbitrary values for the coupling constants in the most general renormalizable Lagrangian. *The Standard Model* is a Standard Model with exactly the values of coupling constants observed in nature. *A Standard Model* has the exact accidental symmetry $U(1)_B \times U(1)_e \times U(1)_\mu \times U(1)_\tau$. This accidental symmetry is present for any values of the parameters in the renormalizable SM Lagrangian (but can be broken by non-renormalizable terms). It is accidental, since we did not explicitly ask for it – it is simply present because we cannot write down renormalizable terms that break it, given the field and gauge content in Eqs. (2)-(4). For *the Standard Model*, because of the actual values of the parameters, there can be additional approximate symmetries.

Isospin is an example of such an approximate symmetry. In QCD interactions one can replace $u$ and $d$ quarks without affecting appreciably the results. For instance, the neutron and proton masses are very close to each other even though, $p \sim uud$, while $n \sim udd$. The reason is not that up and down quark masses would be equal to each other but rather that they are both small, cf. Fig. 1,

$$\frac{|m_u - m_d|}{\Lambda_{\text{strong}}} \ll 1. \tag{15}$$

Here $\Lambda_{\text{strong}} \sim \mathcal{O}(1\text{GeV})$ is the typical scale at which QCD becomes nonperturbative and generates the bulk of the mass for proton and neutron.

## 2.4 Counting physical parameters

The next question we need to address is how one counts the physical parameters. The SM has 19 physical parameters: 3 gauge couplings, 3 lepton masses, 6 quark masses, 4 parameters in the Cabibbo-Kobayashi-Maskawa (CKM) matrix, 2 parameters in the Higgs sector (the Higgs mass and the strength of the self interaction), and the QCD $\theta$ parameter. Physical parameters are parameters that cannot be rotated away by performing phase transformations or flavour rotations.

Let us understand this on the case of charged lepton masses. The charged lepton Yukawa

$$\mathcal{L}_{\text{Yukawa}} \supset -Y_\ell^{ij} \bar{L}_L^i H \ell_R^j + \text{h.c.}, \tag{16}$$

can always be made diagonal and real positive through a bi-unitary transformation, $L_L \to V_L L_L$, $\ell_R \to V_\ell \ell_R$, which gives

$$Y_\ell \to V_L^\dagger Y_\ell V_\ell = \text{diag}(y_e, y_\mu, y_\tau). \tag{17}$$

How many physical parameters are there? The starting point, $Y_\ell$, is described by 9 real and 9 imaginary numbers. The unitary matrices $V_L, V_\ell$ have in total $2 \times (3 \text{ real} + 6 \text{ im.})$ numbers. When we rotate $L_L^i$ and $\ell_R^i$ by the same phase there is no change in $y_{\ell_i}$. That means that 3 phases (im. numbers) have no effect. Thus we have $9 - 2 \times 3 = 3$ real, and $9 - (2 \times 6 - 3) = 0$ imaginary physical parameters. The three real physical parameters are the charged lepton masses, while there are no physical phases.

Extrapolating from this exercise we can postulate the general rule on how to count the physical parameters [2]

$$\# \text{ physical parameters} = \# \text{ parameters} - \# \text{ broken symmetry generators}. \tag{18}$$

Let us check this with a simple example: the spin $1/2$ in a magnetic field. If there is no magnetic field the system has an $SO(3)$ symmetry (3 generators), since the spin can be oriented in an arbitrary direction without changing the energy. The system also has two degenerate eigenstates corresponding to spin up and spin down. In the magnetic field the Zeeman effect splits the two states. The splitting depends on the strength of the magnetic field, $B$. There is thus one physical parameter that controls the splitting. However, the magnetic field in general has three components, and is thus described by 3 parameters, $\vec{B} = B_x \hat{x} + B_y \hat{y} + B_z \hat{z}$. One can use the rotation around $x$ and $y$ axes to align $\vec{B}$ along the $z$ axis, i.e., set $B_x = B_y = 0$. After this is done, making any further rotations around $x$ and $y$ axes would change the $\vec{B}$ component: there are 2 broken symmetry generators. Using the general rule (18) gives that there is $3 - 2 = 1$ physical parameter, as expected.

We can now apply (18) to count the physical parameters in the quark sector of the SM. Using the unitary transformations

$$Q_L \to V_Q Q_L, \qquad u_R \to V_u u_R, \qquad d_R \to V_d d_R, \tag{19}$$

one can bring the Yukawa couplings to the form

$$Y_d = \text{diag}(y_d, y_s, y_b), \qquad Y_u = V_{\text{CKM}}^\dagger \text{diag}(y_u, y_c, y_t), \tag{20}$$

with $V_{\text{CKM}}$ a unitary $3 \times 3$ CKM matrix [12,13]. How many entries in $V_{\text{CKM}}$ are physical? The starting point, the $Y_u, Y_d$ matrices, have $2 \times (9 \text{ real} + 9 \text{ im.})$ parameters. The three unitary matrices, $V_Q, V_u, V_d$ have in total $3 \times (3 \text{ real} + 6 \text{ im.})$ parameters. Finally, there is one global phase corresponding to common phase change $Q_L \to \exp(i\phi)Q_L, u_R \to \exp(i\phi)u_R, d_R \to \exp(i\phi)d_R$, which has no effect. That is, there is one unbroken symmetry generator – the baryon number, while all the other symmetry generators are broken. Using (18) we see that there are $2 \times 9 - 3 \times 3 = 9$ real parameters and $2 \times 9 - (3 \times 6 - 1) = 1$ imaginary physical parameter. These are the 6 quark masses, as well as the 3 mixing angles and one phase describing the CKM matrix.
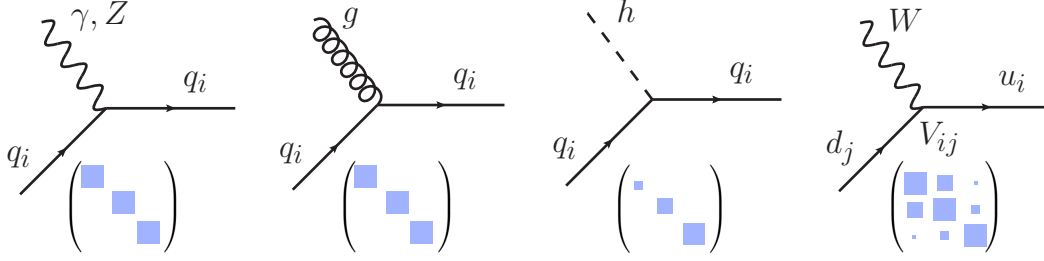
**Fig. 2:** The Feynman diagrams for flavour conserving couplings of quarks to photon, $Z$ boson, gluon and the Higgs (the first three diagrams), and the flavour changing coupling to the $W$ (the last diagram). The $3 \times 3$ matrices are visual representations of couplings in the generation space, with couplings to $\gamma, Z, g$ flavour universal, the couplings to the Higgs flavour diagonal but not universal, and the couplings to $W$ flavour changing and hierarhical.

A conventional parametrization of the CKM matrix is [14]

$$
\begin{aligned}
V_{\text{CKM}} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix},
\end{aligned}
\tag{21}
$$

where $c_{ij} \equiv \cos\theta_{ij}$, $s_{ij} \equiv \sin\theta_{ij}$, so that the CKM matrix is a product of three rotations with one phase inserted in the matrix describing the $\theta_{13}$ rotation. Experimentally, we observe that $\theta_{12} \gg \theta_{23} \gg \theta_{13}$, while $\delta \sim \mathcal{O}(1)$.

As the side benefit of the counting of physical parameters we just performed, we also understand that the flavour breaking due to the Yukawa matrices is as given in Eq. (14). In more detail, if we were to take nonzero just a single Yukawa coupling matrix at the time, the breaking pattern is

- since $Y_\ell \not\propto 1$: $U(3)_L \times U(3)_\ell \to U(1)_e \times U(1)_\mu \times U(1)_\tau$, i.e., the charged lepton family numbers,
- since $Y_u \not\propto 1$: $U(3)_Q \times U(3)_u \to U(1)_u \times U(1)_c \times U(1)_t$, i.e., the up-quark family numbers,
- since $Y_d \not\propto 1$: $U(3)_Q \times U(3)_d \to U(1)_d \times U(1)_s \times U(1)_b$, i.e., the down-quark family number,
- since $[Y_d, Y_u] \neq 0$: $U(1)_q^6 \to U(1)_B$, i.e., the above quark $U(1)$'s further break to a global baryon number.

Note that the final $U(1)$'s are composed both from the $U(1)$ factors in the original $[U(3) = SU(3) \times U(1)]$'s, as well as from the $t^3$ and $t^8$ generators of the $SU(3)$'s. In particular, not all of the $U(1)$ factors in $\mathcal{G}_{\text{flavour}}$ get broken by the Yukawas. The $\mathcal{G}_{\text{flavour}}$ contains five $U(1)$ factors, which can be chosen to be $U(1)^5 = U(1)_Y \times U(1)_B \times U(1)_L \times U(1)_{\text{PQ}} \times U(1)_{\ell_R}$. The $U(1)_Y$ is the hypercharge group, which is gauged, while $B$ and $L$ are the global baryon and lepton numbers. These are not broken by $\mathcal{L}_{\text{Yukawa}}$. The remaining two global $U(1)$'s can be taken to be the Peccei-Quinn symmetry $U(1)_{\text{PQ}}$ ($H$ and $d_R^i, \ell_R^i$ have opposite charges, all others zero), while under $U(1)_{\ell_R}$ only $\ell_R^i$ is charged. The $U(1)_{\text{PQ}}$ is broken by $Y_u \neq 0$, and $U(1)_{\ell_R}$ by $Y_\ell \neq 0$. Had we included neutrino masses in the discussion, these would furthermore break the separate lepton numbers to a common lepton number, $U(1)_L$, if the neutrino masses are Dirac, while Majorana masses also break $U(1)_L$, see appendix A.

## 2.5  The flavour violation as seen in the mass basis

The main message of the discussion so far is: in the SM the flavour structure (flavour breaking) resides in the Yukawa sector of the SM Lagrangian, Eq. (13). If the Yukawa couplings were vanishingly small, the

SM would have had a very large flavour symmetry $\mathcal{G}_{\text{flavour}}$, Eq. (10). In general, the flavor breaking can be parametrized as in Eq. (20), by 6 diagonal Yukawa couplings, and the elements of the CKM matrix, $V_{\text{CKM}}$.

After Higgs obtains the vev, the Yukawa terms give the quark and charge lepton masses,

$$\mathcal{M}_f = Y_f \frac{(v+h)}{\sqrt{2}}. \tag{22}$$

With a field redefinition for the left-handed up quark fields

$$Q_L \to \begin{pmatrix} V^\dagger u_L \\ d_L \end{pmatrix}, \tag{23}$$

we can move the flavour changing interactions to the kinetic term. This gives the SM Lagrangian for the quarks in the mass basis

$$\mathcal{L}_{\text{SM}} \supset (\bar{q}_i \slashed{D}_{\text{NC}} q_i) + \frac{g}{\sqrt{2}} \bar{u}_L^i W^+ V_{\text{CKM}}^{ij} d_L^j + m_{u_i} \bar{u}_L^i u_R^i \left(1 + \frac{h}{v}\right) + m_{d_i} \bar{d}_L^i d_R^i \left(1 + \frac{h}{v}\right) + \text{h.c.} \tag{24}$$

The covariant derivative $D_{\text{NC}}$ contains flavour (generation) universal couplings of photon, gluon and the $Z$. The Higgs has flavour diagonal, yet non-universal, couplings that are proportional to quark masses, while the flavour changing transitions reside in charged currents, with the strength encoded in the CKM matrix, see Fig. 2.

## 2.6  Charged currents vs. neutral currents

In the SM there is a very important distinction between flavour changing neutral and charged currents. Flavour Changing Neutral Currents (FCNCs) are processes in which the quark flavour changes, while the quark charge stays the same. The charged currents change both the flavour and the charge of the quark. A glimpse at the PDG booklet [15] reveals that the probabilities for the two types of processes are strikingly different. The charged currents lead to the dominant weak decays, while the FCNC induced decays are extremely suppressed. Rounding the experimental results, and not showing the errors, a few representative decays are

charged currents:

neutral currents:

$$s \to u\mu^-\bar{\nu}_\mu \;\; : \;\; Br(K^+ \to \mu^+\nu) = 64\%, \qquad s \to d\mu^+\mu^- \;\; : \;\; Br(K_L \to \mu^+\mu^-) = 7 \times 10^{-9},$$
$$b \to c\ell^-\bar{\nu}_\ell \;\; : \;\; Br(B^- \to D^0\ell\bar{\nu}) = 2.3\%, \qquad b \to d\mu^+\mu^- \;\; : \;\; Br(B^- \to K^{*-}\ell^+\ell^-) = 5 \times 10^{-7},$$
$$c \to s\mu^+\nu_\mu \;\; : \;\; Br(D^\pm \to K^0\mu^\pm\nu) = 9\%, \qquad c \to u\ell^+\ell^- \;\; : \;\; Br(D^0 \to \pi^0\ell^+\ell^-) < 1.8 \times 10^{-4},$$

The reason for such a striking difference is that in the SM the charged currents occur at tree level, while FCNCs are forbidden at tree level and only arise at one loop, see Fig. 3. Furthermore, the FCNCs come suppressed by the difference of the masses of the quarks running in the loop, $m_j^2 - m_i^2$. This so called Glashow-Iliopoulos-Maiani (GIM) mechanism [16] is a result of the fact that there is no flavour violation, if all the quark masses are the same.

## 2.7  The CKM matrix

The Cabibbo-Kobayashi-Maskawa (CKM) matrix is very hierarchical in the SM,

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \sim \begin{pmatrix} 1 & 0.2 & 0.004 \\ 0.2 & 1 & 0.04 \\ 0.008 & 0.04 & 1 \end{pmatrix}. \tag{25}$$
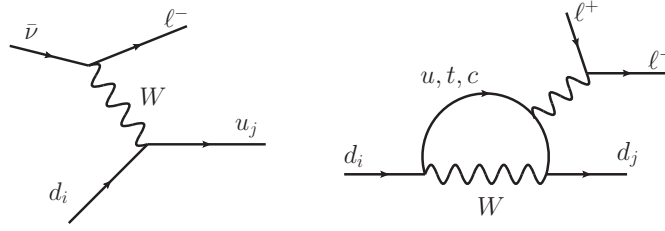
**Fig. 3:** Representative tree level charged current diagram (left) and a loop induced FCNC diagram (right).

In fact, for processes at colliders in many cases the CKM matrix can even be approximated as

$$V_{\text{CKM}} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \text{[collider physicist]} \tag{26}$$

i.e., for many processes at high $p_T$ to a good enough precision the generation number is conserved.

We, on the other hand, are interested precisely in the off-diagonal entries in $V_{\text{CKM}}$. These entries roughly obey a power scaling in $\lambda \equiv |V_{us}| \simeq 0.22$, giving the Wolfenstein parametrization of the CKM matrix [17],

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \tag{27}$$

This parametrization also encodes that the CKM matrix is unitary, $V_{\text{CKM}}^\dagger V_{\text{CKM}} = V_{\text{CKM}} V_{\text{CKM}}^\dagger = 1$. The CKM matrix depends on 3 real parameters and 1 phase. In parametrization of Eq. (21) these were the three mixing angles and the phase $\delta$. In the Wolfenstein parametrization, Eq. (27), these are the three real parameters $\lambda$, $A$, $\rho$, and one imaginary parameter, $\eta$, all counted as being $\mathcal{O}(1)$. A global fit to the flavour observables gives [18]

$$A = 0.825(9), \qquad \lambda = 0.2251(3), \qquad \bar{\rho} = 0.160(7), \qquad \bar{\eta} = 0.350(6), \tag{28}$$

where the modified $\rho, \eta$ parameters were introduced as $\bar{\rho} + i\bar{\eta} = -V_{ud}V_{ub}^*/(V_{cd}V_{cb}^*)$, valid to all orders in $\lambda$. To $\mathcal{O}(\lambda^4)$ we have $\bar{\rho} = \rho(1 - \lambda^2/2)$ and $\bar{\eta} = \eta(1 - \lambda^2/2)$. Note that numerically $\bar{\rho}, \bar{\eta}$ are maybe closer to $\bar{\rho}, \bar{\eta} \sim \mathcal{O}(\lambda)$ than $\bar{\rho}, \bar{\eta} \sim \mathcal{O}(1)$, while at the time when Wolfenstein parametrization was written down this was not known. This can be incorporated in modified expansions [19], though the change in counting only matters at higher orders, not for the leading order expressions in Eq. (27).

## 2.8 Origin of CP violation in the SM

The SM Lagrangian is invariant under the discrete CP symmetry, apart from the Yukawa terms.[1] These transform as (writing explicitly also the hermitian conjugate terms)

$$Y_{ij}\bar{\psi}_L^i H \psi_R^j + Y_{ij}^*\bar{\psi}_R^j H^\dagger \psi_L^i \xrightarrow{\text{CP}} Y_{ij}\bar{\psi}_R^j H^\dagger \psi_L^i + Y_{ij}^*\bar{\psi}_L^i H \psi_R^j. \tag{29}$$

The CP is conserved, if Yukawa couplings are real,

$$Y_{ij}^* = Y_{ij}. \tag{30}$$

Since there is only one physical phase in the CKM, in the SM the CP violation (CPV) is controlled by one parameter, the "CKM phase", which in the Wolfenstein parametrization is the parameter $\eta$. CP is

---

[1]There is another CP violating parameter, the strong CP phase multiplying the QCD anomaly term, $g^2/(32\pi^2)\theta G^{a\mu\nu}\tilde{G}_{\mu\nu}^a$. It is bounded experimentally to be small, $\theta \lesssim 10^{-10}$ and, even if eventually found to be nonzero, is negligible for all the processes discussed in these lectures.
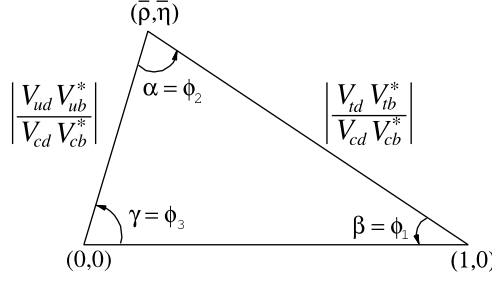
**Fig. 4:** The standard CKM unitarity triangle (from [15]).

thus violated only, if $\eta \neq 0$. This origin of the observed CPV is called the Kobayashi-Maskawa (KM) mechanism [13]. Furthermore, CPT is conserved in any Lorentz invariant Quantum Field Theory, and therefore also in the SM. This means that CPV is equivalent to having T violation – the time reversal is also violated in the SM.

For the existence of CPV in the SM it is crucial that there are at least 3 generations of quarks. Repeating the counting of physical parameters from Sec. 2.4 we can easily convince ourselves that it is possible in the case of 2 generations to make CKM real through field redefinitions. Furthermore, if $Y_u$ and $Y_d$ are "aligned", meaning that they are diagonalized with the same left-handed rotation, then $V_{\text{CKM}} = 1$. This means that in the SM, if there is no flavour violation, there is also no CP violation (ignoring the flavour universal, but numerically negligible $\theta$ term).

The above insights can be encoded in a measure of CP violation, the Jarlskog invariant [20]

$$J_Y \equiv \text{Im} \left( \det \left[ Y_d Y_d^\dagger, Y_u Y_u^\dagger \right] \right). \tag{31}$$

The $J_Y$ is invariant under flavour transformations, $\mathcal{G}_F$, Eq. (10), and is thus basis independent. The CP is conserved, if $J_Y = 0$. We can also write $J_Y$ as

$$J_Y = J_{\text{CP}} \prod_{i>j} \frac{m_i^2 - m_j^2}{v^2/2} \simeq \mathcal{O}(10^{-22}), \tag{32}$$

where the invariant measure of CP violation is

$$J_{\text{CP}} = \text{Im} \left[ V_{us} V_{cb} V_{ub}^* V_{cs}^* \right] = c_{12} c_{23} c_{13}^2 s_{12} s_{23} s_{13} \sin \delta_{\text{KM}} \simeq \lambda^6 A^2 \eta \simeq \mathcal{O}(10^{-5}). \tag{33}$$

The product of masses is

$$\prod_{i>j} \frac{m_i^2 - m_j^2}{v^2/2} = \frac{(m_t^2 - m_c^2)}{v^2/2} \frac{(m_t^2 - m_u^2)}{v^2/2} \frac{(m_c^2 - m_u^2)}{v^2/2} \frac{(m_b^2 - m_s^2)}{v^2/2} \frac{(m_b^2 - m_d^2)}{v^2/2} \frac{(m_s^2 - m_d^2)}{v^2/2}. \tag{34}$$

It would vanish, if any of the two pairs of masses were equal, in which case CP would have been conserved.

## 3 Tests of the CKM structure

### 3.1 The standard CKM unitarity triangle

All flavour transitions in the SM depend on only 4 fundamental parameters, $\lambda$, $A$, $\rho$, and $\eta$. We can test the Kobayashi-Maskawa mechanism by making many measurements, over-constraining the system. One way to visualize a subset of experimental constraints is through the standard CKM unitarity triangle, which tests one out of nine unitarity equations, $V_{\text{CKM}} V_{\text{CKM}}^\dagger = 1$. The standard CKM unitarity triangle is obtained from a product of the first and the third column of the CKM matrix

$$V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* = 0, \tag{35}$$

**Fig. 5:** The evolution of the constraints in the standard CKM unitarity triangle plane from 1995 (left), to just after the start of $B$ factories (middle), to the present (right). Taken from the ckmfitter website [18].



**Fig. 6:** Some of the main CKM constrains and the respective SM diagrams.

which we can rewrite as

$$\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} + 1 + \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} = 0. \tag{36}$$

In terms of the Wolfenstein parameters this sum rule is

$$-\left(\bar{\rho} + i\bar{\eta}\right) + 1 + \left(-1 + \bar{\rho} + i\bar{\eta}\right) = 0. \tag{37}$$

The relation (36) can be interpreted as a sum of three complex numbers that are the sides of a triangle, shown in Fig. 4. There are two common notations for the angles of the standard CKM unitarity triangle: either $\alpha$, $\beta$, $\gamma$ or $\phi_1, \phi_2, \phi_3$, used by the two $B$-factories, BaBar and Belle, respectively. The Belle experiment (1999-2010) at KEK, Japan produced about $\sim 1.5 \times 10^9$ $B$ mesons, while BaBar experiment 1999-2008) at SLAC, USA collected about $\sim 0.9 \times 10^9$ $B$ mesons. The two experiments established that the KM mechanism is the main source of CP violation in the SM. The progression of constraints in the CKM unitarity triangle plane is shown in Fig. 5. We see that there was a big qualitative jump after the start of the $B$ factories, and a very impressive set of improvements in the constraints since then.

The constraints on the standard CKM unitarity triangle are coming from several different meson systems, the $B_d^0, B^+$ mesons from measurements at Belle, BaBar and LHCb, the $B_s$ meson and $\Lambda_b$

**Fig. 7:** Evolving constraints in the $\bar{\rho} - \bar{\eta}$ plane from LHCb measurements and improvements in lattice QCD calculations, alone, with current inputs (2018), and the anticipated improvements from the data accumulated by 2035 (300 fb$^{-1}$ of integrated luminosity). More information on the fits may be found in [11, 21].

baryon from measurements at LHCb, and the kaon physics experiments. Different constraints in the standard CKM unitarity triangle plane are shown in Fig. 6, together with the relevant SM diagrams. The upshot of these results is that the KM mechanism is the dominant origin of CPV. The measurements point to a consistent picture of flavour violation, described by four parameters, $A$, $\lambda$, $\bar{\rho}$, $\bar{\eta}$, with the values given in Eq. (28). Since $\bar{\rho} \lesssim \bar{\eta}$ the CKM phase is large, $\mathcal{O}(1)$. It is given by

$$\gamma = \arctan\left(\bar{\eta}/\bar{\rho}\right) = \arg(V_{ub}^*),\tag{38}$$

where in the last equality we used the common parametrization of the CKM matrix, where the weak phase is moved to the $V_{ub}$ and $V_{ts}$ CKM elements, (27). Experimentally [18],

$$\gamma = (65.4 \pm 1.1)^\circ,\tag{39}$$

so that the weak phase is indeed $\mathcal{O}(1)$ when measured in radians.

The field is undergoing a big upgrade in available statistics. The successor to the Belle experiment, called Belle II, is ramping up right now, with the first physics run expected in early 2019 [10]. Belle II aims to collect about $\sim 8 \times 10^{10}$ $B$ mesons by about 2025, roughly $50\times$ more than Belle did. The LHCb experiment also has ambitious upgrade plans [21]. After the end of Upgrade II in 2035 it may have the statistics that corresponds to roughly $\sim 10^{11}$ or more useful $B$'s (because of hadronic environment this number fluctuates from channel to channel), as well as $B_s$ mesons and heavy baryons, which are also produced in the $pp$ collisions. The constraints on the elements of the CKM matrix are thus set to become much more precise in the future. Fig. 7 (right) shows the improvements that can be achieved by using just the LHCb measurements alone at the end of the high luminosity LHC programme. A similar projection for the improvements using Belle II measurements can be found in Ref. [22].

The constraints in the standard CKM unitarity triangle plot are of two types: the tree level transitions, which are less likely to be affected by new physics, and the loop level transitions, which are more likely to be affected by new physics. In the rest of this section we will choose an example transition for each of the two types of transitions and look at it in detail. This will then lead us to the discussion of new physics searches in Section 4. However, before we do that, we need to introduce several new concepts.

## 3.2 The meson mixing

The term *mixing* denotes that the flavour eigenstates do not equal mass eigenstates, i.e., that the eigenstates of the SM Hamiltonian are composed of states with different flavour compositions. For instance, $B^0 \sim \bar{b}d$ and $\bar{B}^0 \sim b\bar{d}$ are flavour eigenstates but are not mass eigenstates. The mass eigenstates are admixtures of $B^0$ and $\bar{B}^0$.
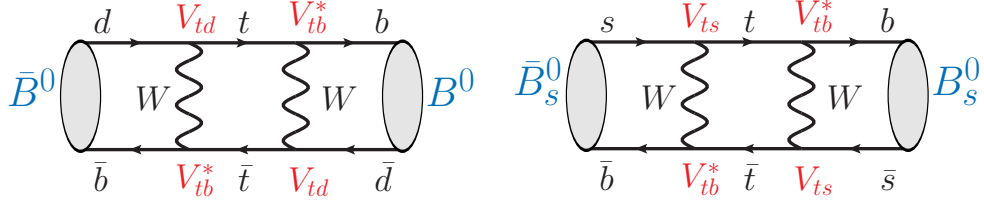
**Fig. 8:** The SM diagrams leading to $B_d$ (left) and $B_s$ mixing (right).

The term *oscillations* denotes that the initial flavour eigenstate time evolves to a different flavour eigenstate. The reason for this is that the flavour eigenstates are composed from two mass eigenstates, each of which evolves slightly differently. The oscillation frequency is the energy splitting, $\omega = \Delta E$. In the rest frame this equals the mass splitting, $\Delta E = \Delta m$, which means that the oscillations are an excellent way to measure small mass splittings.

What kind of mixings between states are possible? A general rule that applies here is: what is not explicitly forbidden is allowed [2]. Using this important rule let us look at two examples:

- Can $B^+ \sim \bar{b}u$ and $B^- \sim b\bar{u}$ mix? The answer is no, since the electric charge is conserved. That is, the $U(1)_{\text{em}}$ gauge symmetry forbids such mixings to all orders in perturbation theory.
- Can $B^0 \sim \bar{b}d$ and $\bar{B}^0 \sim b\bar{d}$ mix? In this case the answer is yes, since nothing forbids it. That is, there is no exact symmetry that forbids this mixing to all orders, so at some order in perturbation theory the mixing will occur. Such FCNCs are forbidden at three level in the SM, but are allowed at 1 loop.

A representative 1 loop weak interactions diagram that mixes $\bar{B}^0 \sim \bar{b}d$ and $B^0 \sim b\bar{d}$ in the SM , is shown in Fig. 8 (left). These diagrams contribute to the flavour off diagonal elements in the Hermitian $\bar{B}^0$, $B^0$ mass matrix, see e.g., [7],[2]

$$\mathcal{M} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \tag{40}$$

written in the flavour basis

$$\begin{pmatrix} |B^0\rangle \\ |\bar{B}^0\rangle \end{pmatrix}. \tag{41}$$

The off-diagonal elements are much smaller than the diagonal ones, so that the mass matrix has the form,

$$\mathcal{M} \propto \begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix}. \tag{42}$$

CPT guarantees $M_{11} = M_{22}$. If CP is conserved, then also $M_{12} = M_{21}$. Numerically, $M_{11} = M_{22} \simeq m_B$, while $M_{12,21} \ll M_{11}$. If CP is conserved, the mass eigenstates are

$$|B_{L,H}\rangle = \frac{1}{\sqrt{2}}\big(|B^0\rangle \pm |\bar{B}^0\rangle\big), \tag{43}$$

where we used the phase convention

$$CP|B^0\rangle = |\bar{B}^0\rangle, \qquad CP|\bar{B}^0\rangle = |B^0\rangle, \tag{44}$$

such that

$$CP|B_{L,H}\rangle = \pm|B_{L,H}\rangle. \tag{45}$$

---

[2]Note that our phase conventions differ from [7] by a sign. The results in [7] are obtained by replacing $q \to -q$, $|\bar{B}^0\rangle \to -|\bar{B}^0\rangle$.

That is, the mass eigen-states are maximally mixed, exactly what we are used to for eigenstates of the matrices of the form in Eq. (42).

For $B^0 - \bar{B}^0$ meson system the discussion deviates from the above results in two important ways. The first complication is that the CP is violated (but CPT still conserved). Then $M_{12} \neq M_{21}$, while $M_{11} = M_{22}$, in which case the two mass eigenstates are

$$|B_{L,H}\rangle = p|B^0\rangle \pm q|\bar{B}^0\rangle. \tag{46}$$

For CP conserving case $p = q = 1/\sqrt{2}$.

The other complication is that $B^0$ and $\bar{B}^0$ decay. We can describe this through a non-unitary evolution of a two-state system, given by a non-hermitian Hamiltonian

$$\mathcal{H} = M + i\Gamma, \tag{47}$$

so that the time evolution of a two-state system is described by

$$i\frac{d}{dt}\begin{pmatrix}|B^0(t)\rangle \\ |\bar{B}^0(t)\rangle\end{pmatrix} = \mathcal{H}\begin{pmatrix}|B^0(t)\rangle \\ |\bar{B}^0(t)\rangle\end{pmatrix} = \begin{pmatrix}M_{11} + i\Gamma_{11}, & M_{12} + i\Gamma_{12} \\ M_{21} + i\Gamma_{21}, & M_{22} + i\Gamma_{22}\end{pmatrix} \cdot \begin{pmatrix}|B^0(t)\rangle \\ |\bar{B}^0(t)\rangle.\end{pmatrix} \tag{48}$$

The $\Gamma$ matrix encodes the effects of $B^0$ and $\bar{B}^0$ decays on the time evolution. The non-unitary evolution describes the "disappearance" of $B^0$ and $\bar{B}^0$ states due to decays into final particles, i.e., outside of the two-state system, $|B^0\rangle$, $|\bar{B}^0\rangle$. The eigenstates of $\mathcal{H}$ are still given by Eq. (46), though now in general $|B_L\rangle$ and $|B_H\rangle$ are no longer orthogonal.

## 3.3 Different ways of measuring the CP violation

CP violation is an inherently quantum mechanical effect. As we saw in Section 2.8 it is intimately tied to the existence of a physical phase in the Lagrangian. In order to be sensitive to a phase an interference is needed. Thus, CP violating observables necessarily require some kind of interference. Depending on the type of interference there are three distinct categories of CP violating observables

1. *CPV in the decay*, also called *direct CPV*, occurs when there is interference between different contributions to the decay amplitudes so that

$$|A_f| \neq |\bar{A}_f|. \tag{49}$$

Here we used the short-hand notation

$$A_f \equiv \langle f|\mathcal{H}|B^0\rangle, \qquad \bar{A}_f \equiv \langle f|\mathcal{H}|\bar{B}^0\rangle. \tag{50}$$

2. *CPV in mixing* occurs when there is interference between $M_{12}$ and $\Gamma_{12}$ in the time evolution of the two-state system. This arises when

$$|q/p| \neq 1, \tag{51}$$

and corresponds to interference between different ways to oscillate between $B^0$ and $\bar{B}^0$ states, either through dispersive matrix elements or through absorptive ones.

3. *CPV in interference between decays with and without mixing*, arises when

$$\text{Im}\,\lambda_f \neq 0 \tag{52}$$

where

$$\lambda_f \equiv \frac{q}{p}\frac{\bar{A}_f}{A_f}. \tag{53}$$

Here the interference is between two different paths of $B^0$ to decay to the final states $f$, see Fig. 9. The two paths are either through direct decay, proportional to $A_f$, or by first oscillating to $\bar{B}^0$, which then decays to $f$, giving a contribution proportional to $(q/p)\bar{A}_f$.
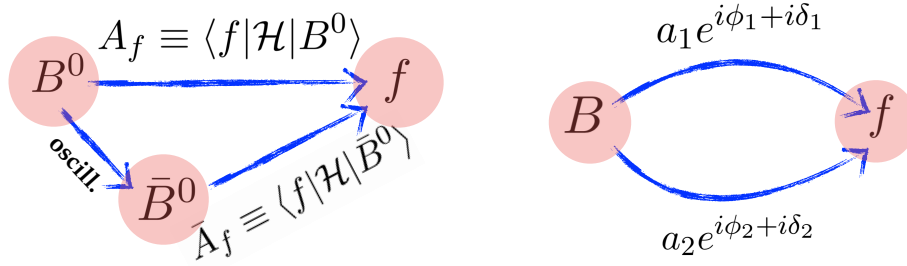
**Fig. 9:** Left: The two different paths for a $B^0$ meson to decay to a final sate $f$. Right: two interfering amplitudes are required to have direct CPV.

In the next few subsections we will look at two examples: the CPV in the decay, essential to measure the angle $\gamma$ of the standard CKM unitarity triangle, and the determination of angle $\beta$, which relies on the CPV in interference between decays with and without mixing for $B_d$ mesons.

When discussing CPV, it is important to remember that not all of the phases are CP violating. An elementary example is the double slit experiment, in which the interference pattern arises because there is a phase difference between two waves due to different paths, $e^{i\delta} = e^{ik\Delta r}$. This phase difference is not CP violating, since it does not depend on whether the double slit experiment is done with particles or antiparticles.

We thus distinguish two different types of phases. The *weak phases* are the (physical) phases that appear in the Lagrangian. The weak phases violate CP, just as the CKM phase in the weak interaction part of the SM Lagrangian violates CP. The *strong phase* is the name used for CP conserving phases. An example of such a strong phase is, for instance, the phase shift resulting from rescattering of particles due to QCD/strong interactions. Imagine a thought experiment, in which we collide two pion beams with the total center of mass energy close to the rho meson mass. The $\pi^+\pi^0 \to \rho^+ \to \pi^+\pi^0$ and $\pi^-\pi^0 \to \rho^- \to \pi^-\pi^0$ scatterings both result in the same complex Breit-Wigner scattering amplitude

$$A \propto \frac{1}{p^2 - m^2 + im\Gamma}. \tag{54}$$

The imaginary term in the propagator is due to on-shell rescattering through decay products of $\rho$. These are CP conserving processes, and so is the resulting phase, $\arg(A)$. This phase does not change sign, when we exchange $\pi^+ \leftrightarrow \pi^-$.

### 3.4 CPV in the decay

We start with the CPV in the decay (or the so called direct CP violation). The CPV observable is the decay asymmetry

$$\mathcal{A}_f \equiv \frac{\Gamma(\bar{B} \to \bar{f}) - \Gamma(B \to f)}{\Gamma(\bar{B} \to \bar{f}) + \Gamma(B \to f)} = \frac{1 - |A_f/\bar{A}_f|^2}{1 - |A_f/\bar{A}_f|^2}, \tag{55}$$

where $A_f$ are defined in (50). In order to have non-vanishing CP asymmetry, $\mathcal{A}_f \neq 0$, the $B \to f$ decay amplitude needs to receive contributions from (at least) two different terms with differing weak, $\phi_{1,2}$, and strong phases, $\delta_{1,2}$, see also Fig. 9 (right)

$$A_f = a_1 e^{i\phi_1 + i\delta_1} + a_2 e^{i\phi_2 + i\delta_2}, \tag{56}$$

$$\bar{A}_f = a_1 e^{-i\phi_1 + i\delta_1} + a_2 e^{-i\phi_2 + i\delta_2}. \tag{57}$$

The weak phases are due to the CKM phase in the SM Lagrangian and change the sign under CP transformation, while the strong phases are due to on-shell rescattering of particles (pions, etc) and are thus
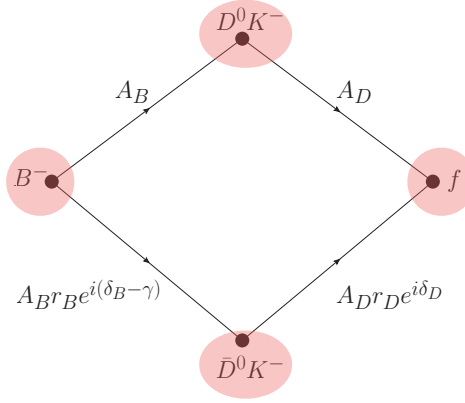
**Fig. 10:** The two interfering amplitudes in the $B^- \to D[\to f]K^-$ decays. The decay amplitudes for CP conjugate decay $B^+ \to D[\to \bar{f}]K^+$ are obtained by exchanging $\{B^-, K^-, f\} \to \{B^+, K^+, \bar{f}\}$, $D^0 \leftrightarrow \bar{D}^0$, and $\gamma \to -\gamma$ in the above.

CP even, the same as QCD interactions. The CP asymmetry is, in the simplifying limit $a_2/a_1 \ll 1$,

$$\mathcal{A}_f = \frac{a_2}{a_1}\sin(\phi_2 - \phi_1)\sin(\delta_2 - \delta_1) + \mathcal{O}(a_2^2/a_1^2). \tag{58}$$

The CP asymmetry vanishes in the limit where either (i) there is only one contribution to the amplitude, $a_2 \to 0$, and/or (ii) if the weak phase difference vanishes, $\phi_2 - \phi_1 \to 0$, and/or (iii) if the strong phase difference vanishes, $\delta_2 - \delta_1 \to 0$.

### 3.5 Measuring the CKM angle $\gamma$

The measurements of CKM unitarity triangle angle $\gamma$ use the decays in which there is interference between $b \to c\bar{u}s$ and $b \to u\bar{c}s$ transitions [23–26]. This happens for instance in the $B^- \to [D \to f]K^-$ decay chain. The $B^- \to D^0 K^-$ decay is due to the $b \to c\bar{u}s$ transition, while the $B^- \to \bar{D}^0 K^-$ decay is mediated by the $b \to u\bar{c}s$ transition, which is proportional to $V_{ub} \propto e^{-i\gamma}$, see the top two diagrams in Fig. 6. If the $D^0$ and $\bar{D}^0$ decay to the same final state, such as $f = \pi^+\pi^-$, $K^+K^-$, $K_S\pi^+\pi^-$, the two decay amplitudes interfere, giving sensitivity to the phase $\delta_B - \gamma$. Our notation is defined in Fig. 10, with $\delta_{B,D}$ the strong phases, while $\gamma$ is a weak phase and changes sign under CP conjugation.

In order to extract $\gamma$ both the rates for $B^- \to [D \to f]K^-$ and its CP conjugated mode $B^+ \to [D \to \bar{f}]K^+$ need to be measured. The interference terms in the two rates are proportional to $\delta_B + \delta_D - \gamma$ and $\delta_B + \delta_D + \gamma$, respectively. The difference of the two thus gives the quantity we are after, $\gamma$, if the hadronic parameters, $A_{B,D}, r_{B,D}, \delta_{B,D}$ are known. Note that the direct CPV asymmetries, Eq. (55),

$$\mathcal{A}_f \propto r_B r_D \sin(\delta_B + \delta_D)\sin\gamma, \tag{59}$$

are crucial. If $\mathcal{A}_f$ vanish, so thus the sensitivity to $\gamma$. The measurement of $\gamma$ requires both $r_B, r_D$ and the strong phases to be nonzero.

Amazingly, all the hadronic inputs can be measured experimentally. The $A_D$ and $r_D$ are obtained from $D^{*+} \to [D^0 \to f]\pi^+$ decays where the charge of $\pi^+$ tags the flavour of $D^0$, i.e., $\bar{D}^0$ would be accompanied by a $\pi^-$. Choosing $N_f$ different final states leaves us with $4 + N_f$ unknowns: $\gamma, A_B, r_B, \delta_B, \delta_D$. On the other hand, we can measure $2N_f$ decay branching ratios, $B^- \to [D \to f]K^-$ and $B^+ \to [D \to f]K^+$ (taking $f$ not to be CP conjugate final state for simplicity, such as bins in $K_S\pi^+\pi^-$ Dalitz plot). For $N_f \geq 4$ there is enough information to extract all the unknowns. The situation is in fact even better, since $\delta_D$ can be measured at CLEO and BESS III from entangled decays $\psi(3770) \to D^0\bar{D}^0$, improving the precision with which $\gamma$ is extracted.
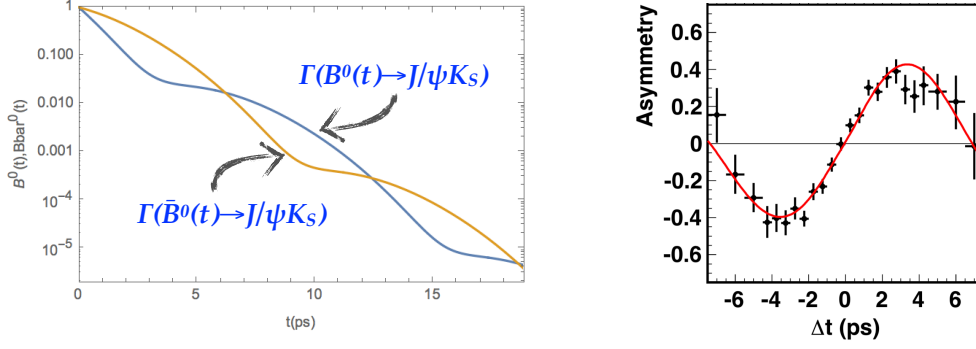
**Fig. 11:** Left: the time dependent $\bar{B}^0(t) \to J/\psi K_S$ (orange) and $B^0(t) \to J/\psi K_S$ (blue) decay rates. Right: the time dependent CP asymmetry (taken from [30]).

That all the hadronic uncertainties can be obtained experimentally makes this approach a very powerful tool. It means that the angle $\gamma$ can be extracted with basically no theory uncertainties. The theoretical corrections arise only from one loop electroweak corrections, limiting the ultimate precision with which $\gamma$ can be extracted up to miniscule $\gamma_{\rm th} < 10^{-7}$ [27, 28]. The experimental error bars will be larger than this for a long time. At present they are at $\delta\gamma \lesssim 6°$ [29].

### 3.6 CPV in interference between decays with and without mixing

The state that is created at $t = 0$ as the $\bar{B}^0$ [or $B^0$] time evolves according to

$$\frac{d}{dt}\Gamma(\bar{B}^0(t)[B^0(t)] \to f_{\rm CP}) \propto e^{-\Gamma t}\left[\frac{1}{2}\left(1 + |\lambda_f|^2\right) \pm S_f \sin(\Delta mt) \mp C_f \cos(\Delta mt)\right], \quad (60)$$

with,

$$S_f \equiv \frac{2\,{\rm Im}\,\lambda_f}{1 + |\lambda_f|^2}, \qquad C_f \equiv \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2}. \quad (61)$$

Here we assumed for simplicity that the final state is a CP eigenstate, $f_{CP}$, such as $f_{CP} = J/\psi K_S$. We also used that the mass splitting between the two mass eigenstates is much bigger than the difference between the two decay widths, $\Delta\Gamma \ll \Delta m$, so that it can be neglected, setting $|q/p| = 1$. The time evolution is plotted in Fig. 12 (left). The exponential decay is modulated by an oscillatory behaviour as $B^0$ converts to $\bar{B}^0$ and back (and vice versa), with the frequency of the oscillations given by the mass splitting, $\Delta m$.

The difference between the two decay rates is the time dependent CP asymmetry

$$\mathcal{A}_{f_{CP}}(t) \equiv \frac{\frac{d}{dt}\Gamma[\bar{B}^0(t) \to f_{\rm CP}] - \frac{d}{dt}\Gamma[B^0(t) \to f_{\rm CP}]}{\frac{d}{dt}\Gamma[\bar{B}^0(t) \to f_{\rm CP}] + \frac{d}{dt}\Gamma[B^0(t) \to f_{\rm CP}]}, \quad (62)$$

and is described by a purely oscillatory behaviour, see Fig. 12 (right),

$$\mathcal{A}_{f_{CP}}(t) = S_f \sin(\Delta mt) - C_f \cos(\Delta mt). \quad (63)$$

The coefficient of $\cos(\Delta mt)$ is nonzero, if there is direct CPV, since $C_f = -A_f$ for $|q/p| = 1$. The coefficient of $\sin(\Delta mt)$ is nonzero if there is CPV in interference between decays with and without mixing, cf. Eqs. (52) and (61). We will see that $S_f$ is an important observable in searches for New Physics (NP). In the SM it is a measure of the CKM unitarity triangle angle $\beta$.

### 3.7 The measurement of angle $\beta$

The $q/p$ does not depend on the final state $f$, and is the property of the $B^0 - \bar{B}^0$ system. In the SM it is given by the ratio of one loop diagram in Fig. 8 and its complex conjugated version, so that

$$\frac{q}{p} = e^{-i\phi_B} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}, \tag{64}$$

with hadronic matrix elements cancelling in the ratio.

The decay amplitudes $A_f$, $\bar{A}_f$ do depend on the final state. However, a simplification occurs for $B^0 \to J/\psi K_S$ and other decays that are dominated by a single amplitude, in this case due to the tree level $b \to c\bar{c}s$ transition. In the ratio $\bar{A}_f/A_f$ the hadronic matrix elements largely cancel. To a good approximation it is given by the ratio of the CKM elements

$$\frac{\bar{A}_{J/\psi K_S}}{A_{J/\psi K_S}} = \eta_f \frac{V_{cb} V_{cs}^*}{V_{cb}^* V_{cs}} + \cdots, \tag{65}$$

with $\eta_f = -1$ the CP of $J/\Psi K_S$, and the ellipses the corrections due to penguin diagrams that depend on a different product of CKM elements. Therefore,

$$\lambda_{J/\psi K_S} = \eta_f \frac{V_{tb}^* V_{td} V_{cb} V_{cs}^*}{V_{tb} V_{td}^* V_{cb}^* V_{cs}} = \eta_f e^{-i2\beta}, \tag{66}$$

and thus

$$\mathrm{Im}\, \lambda_{J/\psi K_S} = \sin 2\beta. \tag{67}$$

The measurement of $\sin 2\beta$ was the flagship measurement of the $B$ factories which showed that the CP violating phase in the SM is large, cf. Fig. 5.

## 4 New Physics searches

So far we looked at the measurements of the SM parameters. We now turn to a different question: how does one search for New Physics (NP)? Before we tackle this question let us first answer a seemingly unrelated question: why is the weak force weak? The weak and strong interactions are similar in many respects. They are both nonabelian gauge interactions, and at high energies, of a few 100 GeV, they even have coupling constants that are not that different in size. However, at low energies they exhibit very different strengths. The strong force gives rise to a strong binding potential, while the weak force results only in a very weak short range potential. The decays that proceed through strong interactions such as $\rho^+ \to \pi^+ \pi^0$ occur at times scales $\sim 10^{-23}$ s, while the weak decays are much slower, from $\sim 10^{-12}$ s for $B$ decays to hundreds of seconds in the case of neutron beta decay.

The reason for this disparity is that the strength of the interaction is governed both by the size of the couplings and the mass of the force carriers. The more massive the carrier the shorter the range of the potential, and the weaker the interaction at low energies. The weak force is weak because the force carriers, $W$ and $Z$ are heavy, with masses equal to $80.4$ GeV and $91.2$ GeV, respectively. The neutron beta decay width is highly suppressed, because the available energy in the decay, $\sim (m_p - m_n)$ up to corrections from electron mass, is much smaller than the mass of the force carrier, the $W$ boson,

$$\Gamma(n \to p e \bar{\nu}_e) \propto \frac{(m_p - m_n)^5}{m_W^4} \sim 10^{-20}(m_p - m_n). \tag{68}$$

This detour lead us to an important insight: through rare (or slowly occurring) processes we can probe heavy mediators. Historically, the weak nuclear decays were the first sign of a new force with a heavy mediator, the $W$ boson. Other processes could, in a similar way, hint at new forces beyond the
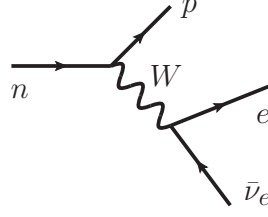
**Fig. 12:** The neutron beta decay proceeds through a tree level exchange of the $W$ boson.

SM. We thus arrived at the recipe for indirect searches: identify processes that are rare in the SM and then search for deviations from the SM predictions.

A good target are the Flavour Changing Neutral Current (FCNC) processes. In the SM there are no FCNCs at tree level – the gluon, photon, $Z$, and Higgs tree level exchanges are strictly flavour conserving, cf. Fig. 2. The FCNC processes, such as meson mixings, arise only at loop level and are thus suppressed, see Section 2.6. The FCNC processes can be easily modified by NP, either through tree level or loop level NP contributions. Taking $B_s$ mixing as an example the tree level NP contributions will have the form $\propto g_{sb}^2/M_{\rm NP}^2$, where $g_{sb}$ is the NP coupling to $b$ and $s$ quarks, and $M_{\rm NP}$ the mass of the new mediator. The NP contributions thus vanish if the NP is very heavy, $M_{\rm NP} \to \infty$, or if the flavour violating coupling constants are small, $g_{sb} \to 0$.

In the rest of this section we explore in more detail the two main ways of searching for beyond the SM physics in flavour: by measuring the meson mixing amplitudes, and by measuring rare decays such as $b \to s\ell^+\ell^-$.

### 4.1 New physics searches using meson mixings

There are four neutral meson systems that mix through weak interactions at 1 loop: $K^0 - \bar{K}^0$ ($\bar{s}d \leftrightarrow s\bar{d}$), $D^0 - \bar{D}^0$ ($c\bar{u} \leftrightarrow \bar{c}u$), $B^0 - \bar{B}^0$ ($\bar{b}d \leftrightarrow b\bar{d}$), and $B_s^0 - \bar{B}_s^0$ ($\bar{b}s \leftrightarrow b\bar{s}$). We will mainly focus on $B^0 - \bar{B}^0$ and $B_s^0 - \bar{B}_s^0$ systems, which are dominated by the $W$–top quark loop, Fig 8.

Since $m_{t,W} \gg m_B$ the top and $W$ can be integrated out, leading to the $B_d - \bar{B}_d$ mixing effective weak Hamiltonian [7]

$$\mathcal{H}_{\rm eff}^d = \frac{G_F^2}{16\pi^2} m_W^2 \eta_B S_0 \left(V_{tb}^* V_{td}\right)^2 \left(\bar{b}d\right)_{V-A} \left(\bar{b}d\right)_{V-A} + {\rm h.c.}, \tag{69}$$

where $G_F \simeq 1.166 \cdot 10^{-5}\,{\rm GeV}^{-2}$ is the Fermi constant, and $\eta_B S_0 \simeq 1.26$ is the product of a properly normalized loop function and the QCD correction factor. The effective weak Hamiltonian is local, i.e., it corresponds to the potential that acts only at a point. This is a result of taking the weak mediators to be infinitely heavy. Another way of writing the effective Hamiltonian is

$$\mathcal{H}_{\rm eff}^d = \frac{1}{\Lambda_{\rm MFV}^2} \left(V_{tb}^* V_{td}\right)^2 \left(\bar{b}_L \gamma^\mu d_L\right) \left(\bar{b}_L \gamma_\mu d_L\right) + {\rm h.c.}, \tag{70}$$

where the dimensionful prefactor,

$$\Lambda_{\rm MFV} = \frac{2\pi}{G_F m_W \sqrt{\eta_B S_0}} \simeq 6.0\,{\rm TeV}, \tag{71}$$

is significantly larger than the weak scale, $m_W \simeq 80.2\,{\rm GeV}$, because we absorbed in it the loop factor, $1/16\pi^2$ (up to a factor of 4 that went into a redefinition of the operator). For $B_s$ mixing the CKM factors in the weak vertices change, cf. Fig. 8, so that one has instead

$$\mathcal{H}_{\rm eff}^s = \frac{1}{\Lambda_{\rm MFV}^2} \left(V_{tb}^* V_{ts}\right)^2 \left(\bar{b}_L \gamma^\mu s_L\right) \left(\bar{b}_L \gamma_\mu s_L\right) + {\rm h.c.}. \tag{72}$$

The top and $W$ in the loop are much heavier than the available energy in the mixing – the $B$ meson mass. The top and $W$ lines in the diagram are thus always off-shell, and so the dominant diagram only contributes to the dispersive part of the mixing amplitude,

$$M_{12}^d = \frac{1}{2m_B} \langle \bar{B}_d^0 | \mathcal{H}_{\text{eff}}^d | B_d^0 \rangle^*. \tag{73}$$

The absorptive part of the mixing amplitude, $\Gamma_{12}$, receives contributions from the subleading amplitudes with $c$ and $u$ quarks running in the loop.

When NP is present, then $\mathcal{H}_{\text{eff}}^q = \mathcal{H}_{\text{eff},q}^{\text{SM}} + \mathcal{H}_{\text{eff},q}^{\text{NP}}$, and we can write

$$M_{12}^q = M_{12,q}^{\text{SM}} + M_{12,q}^{\text{NP}} = M_{12,q}^{\text{SM}} \Big[ 1 + \big(A_q^{\text{NP}}/A_q^{\text{SM}}\big) e^{i\phi_q^{\text{NP}}} \Big], \qquad q = d, s, \tag{74}$$

for $B_d$ and $B_s$ systems, respectively. The above parametrization is completely general, as long as NP is heavy, so that it does not appear in the decays of $B_{d,s}$ mesons. There is no NP contribution, when $A_{d,s}^{\text{NP}} = 0$. If $\phi_{d,s}^{\text{NP}} \neq 0$ this means that there are new CP violating phases in the NP contribution, beyond the CKM one. At present $A_{d,s}^{\text{NP}}/A_{d,s}^{\text{SM}}$ of about 0.2 are still allowed, depending on the NP phase. With the future measurements at Belle II and LHCb this will be drastically improved to less than about 0.05, see Fig. 13.

What does this mean in terms of bounds on NP masses? Let us assume that NP has the same $(V-A) \times (V-A)$ structure as the SM, so that the effective Hamiltonian is ($q = d, s$)

$$\mathcal{H}_{\text{eff}} = \left( \frac{\big(V_{tb}^* V_{tq}\big)^2}{\Lambda_{\text{MFV}}^2} + \frac{C_{\text{NP}}}{\Lambda_{\text{NP}}^2} \right) \big(\bar{b}_L \gamma^\mu q_L\big)\big(\bar{b}_L \gamma_\mu q_L\big) + \text{h.c.}. \tag{75}$$

For instance, the new physics contribution, $C_{\text{N}}/\Lambda_{\text{NP}}^2$ could be due to the $Z'$ exchange. This would give for the effective Hamiltonian

$$\mathcal{H}_{\text{eff}} = i(ig_{Z'})^2 \big(\bar{b}_L \gamma_\mu q_L\big) \frac{-ig^{\mu\nu}}{q^2 - m_{Z'}^2} \big(\bar{b}_L \gamma_\nu q_L\big) \rightarrow \frac{g_{Z'}^2}{m_{Z'}^2} \big(\bar{b}_L \gamma^\mu q_L\big)\big(\bar{b}_L \gamma_\mu q_L\big), \tag{76}$$

where $g_{Z'}$ is the flavour violating $Z'$ coupling to quarks, $m_{Z'}$ the $Z'$ mass, and in obtaining the last expression we used that $q^2 \ll m_{Z'}^2$. For the NP Wilson coefficient we thus have

$$\frac{C_{\text{NP}}}{\Lambda_{\text{NP}}^2} = \frac{g_{Z'}^2}{m_{Z'}^2}. \tag{77}$$

If $g_{Z'} = 1$, then $\Lambda_{\text{NP}}$ can be identified with $m_{Z'}$ for $C_{\text{NP}} = 1$.

In general NP will not have the $V - A$ structure. However, the choice of possible operator structure is still quite limited. The general dimension 6 operator basis for meson mixing contributions is [32]

$$\mathcal{H}_{\text{eff}}^{\text{NP}} = \sum_i \frac{C_i}{\Lambda_{\text{NP},B_q}^2} Q_{i,q}, \tag{78}$$

where

$$\begin{aligned}
Q_{1,q} &= (\bar{b}_L \gamma^\mu q_L)(\bar{b}_L \gamma^\mu q_L), \\
Q_{2,q} &= (\bar{b}_R q_L)(\bar{b}_R q_L), \\
Q_{3,q} &= (\bar{b}_R^\alpha q_L^\beta)(\bar{b}_R^\beta q_L^\alpha) \\
Q_{4,q} &= (\bar{b}_R q_L)(\bar{b}_L q_R), \\
Q_{5,q} &= (\bar{b}_R^\alpha q_L^\beta)(\bar{b}_L^\beta q_R^\alpha),
\end{aligned} \tag{79}$$
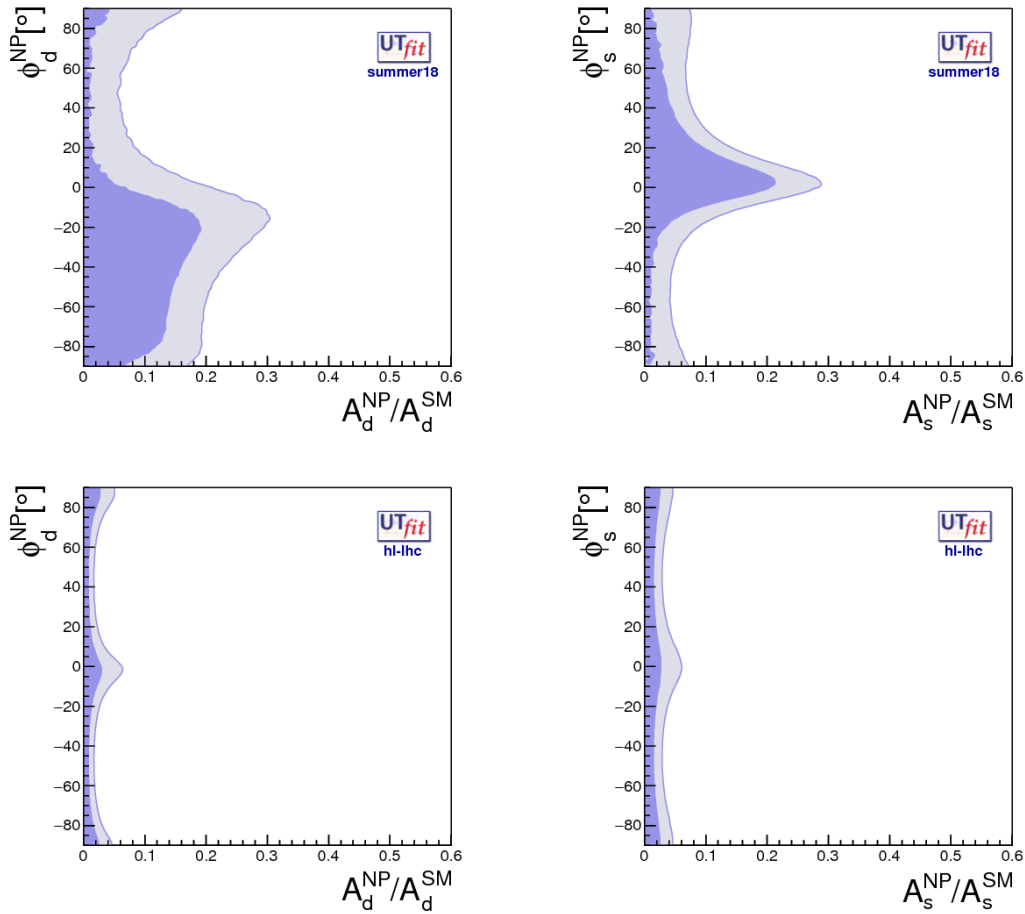
**Fig. 13:** Present (upper panels) and future (lower panels) constraints, at the end of Belle II, with 50 ab$^{-1}$ of integrated luminosity, and LHCb with 300 fb$^{-1}$, on the NP contributions to the $B_d$ (left) or $B_s$ (right) mixing amplitudes [31].

along with three other operators obtained from $Q_{i,q}$, $i = 1, 2, 3$ by replacing $L \leftrightarrow R$ (the bounds on these parity related operators are the same as for $Q_{i,q}$, $i = 1, 2, 3$, though). The operators for other meson systems are obtained through trivial replacements of quark flavours. Taking $|C_i| = 1$, the present bounds on the NP scale, $\Lambda_{\mathrm{NP}}$ are shown with ligher colors in Fig. 14 (left). The future projections to the end of LHCb Upgrade II are shown with darker colors. A jump in the mass reach is clearly visible even on the logarithmic scale.

Different colours in Fig. 14 denote different meson systems: green bars denote the constraints from $K^0 - \bar{K}^0$ mixing, yellow from $D^0 - \bar{D}^0$ mixing, in both cases assuming maximal new weak phase relative to the SM; orange bar denotes constraints from $B^0 - \bar{B}^0$, marginalized over the weak phase; red (blue) from $B_s^0 - \bar{B}_s^0$ system assuming no (maximal) NP phase. Extraction of constraints from $K - \bar{K}$ and $D - \bar{D}$ mixing is more complicated than for $B_q - \bar{B}_q$, since in these two cases the long distance contributions from light quarks running in the loop are important.

The bounds on $\Lambda_{\mathrm{NP}}$ are strikingly different for the various meson systems. This is easy to understand by considering the CKM suppression of the SM contributions, since the precision of experimental measurements and theoretical predicitons is typically at a fraction of the SM amplitude. Demanding for illustration that the contribution from NP is at most 20% of the short-distance SM this would give, for
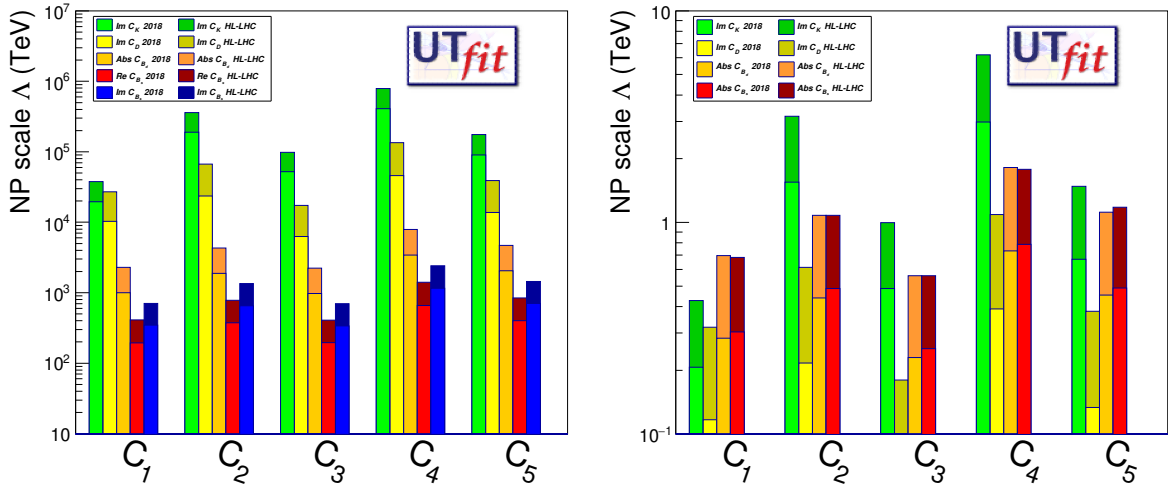
**Fig. 14:** Present constraints (lighter) and expected constraints at the end of LCHb upgrade II (darker) on the NP scale, $\Lambda_{\mathrm{NP}}$, from the UTfit NP analysis for different meson systems. The right panel shows constraints assuming NP is weakly coupled, has MFV structure of couplings, and enters observables only at one loop, see text for details (from [11]).

the operator $Q_{1,q}$

$$
\text{for NP=20\% SM,} \quad K - \bar{K}: \quad (\underbrace{V_{ts}^*}_{\lambda^2}\underbrace{V_{td}}_{\lambda^3})^2 \quad \Rightarrow \quad \Lambda_{\mathrm{NP}} \gtrsim 4 \cdot 10^4 \text{ TeV},
$$

$$
\text{for NP=20\% SM,} \quad B_d - \bar{B}_d: \quad (\underbrace{V_{tb}^*}_{1}\underbrace{V_{td}}_{\lambda^3})^2 \quad \Rightarrow \quad \Lambda_{\mathrm{NP}} \gtrsim 1.5 \cdot 10^3 \text{ TeV}, \tag{80}
$$

$$
\text{for NP=20\% SM,} \quad B_s - \bar{B}_s: \quad (\underbrace{V_{tb}^*}_{1}\underbrace{V_{ts}}_{\lambda^2})^2 \quad \Rightarrow \quad \Lambda_{\mathrm{NP}} \gtrsim 3 \cdot 10^2 \text{ TeV}.
$$

roughly in agreement with the constraints shown in Fig. 14.

Note that the interpretation of the bounds in term of NP scale crucially depends on the assumed flavour structure in the dimensionless Wilson coefficient, $C_i$. If the NP contribution also follows the SM CKM suppression, this is referred to as Minimal Flavour Violation (MFV). Fig. 14 (right) shows the bounds for the case of MFV type NP running in the loop, i.e., the Wilson coefficients were set to $C_a = (V_{ti}^*V_{tj})^2 g^4/16\pi^2$, with $V_{ti}, V_{tj}$ the appropriate SM CKM coefficients, and $g$ the weak coupling constant. We see that even for a weakly coupled NP that has the MFV flavour structure and only contributes at 1 loop, the bounds are in the few 100 GeV to few TeV range.

## 4.2 New physics searches using rare decays

We turn next to the other main pathway to searching for new physics - searching for deviations in rare decays. Here the benefit is that there are many observables in flavour physics: the branching ratios, asymmetries, distributions, ... There is also a choice of different parent particles as well as many possible final states. The abundance of observables is clearly illustrated by opening the "bible" of particle physics, the Particle Data Group (PDG) book [15]. Even the condensed version, the PDG booklet, clocks out at more than 170 pages.

To shorten the discussion we will focus on the processes that are at present showing deviations from the SM expectations. The present experimental situation can then succinctly be described in the
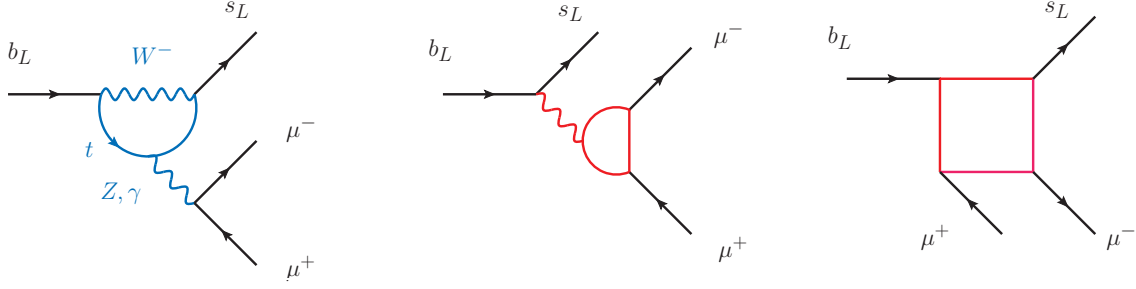
**Fig. 15:** A representative SM diagram for $b \to s\mu^+\mu^-$ transition (left), and the representative loop level NP contributions (middle and right).

following way. There are many different transitions that were measured, all of which agree with the SM expectations within experimental and theoretical errors. There are only two sets of quark level transitions that are showing $\sim 4\sigma$ deviations from the SM: the $b \to c\tau\nu$ and $b \to s\mu^+\mu^-$ transitions.[3] The apparent NP scale that explains the deviations is quite different in the two cases. For instance, if the NP is due to the following $V - A$ operator

$$\mathcal{L}_{\text{NP}} \supset \frac{1}{\Lambda_{\text{NP}}^2} \left( \bar{Q}_i \gamma^\mu \sigma^A Q_j \right) \left( \bar{L}_k \gamma_\mu \sigma^A L_l \right), \tag{81}$$

then $\Lambda_{\text{NP}} \sim 3$ TeV in order to explain the deviations in $b \to c\tau\nu$ transitions, and $\Lambda_{\text{NP}} \sim 30$ TeV in order to explain the $b \to s\mu^+\mu^-$ anomalies. We discuss next the possible NP explanations for each of the two.

### 4.3 New physics searches in $b \to s\mu^+\mu^-$ transitions

The upshot of the observed $b \to s\mu^+\mu^-$ anomaly is: choosing only the theoretically clean observables the excess is at the $\sim 4\sigma$ level. From the NP perspective the scale required to explain the anomaly makes sense, since it is high enough to avoid many of the experimental constraints. The models that explain the anomaly do, however, face I.I. Rabi's question:" Who ordered that?", when the muon was first discovered [36].

The FCNC $b \to s\ell^+\ell^-$ transitions are generated at 1-loop in the SM. A representative diagram in the SM is shown in Fig. 15 (left). Integrating out the heavy degrees of freedom, $W, Z, t$, gives the following effective Hamiltonian [37–39]

$$\mathcal{H}_{\text{eff}} = G_F V_{tb} V_{ts}^* \frac{\alpha}{4\pi} \left[ C_9 \left( \bar{s}_L \gamma^\mu b_L \right) \left( \bar{\ell} \gamma_\mu \ell \right) + C_{10} \left( \bar{s}_L \gamma^\mu b_L \right) \left( \bar{\ell} \gamma_\mu \gamma_5 \ell \right) \right], \tag{82}$$

where in the SM $C_9^{\text{SM}} \simeq -C_{10}^{\text{SM}}$, i.e., the SM diagrams give to a good approximation a $V - A$ structure of the leptonic current.

Another prediction of the SM is that the rates for the $b \to se^+e^-$ and $b \to s\mu^+\mu^-$ transitions should be equal to each other as soon as we are reasonably far above the muon production threshold so that the effect of muon mass on the available phase space can be neglected. The SM prediction of Lepton Flavour Universality (LFU) is deeply engrained in the structure of the theory, since it is a consequence of the fact that the electroweak gauge group is the same for all three generations.

The prediction of LFU can be tested experimentally by forming theoretically clean observables such as the ratios of $b \to s\mu\mu$ to $b \to see$ rates,

$$R_{K^{(*)}} = \frac{Br(B \to K^{(*)}\mu\mu)}{Br(B \to K^{(*)}ee)}. \tag{83}$$

---

[3]There are other interesting deviations, e.g., the $\sim 3\sigma$ deviation in $\epsilon'/\epsilon$, see, e.g., [33–35].
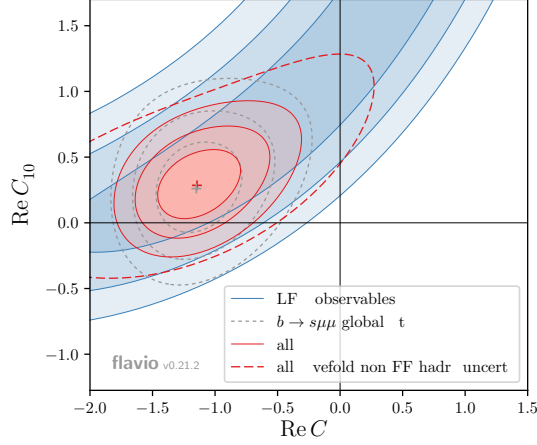
**Fig. 16:** The fit for NP contributions $C_{9,10}^{\mu}$ to Wilson coefficients in (82) from LFU only observables (red) or including all the observables (blue). The SM point is $C_{9,10}^{\mu} = 0$. (from [54].)

In the ratios the uncertainties from hadronic inputs (the form facors) cancel to a very good approximation. Above the muon threshold they are equal to 1 within a percent, and are also presicely predicted close to the muon threshold [40–47]. Experimentally, on the other hand $R_{K^{(*)}} \sim 0.7$ [48–51], violating LFU with a significance of $2.2 - 2.6\sigma$ in each of the three most precise measurements (the measurements are at different dilepton invariant masses). A combined significance for the discrepancy with the SM is $\sim 4\sigma$ [52–57]. The most precise measurements are due to LHCb, which dominates the world averages for $R_{K^{(*)}}$.

LFU ratios is not the only experimental information about the $b \to s\ell\ell$ transitions. In principle there is much more information available, branching ratios for different choices of initial and final state mesons, $Br(B \to K^{(*)}\mu\mu)$, $Br(B_s \to \phi\mu\mu)$, $Br(B \to X_s\mu\mu)$, angular observables in $B^0 \to K^{*0}\mu\mu, B_s \to \phi\mu\mu$, etc. However, the interpretation of these is much more sensitive to hadronic inputs. It requires form factor predictions (now coming from QCD sum rules), the estimate of charm loops, nonfactorizable contributions, etc. Using the best available estimates for these inputs the favored interpretation is that the NP is mostly in muons [52–57]. Furthermore, the picture obtained from such global fits to data seems to be in agreement with the LFU only determination, see Fig. 16 for a simultaneous fit to NP contributions in $C_{9,10}$.

If the anomaly is due to NP we thus already have a significant amount of information about it. First of all, there are only four dimension 6 operators that can explain $R_K$ [58]

$$\mathcal{O}_9^{(')\ell} = \frac{\alpha}{4\pi}\left(\bar{s}\gamma^\mu P_{L(R)}b\right)\left(\bar{\ell}\gamma_\mu\ell\right), \qquad \mathcal{O}_{10}^{(')\ell} = \frac{\alpha}{4\pi}\left(\bar{s}\gamma^\mu P_{L(R)}b\right)\left(\bar{\ell}\gamma_\mu\gamma_5\ell\right). \qquad (84)$$

The other operators are either constrained by $B_s \to \ell\ell$ as is the case for scalar currents, or come from further suppressed dimension 8 operators before electroweak symmetry is broken, as is the case for tensor operators.

Since the $K$ and $K^*$ in the final states differ in their spin-parity quantum numbers, one is pseudoscalar, the other vector meson, the ratios $R_K$ and $R_{K^*}$ give complementary information. For instance, for the central $q^2$ bins we have [52]

$$R_K \simeq 1 + 2\frac{\mathrm{Re}C_{b_{L+R}(\mu-e)_L}^{\mathrm{BSM}}}{C_{b_L\mu_L}^{\mathrm{SM}}}, \qquad R_{K^*} \simeq R_K - 4p\frac{\mathrm{Re}C_{b_R(\mu-e)_L}^{\mathrm{BSM}}}{C_{b_L\mu_L}^{\mathrm{SM}}}, \qquad (85)$$

when expanded to linear order in the BSM contributions to the Wilson coefficients (here $p \simeq 0.86$ is the polarization fraction of $K^*$). The resulting predictions for several choices of chirality in the NP

New physics in $\mu$ ⠀⠀⠀⠀⠀⠀ New physics in $e$
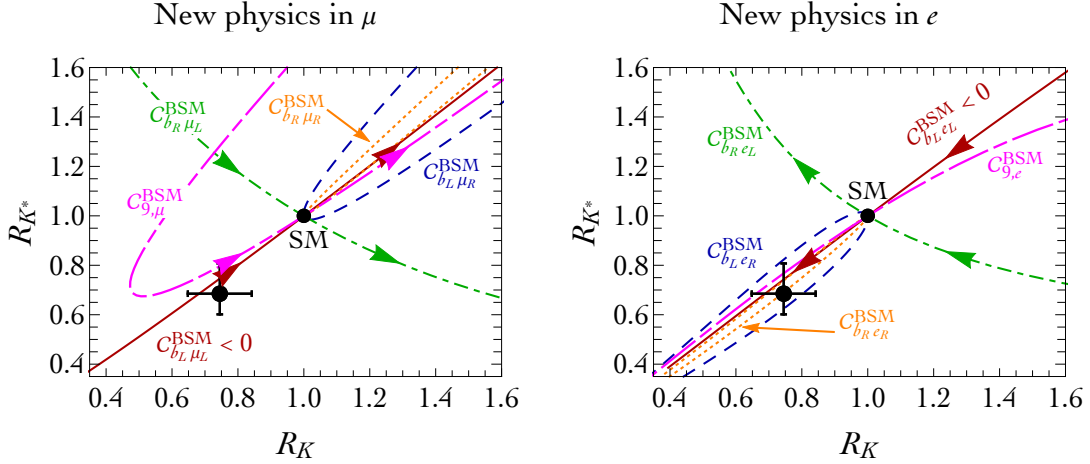


**Fig. 17:** The predictions for the central $q^2$ bins in $R_K$ and $R_{K^*}$ for several NP scenarios, assuming NP is only in muons (left) or only in electrons (right). The experimental values are given by the black error bars, the SM value by a point (from [52]).

contributions to the Wilson coefficients are shown in Fig. 17. Using just the "clean" observables, $R_K$ and $R_{K^*}$, NP can be either due to a deficit in muons or an increase in the electron channel. In both cases the operators with $(\bar{s}\gamma^\mu b)_L$ current can explain the anomaly, with significant freedom for the chirality of the leptonic current. For electrons also the NP due to $(\bar{s}\gamma^\mu b)_R(\bar{e}\gamma_\mu e)_R$ is possible. In this case the NP contribution enters only quadratically to a good approximation, since there is almost no interference with the SM predominantly $V-A$ leptonic current. The NP thus increases the rate for the electron channel, reducing the $R_{K^{(*)}}$ ratios below 1. It is only once additional observables, such as the absolute branching ratios, are taking into account that the possibility of NP right-handed currents is disfavored. These additional observables do require theoretical inputs and are subject to hadronic uncertainties.

For the remainder of this section let us assume that there is NP in $b \to s\mu^+\mu^-$. What kind of NP can explain it? There is significant freedom in the NP interpretations, since the associated scale is quite high. The Wilson coefficients shown in Fig. 16 are normalized to

$$V_{tb}V_{ts}^* \frac{\alpha}{4\pi v^2} C_I = \frac{C_I}{(36 \text{ TeV})^2}. \tag{86}$$

The NP scale of $\sim 30$ TeV is high enough that the NP can enter either at tree level, or even only at one loop level. The tree level NP models are of two distinct types. The mediator can be *(i)* a $Z'$, either an $SU(2)_L$ singlet or part of a triplet [59–61], or *(ii)* a leptoquark with either spin 0 or spin 1 [57,62]. The diagrams for the two types of mediators are shown in Fig. 18 middle and right, respectively.

There are 4 different possible charge assignments under the SM gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$ for a scalar leptoquark, and 3 for a vector leptoquark [57]. However, only one scalar leptoquark, $S_3 \sim (\bar{3}, 3, 1/3)$, and only two vector leptoquarks, $V_1 \sim (3, 1, 2/3)$ and $V_3 \sim (3, 3, -2/3)$, lead to $R_K \simeq R_{K^*} < 1$ in agreement with the data. All three predict $C_9^\mu = -C_{10}^\mu$. At 1-loop the leptoquarks contribute to $B_s - \bar{B}_s$ mixing, correcting the mass splitting by $\Delta m_{B_s} \propto (YY^*)^2/M^2$, where $M$ is the leptoquark mass, and $Y$ the relevant couplings to the SM fermions. The corrections to $R_{K^{(*)}}$, on the other hand, scale as $R_{K^{(*)}} - 1 \propto YY^*/M^2$. This means that the value of $Y$ required to explain $R_{K^{(*)}}$ grows faster with leptoquark mass than does the value of $Y$ still allowed by the $B_s - \bar{B}_s$ mixing constraints. In other words, the bound on allowed NP in $B_s - \bar{B}_s$ mixing implies un upper bound on the leptoquark mass, $M \lesssim 40$ TeV, 45 TeV, 20 TeV, for leptoquarks $S_3, V_1, V_3$, respectively [57].

The bounds on allowed NP contributions to $B_s$ mixing also imply a nontrivial constraint on the
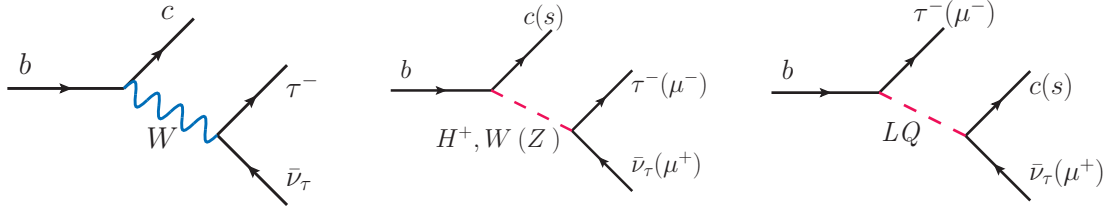
**Fig. 18:** The SM diagrams for $b \to c\tau\nu$ transition (left), and the possible tree level NP contributions to $b \to c\tau\nu$ or $b \to s\mu\mu$ transitions (middle and right).

models with $Z'$, since this contributes at tree level, giving [59, 63]

$$\frac{g_{bsZ'}}{m_{Z'}} \lesssim \frac{0.01}{2.5 \text{ TeV}}. \tag{87}$$

Thus a 2.5 TeV $Z'$ has to have a relatively small, but not extremely small, flavour violating coupling, $g_{bsZ'} \lesssim 0.01$ (comparable, for instance, with $|V_{ts}| \simeq 0.04$). This also means that the $Z'$ has to have sizeable couplings to muons. If the coupling is to left-handed muons, this implies nontrivial constraints from neutrino trident production in neutrino scattering on nuclei, i.e., from bounds on the process $\nu N \to \nu N \mu^+ \mu^-$ mediated by a $Z'$ [60, 64]. For couplings to left-handed muons the $b \to s\mu^+\mu^-$ is also accompanied by a $b \to s\bar{\nu}\nu$ signal, giving stringent constraints on the parameter space. Another important constraint are the $Z'$ searches at the LHC.

To recap, the NP explanations of the $b \to s\mu^+\mu^-$ anomaly should lead to new signals in a number of observables. The present constraints give meaningful bounds on the models already, but they are not too constraining. For instance, simply raising the mass of $Z'$ avoids the high $p_T$ constraints at the LHC. The bounds are more stringent for loop induced models [65–69], Fig. 15 (middle and right), since there the NP particles need to be lighter, below about a TeV.

### 4.4   New physics searches in $b \to c\tau\nu$ transitions

The $b \to c\tau\nu$ flavour anomaly is similarly very clean theoretically [70], and the disagreement with the SM predictions is also about $\sim 4\sigma$. However, the NP effect is large, $\mathcal{O}(20\%)$ of the SM tree level contribution given in Fig. 18 (left). This means that the scale of NP needs to be low, and consequently the NP interpretations are often in conflict with the other constraints.

The two main observables are

$$R(D^{(*)}) = \frac{\Gamma(\bar{B} \to D^{(*)}\tau\bar{\nu})}{\Gamma(\bar{B} \to D^{(*)}\ell\bar{\nu})}, \qquad \ell = \mu, e, \tag{88}$$

where $\bar{B}^+ \sim b\bar{u}$, $D \sim c\bar{u}$, etc. The SM predictions are shown in 19. Even though these are flavour universality ratios, the SM predictions are well below 1, because the $b \to c\tau\nu$ decays have much less final state phase space available due to the large $\tau$ mass. The thing to note is that the trend $R(D^{(*)})_{\text{exp}} > R(D^{(*)})_{\text{SM}}$ is seen in several experiments. Furthermore, the theoretical predicitons are well under control. Another comment is that, since the neutrino is not seen in the expriments, it does not need to be the SM neutrino. It could be a new state, possibly even of right-handed chirality [71–74].

What kind of NP could explain this anomaly? The most obvious candidates are ruled out. Theoretical bias would have been that the new charged currents are either due to a charged Higgs, $H^+$, or a new vector boson, $W'$, see Fig. 18 (middle). The charged Higgs option is in conflict with total $B_c$ lifetime [87], the $b \to c\tau\nu$ leptonic mass distributions, and searches in $pp \to \tau^+\tau^-$ [88]. The $W'$ is
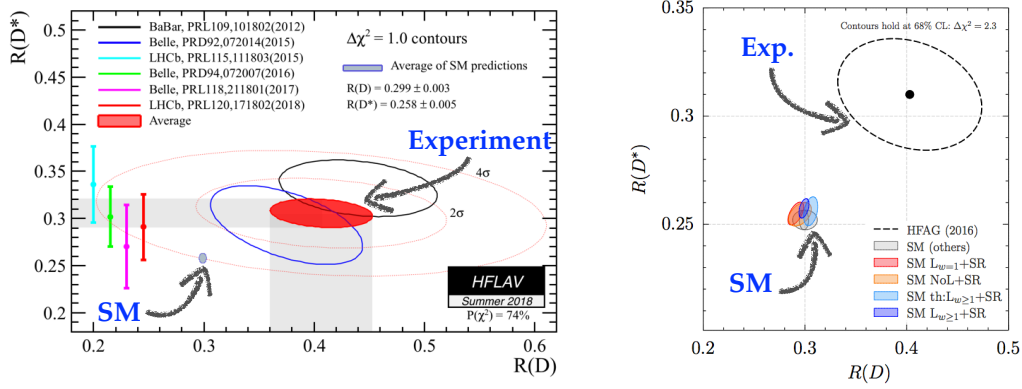
**Fig. 19:** Left panel: the measurements of $R(D)$ and $R(D^*)$ by different experiments [75–82], with the world average shown in red and the SM prediction in blue [83–86]. Right panel shows that the variation in the SM predictions is small even when some of the theoretical constraints are relaxed (from [84]).

excluded by $pp \to \tau$+MET searches at the LHC [89], while in addition the $pp \to \tau^+\tau^-$ in conjunction with $B$ mixing constraints exclude the related $Z'$.

There are several viable leptoquark solutions, both with SM neutrinos [90] and for right-handed neutrinos [91]. The vector leptoquark $V_1$ also allows to simultaneously explain the $b \to c\tau\nu$ and the $b \to s\mu^+\mu^-$ anomalies [92]. A simultaneous explanation is also possible, if there are more than one scalar leptoquarks contributing [93].

### 4.5 Other modes

Besides the two quark level transitions that are showing experimental discrepancies there are a number of other rare decays that are important probes of NP. The useful rare decays are such that we can predict them precisely and that NP contributions are possible or even likely. The modes with only one final state hadron, $K \to \pi\nu\nu$, $B \to K\ell\ell$,..., fall into this category. The hadronic matrix elements for these decays are easier to predict than for the fully hadronic decays. Another example are inclusive decays, where one sums over all hadronic final states, which are also easier to predict theoretically. We look at one important example for each of these two categories.

The inclusive $b \to s\gamma$ decay is a classic example of a GIM suppressed loop induced SM process. The loop contributions that do not depend on masses of the quarks running in the loop cancel due to CKM unitarity, $M \propto \sum_i V_{ib}^* V_{is} = 0$. The first nonzero contribution is thus proportional to mass differences of the quarks on the internal line in Fig. 20 (right). The SM contribution is finite, since it is described by the effective Hamiltonian of dimension 5 [38]

$$\mathcal{H}_{\text{eff}} = -\frac{G_F}{\sqrt{2}} V_{ts}^* V_{tb} C_{7\gamma}(m_b) \frac{e}{4\pi^2} m_b \big(\bar{s}_L \sigma^{\mu\nu} b_R\big) F_{\mu\nu}. \tag{89}$$

In the renormalizable SM Lagrangian there is no such counter-term, thus the contribution needs to be finite. The operator in (89) is chirality flipping. In the SM the chirality flip occurs on the external leg, and is thus proportional to the $b$ quark mass, $m_b$. NP contributions, on the other hand, can have the chirality flip on the internal line, leading to a relative enhancement of the NP contributions compared to the SM. This happens for instance in the Minimal Supersymmetric Standard Model (MSSM) for the gluino-squark diagram, or for the exchange of a charged Higgs in the loop. The measurements of $b \to s\gamma$ are therefore very sensitive to such NP contributions. A great theoretical effort has thus been devoted to obtain a precise theoretical prediction for the SM $b \to s\gamma$ rate [94].
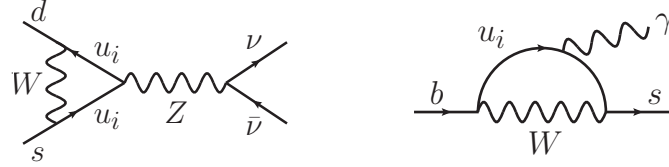
**Fig. 20:** The representative SM diagrams for $K \to \pi\nu\bar{\nu}$ (left) and $b \to s\gamma$ transitions (right).

The decays $K^+ \to \pi^+\nu\bar{\nu}$ and $K_L \to \pi^0\nu\bar{\nu}$ stand out, since these are one of the few rare decays in kaon sector that are the golden modes for NP searches. They are suppressed by a loop factor, CKM factors, and the GIM mechanism. They are also extremely well predicted theoretically [95]

$$Br(K^+ \to \pi^+\nu\bar{\nu}) = (8.4 \pm 1.0) \times 10^{-11},$$
$$Br(K_L \to \pi^0\nu\bar{\nu}) = (3.4 \pm 0.6) \times 10^{-11},$$

(90)

because the hadronic matrix elements, $\langle\pi|(\bar{s}d)_{V-A}|K\rangle$, are known precisely – they are extracted from data on $K^+ \to \pi^0 e^+\nu$ using isospin symmetry. The largest uncertainties are from the CKM inputs, $V_{cb}$ and $\gamma$, which will be improved in the future.

The experimental challenge is that the two processes are very rare. They are set to be measured by the NA62 experiment at CERN [96], and KOTO at J-PARC [97], respectively, even if the rates are at the SM values. On the positive note, since these decays are so suppressed, the scales probed are very high, $\sim 10^3$ TeV for $Z'$ models with $\mathcal{O}(1)$ couplings. On top of this, $K_L \to \pi^0\nu\bar{\nu}$ is also CP violating.

### 4.6 The future of NP searches with rare decays

The NP searches with rare decays, as well as the tests of the CKM unitarity will receive a significant boost with the upcoming Belle II and LHCb upgrades. Belle II expects to collect 50 times the Belle dataset. First collisions were seen in May 2018, and the first $B$ physics run is expected in March 2019. LHCb after upgrade II aims for roughly 100 times the present data set with an upgraded detector. A rule of thumb on the improved NP reach gives, for instance for Belle II, that the reach in $\Lambda_{\mathrm{NP}}$ will be improved by $\sim \sqrt[4]{50} = 2.7\times$. Similar if not larger increase applies to LHCb Upgrade II sensitivity improvements. This is a similar jump in energy reach as going from 13TeV LHC to a 35TeV LHC!

Among other things this also means that, if the two anomalies discussed in Sections 4.3, 4.4 are not mere statistical fluctuations, we should have available measurements with $5\sigma$ significance in a relatively near future.

### 5 Higgs and flavour

In the SM all flavour structure is due to the Higgs Yukawa couplings, $y_f = \sqrt{2}m_f/v$. The very hierarchical values of fermion masses therefore imply similarly very hierarchical Yukawa couplings. How well have we tested this? There are a number of tests that are experimentally accessible to different degrees of accuracy [98]

1. proportionality: is $y_{ii} \propto m_i$?
2. factor of proportionality: is $y_{ii}/m_i = \sqrt{2}/v$?
3. diagonality (flavour violation): is $y_{ij} = 0$ for $i \neq j$?
4. reality (CP violation): is $\mathrm{Im}(y_{ij}) = 0$?

Each of these questions probes a slightly different set of NP models. The proportionality, $y_{ii} \propto m_i$, and factor of proportionality, $y_{ii}/m_i = \sqrt{2}/v$, are relatively well tested for 3rd generation fermions, i.e.,
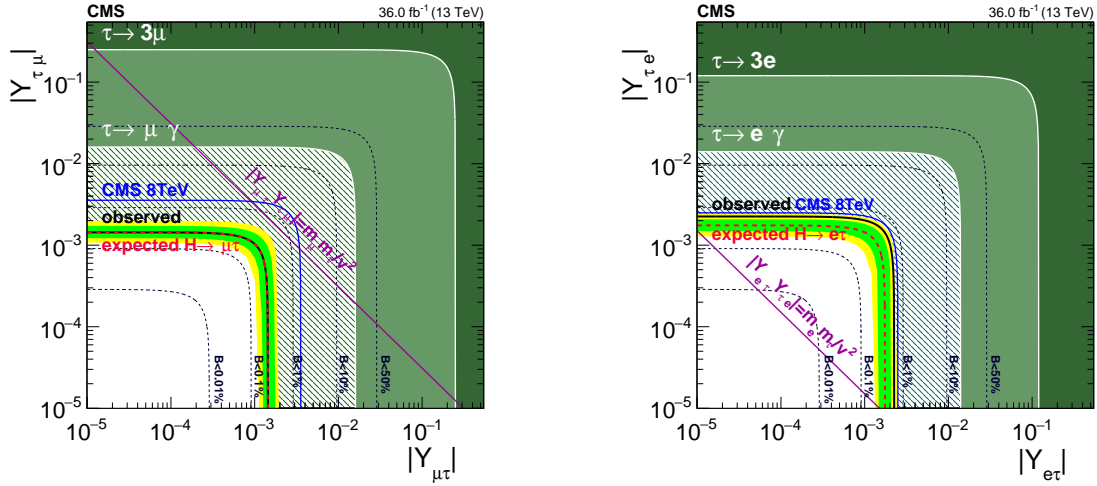
**Fig. 21:** The constraints on the Higgs flavour violating couplings to $\tau\mu$ (left) and $\tau e$ (right) (from [105]).
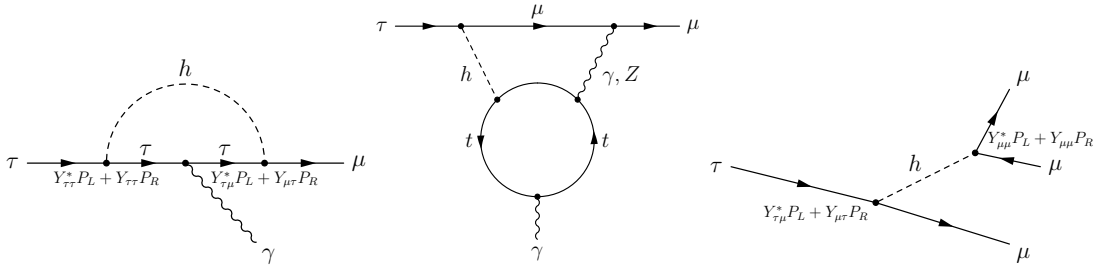


**Fig. 22:** Representative diagrams for $\tau \to \mu\gamma$ induced by Higgs flavour violating couplings at 1 loop (left) and 2 loops (middle), as well as the tree level diagram leading to $\tau \to 3\mu$ (from [106]).

the Higgs couplings to top, bottom and tau. Experimentally much more difficult question is how Higgs couples to the first two generations. This is difficult to address since the SM Yukawa couplings are so small. A more modest question is: can we show that the couplings are hierarchical? The answer is already now a positive one, though for quarks this is achieved with some assumptions. Experimentally [99–102],

$$\frac{Y^{\text{exp}}_{e(\mu)}}{Y^{\text{exp}}_{\tau}} < 0.22(0.10), \qquad \frac{Y^{\text{exp}}_{u(c)}}{Y^{\text{exp}}_{t}} \lesssim 0.04, \qquad \frac{Y^{\text{exp}}_{d(s)}}{Y^{\text{exp}}_{t}} < 0.7(6), \qquad (91)$$

where the bounds for leptons come from direct measurements, on up quarks from a global fit, and on down quarks from Higgs $p_T$ distributions (global fit).

Pushing these bounds to the SM values would be very challenging, if not impossible. The one bright exception is the muon Yukawa, which will become accessible at the high-luminosity LHC as the only one among the first two generations of fermions [103]. This is quite exciting, since it is easy to imagine that part of the muon mass comes not from the SM Higgs vev, but from new small sources of the electroweak breaking (see, e.g., [104]). The muon Yukawa could deviate significantly from the SM, in the extreme case it could even be zero.

Another important NP test are searches for flavour violating Higgs couplings. In the SM Higgs couplings are flavour diagonal (up to very small 1-loop corrections). Discovering flavour violating Higgs couplings would thus immediately mean New Physics. For charged lepton final states these couplings are accessible directly, by searching for $h \to \tau\mu, \tau e$ decays [106, 107]. The resulting bounds are shown

in Fig. 21. If the NP corrections come from dimension 6 operators then the Higgs Yukawa couplings are,

$$Y_{ij} = \frac{m_i}{v}\delta_{ij} + \frac{v^2}{\sqrt{2}\Lambda^2}\hat{\lambda}_{ij}. \tag{92}$$

The present bounds give for the NP scale, $\Lambda_{\mu\tau} > 5.5$ TeV, $\Lambda_{e\tau} > 4.4$ TeV, taking $\hat{\lambda}_{ij} = 1$. The Higgs decay measurements thus already probe interesting NP scales. There are also indirect bounds on flavour violating Higgs Yukawas that come from charged lepton FCNC transition. The $\tau \to \mu\gamma$ and $\tau \to 3\mu$ are induced by the diagrams shown in Fig. 22. While these lead to less stringent constraints on flavour violating couplings of the Higgs, see Fig. 21, this is not the case for $h \to \mu e$ decays, where the bounds on $\mu \to e\gamma$ limits the branching ratio to $Br(h \to \mu e) \lesssim 10^{-8}$, barring cancellations.

## 6 Conclusions

In the SM the flavour violation and CP violation are due to the Higgs couplings to the charged fermions. Experimentally, we know that the CKM is the dominant source of flavour violation in Nature, with the CKM phase responsible for the bulk of the CP violation in quark transitions. New physics contributions at the level of $\mathcal{O}(20\%)$ of the SM amplitude are still allowed, e.g., in the meson mixing.

Most of the measured flavoured transitions agree with the SM predictions, with the possible exception of two quark level transitions, $b \to s\mu^+\mu^-$ and $b \to c\tau\nu$, which show $\sim 4\sigma$ discrepancies with the SM predictions. If true, this would imply many new signals in both high $p_T$ processes measured by CMS and ATLAS, as well as in precision flavour experiments LHCb, Belle II, NA62, KOTO, the muon $g-2$ experiment, etc.

There are many excellent reviews and books that go beyond the scope of these lectures, some of which were mentioned in the Introduction. A good starting point for exploring the scope of future flavour programmes at LHCb and Belle II can be found in [10,21], and for general flavour physics possibilities at high-luminosity LHC in [11]. A good starting point for a study of new physics models that are bounded by flavour physics measurements is the introductory book [108], or the somewhat more detailed, albeit older Ref. [8].

## Appendices

## A Nonzero neutrino masses

When the neutrino masses are included, the counting of physical parameters in the SM changes from what was given in Section 2.4. We show this for (i) the case that the neutrinos are Majorana fermions and (ii) the neutrinos are Dirac.

For the case of Dirac neutrinos we enlarge the SM field content, Eqs. (3), (4) by three right-handed neutrinos that are complete singlets under the SM gauge group,

$$\nu_{R,i} \sim (1,1)_0. \tag{A.1}$$

The Yukawa interaction Lagrangian is then enlarged by the

$$\mathcal{L}_{\text{Yukawa}} \supset -Y_\nu^{ij}\bar{L}_L^i H^c \nu_R^j + \text{h.c.}, \tag{A.2}$$

while we assume that the Majorana mass terms, $m_{ij}\bar{\nu}_{R,i}^c\nu_{R,j}$, are forbidden by the conservation of total lepton number which this term would violate by two units. The counting of the physical parameters for

the leptons is now completely analogous to the counting we did for the quarks in Section 2.4. Using unitary transformations

$$L_L \to V_L L_L, \qquad \ell_R \to V_\ell \ell_R, \qquad \nu_R \to V_\nu \nu_R, \tag{A.3}$$

one can bring the lepton Yukawa couplings to the form

$$Y_\ell = \mathrm{diag}(y_e, y_\mu, y_\tau), \qquad Y_\nu = V_{\mathrm{PMNS}}^\dagger \mathrm{diag}(y_\nu^1, y_\nu^2, y_\nu^3). \tag{A.4}$$

The $3 \times 3$ unitary Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [109, 110] is the analogue of the CKM matrix for the quarks. It has three mixing angles and one physical phase. The remaining six real parameters are the three charged lepton masses and the three neutrino masses.

This agrees with the counting of physical parameters that follows from the general rule in Eq. (18). The $Y_\ell, Y_\nu$ matrices have $2 \times (9\ \mathrm{real} + 9\ \mathrm{im.})$ parameters, while the three unitary matrices, $V_L, V_\ell, V_\nu$ have in total $3 \times (3\ \mathrm{real} + 6\ \mathrm{im.})$ parameters. Out of these one corresponds to an unbroken generator, the lepton number, under which all the lepton fields change by the same phase, $L_L \to \exp(i\phi)L_L, \ell_R \to \exp(i\phi)\ell_R, \nu_R \to \exp(i\phi)\nu_R$. Using (18) there are $2 \times 9 - 3 \times 3 = 9$ real parameters, the six leptonic masses and three PMNS mixing angles, and $2 \times 9 - (3 \times 6 - 1) = 1$ imaginary physical parameter, the phase in the PMNS matrix, as anticipated.

If the neutrinos are Majorana, the field content is the same as for the SM with neutrino masses set to zero, Eqs. (3), (4). In this case the neutrino masses come from dimension 5 Weinberg operator after the Higgs obtains a vev. In two-component notation this is

$$\mathcal{L}_{\mathrm{eff}} \supset -\frac{c_{ij}}{\Lambda}(H^{c\dagger}L_i)(H^{c\dagger}L_j) + \mathrm{h.c.}. \tag{A.5}$$

The coefficient $c_{ij}$ form a $3 \times 3$ symmetric complex matrix, which is described by 6 real and 6 imaginary entries. In addition, there are the 9 real and 9 imaginary parameters that describe the charge lepton Yukawa matrix, $Y_\ell$, Eq. (16). The generators of unitary transformations $L_L \to V_L L_L$, $\ell_R \to V_\ell \ell_R$ are now completely broken by the Weinberg operator in conjuction with the charged lepton Yukawa couplings. This means that we have broken generators described by $2 \times (3\ \mathrm{real} + 6\ \mathrm{im.})$ parameters. From the general rule (18) it then follows that we have $9 + 6 - 2 \times 3 = 9$ real and $9 + 6 - 2 \times 6 = 3$ imaginary physical parameters. The nine real parameters are the three charged lepton masses, three neutrino masses, and the three mixing angles of the PMNS matrix. The three physical phases are the phase in the PMNS matrix, and the two Majorana phases in the Majorana mass matrix, $M_\nu = \mathrm{diag}(m_1, m_2 e^{i\phi_2}, m_3 e^{i\phi_3})$.

Other options for neutrino mass matrix are possible. There could be just one, two or more than three sterile neutrinos, $\nu_{R,i}$. For an introduction of phenomenological implications see, e.g., [111]. In the case where there are only Dirac mass terms, these break the individual lepton flavour numbers $U(1)_e \times U(1)_\mu \times U(1)_\tau$ down to a total lepton number $U(1)_L$. Majorana mass terms, such as the Weinberg operator in (A.5), break also the $U(1)_L$.

# References

[1] T. E. Browder, T. Gershon, D. Pirjol, A. Soni, and J. Zupan, Rev. Mod. Phys. **81**, 1887 (2009), 0802.3201.

[2] Y. Grossman and P. Tanedo, Just a Taste: Lectures on Flavor Physics, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Anticipating the Next Discoveries in Particle Physics (TASI 2016) Boulder, CO, USA, June 6-July 1, 2016*, 2017, 1711.03624.

[3] M. Blanke, Introduction to Flavour Physics and CP Violation, in *Proceedings, 2016 European School of High-Energy Physics (ESHEP2016): Skeikampen, Norway, June 15-28 2016*, pp. 71–100, 2017, 1704.03753.

[4] O. Gedalia and G. Perez, Flavor Physics, in *Physics of the large and the small, TASI 09, proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, Colorado, USA, 1-26 June 2009*, pp. 309–382, 2011, 1005.3106.

[5] Y. Nir, Probing new physics with flavor physics (and probing flavor physics with new physics), in *Prospects in Theoretical Physics (PiTP) summer program on The Standard Model and Beyond IAS, Princeton, NJ, June 16-27, 2007*, 2007, 0708.1872.

[6] J. F. Kamenik, Flavour Physics and CP Violation, in *Proceedings, 2014 European School of High-Energy Physics (ESHEP 2014): Garderen, The Netherlands, June 18 - July 01 2014*, pp. 79–94, 2016, 1708.00771.

[7] U. Nierste, Three Lectures on Meson Mixing and CKM phenomenology, in *Heavy quark physics. Proceedings, Helmholtz International School, HQP08, Dubna, Russia, August 11-21, 2008*, pp. 1–38, 2009, 0904.1869.

[8] G. C. Branco, L. Lavoura, and J. P. Silva, Int. Ser. Monogr. Phys. **103**, 1 (1999).

[9] A. J. Buras, (2011), 1102.5650.

[10] Belle II, W. Altmannshofer *et al.*, (2018), 1808.10567.

[11] A. Cerri *et al.*, (2018), 1812.07638.

[12] N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963), [,648(1963)].

[13] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).

[14] L.-L. Chau and W.-Y. Keung, Phys. Rev. Lett. **53**, 1802 (1984).

[15] Particle Data Group, M. Tanabashi *et al.*, Phys. Rev. **D98**, 030001 (2018).

[16] S. L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. **D2**, 1285 (1970).

[17] L. Wolfenstein, Phys. Rev. Lett. **51**, 1945 (1983).

[18] CKMfitter Group, J. Charles *et al.*, Eur. Phys. J. **C41**, 1 (2005), hep-ph/0406184, Sept 2018 update, available from http://ckmfitter.in2p3.fr.

[19] Y. H. Ahn, H.-Y. Cheng, and S. Oh, Phys. Lett. **B703**, 571 (2011), 1106.0935.

[20] C. Jarlskog, Phys. Rev. Lett. **55**, 1039 (1985).

[21] LHCb, R. Aaij *et al.*, (2018), 1808.08865.

[22] J. Charles *et al.*, Phys. Rev. **D89**, 033016 (2014), 1309.2293.

[23] M. Gronau and D. Wyler, Phys. Lett. **B265**, 172 (1991).

[24] M. Gronau and D. London, Phys. Lett. **B253**, 483 (1991).

[25] A. Giri, Y. Grossman, A. Soffer, and J. Zupan, Phys. Rev. **D68**, 054018 (2003), hep-ph/0303187.

[26] D. Atwood, I. Dunietz, and A. Soni, Phys. Rev. Lett. **78**, 3257 (1997), hep-ph/9612433.

[27] J. Brod and J. Zupan, JHEP **01**, 051 (2014), 1308.5663.

[28] J. Brod, Phys. Lett. **B743**, 56 (2015), 1412.3173.

[29] LHCb Collaboration, CERN Report No. LHCb-CONF-2017-004. CERN-LHCb-CONF-2017-004, 2017 (unpublished).

[30] I. Adachi *et al.*, Phys. Rev. Lett. **108**, 171802 (2012), 1201.4643.

[31] M. Ciuchini, private communication.

[32] UTfit, M. Bona *et al.*, JHEP **03**, 049 (2008), 0707.0636.

[33] A. J. Buras, M. Gorbahn, S. Jäger, and M. Jamin, JHEP **11**, 202 (2015), 1507.06345.

[34] T. Blum *et al.*, Phys. Rev. **D91**, 074502 (2015), 1502.00263.

[35] A. J. Buras, JHEP **04**, 071 (2016), 1601.00005.

[36] M. Bartusiak, The New York Times **September 27** (1987).

[37] B. Grinstein, M. J. Savage, and M. B. Wise, Nucl. Phys. **B319**, 271 (1989).

[38] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, Rev. Mod. Phys. **68**, 1125 (1996), hep-

ph/9512380.

[39] K. G. Chetyrkin, M. Misiak, and M. Munz, Phys. Lett. **B400**, 206 (1997), hep-ph/9612313, [Erratum: Phys. Lett.B425,414(1998)].

[40] M. Bordone, G. Isidori, and A. Pattori, Eur. Phys. J. **C76**, 440 (2016), 1605.07633.

[41] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, JHEP **06**, 092 (2016), 1510.04239.

[42] B. Capdevila, S. Descotes-Genon, J. Matias, and J. Virto, JHEP **10**, 075 (2016), 1605.03156.

[43] B. Capdevila, S. Descotes-Genon, L. Hofer, and J. Matias, JHEP **04**, 016 (2017), 1701.08672.

[44] N. Serra, R. Silva Coutinho, and D. van Dyk, Phys. Rev. **D95**, 035029 (2017), 1610.08761.

[45] A. Bharucha, D. M. Straub, and R. Zwicky, JHEP **08**, 098 (2016), 1503.05534.

[46] W. Altmannshofer, C. Niehoff, P. Stangl, and D. M. Straub, Eur. Phys. J. **C77**, 377 (2017), 1703.09189.

[47] S. Jäger and J. Martin Camalich, Phys. Rev. **D93**, 014028 (2016), 1412.3183.

[48] LHCb, R. Aaij *et al.*, Phys. Rev. Lett. **113**, 151601 (2014), 1406.6482.

[49] LHCb, R. Aaij *et al.*, JHEP **08**, 055 (2017), 1705.05802.

[50] Belle, J. T. Wei *et al.*, Phys. Rev. Lett. **103**, 171801 (2009), 0904.0770.

[51] BaBar, J. P. Lees *et al.*, Phys. Rev. **D86**, 032012 (2012), 1204.3933.

[52] G. D'Amico *et al.*, JHEP **09**, 010 (2017), 1704.05438.

[53] L.-S. Geng *et al.*, Phys. Rev. **D96**, 093006 (2017), 1704.05446.

[54] W. Altmannshofer, P. Stangl, and D. M. Straub, Phys. Rev. **D96**, 055008 (2017), 1704.05435.

[55] B. Capdevila, A. Crivellin, S. Descotes-Genon, J. Matias, and J. Virto, JHEP **01**, 093 (2018), 1704.05340.

[56] T. Hurth, F. Mahmoudi, D. Martinez Santos, and S. Neshatpour, Phys. Rev. **D96**, 095034 (2017), 1705.06274.

[57] G. Hiller and I. Nisandzic, Phys. Rev. **D96**, 035003 (2017), 1704.05444.

[58] R. Alonso, B. Grinstein, and J. Martin Camalich, Phys. Rev. Lett. **113**, 241802 (2014), 1407.7044.

[59] W. Altmannshofer and D. M. Straub, Eur. Phys. J. **C73**, 2646 (2013), 1308.1501.

[60] W. Altmannshofer, S. Gori, M. Pospelov, and I. Yavin, Phys. Rev. **D89**, 095033 (2014), 1403.1269.

[61] A. Greljo, G. Isidori, and D. Marzocca, JHEP **07**, 142 (2015), 1506.01705.

[62] G. Hiller and M. Schmaltz, JHEP **02**, 055 (2015), 1411.4773.

[63] W. Altmannshofer and D. M. Straub, Eur. Phys. J. **C75**, 382 (2015), 1411.3161.

[64] W. Altmannshofer, S. Gori, M. Pospelov, and I. Yavin, Phys. Rev. Lett. **113**, 091801 (2014), 1406.2332.

[65] J. F. Kamenik, Y. Soreq, and J. Zupan, Phys. Rev. **D97**, 035002 (2018), 1704.06005.

[66] G. Bélanger and C. Delaunay, Phys. Rev. **D94**, 075019 (2016), 1603.03333.

[67] B. Gripaios, M. Nardecchia, and S. A. Renner, JHEP **06**, 083 (2016), 1509.05020.

[68] M. Bauer and M. Neubert, Phys. Rev. Lett. **116**, 141802 (2016), 1511.01900.

[69] D. Bečirević and O. Sumensari, JHEP **08**, 104 (2017), 1704.05835.

[70] S. Fajfer, J. F. Kamenik, and I. Nisandzic, Phys. Rev. **D85**, 094025 (2012), 1203.2654.

[71] P. Asadi, M. R. Buckley, and D. Shih, JHEP **09**, 010 (2018), 1804.04135.

[72] A. Greljo, D. J. Robinson, B. Shakya, and J. Zupan, JHEP **09**, 169 (2018), 1804.04642.

[73] D. Bečirević, S. Fajfer, N. Košnik, and O. Sumensari, Phys. Rev. **D94**, 115021 (2016), 1608.08501.

[74] G. Cvetič, F. Halzen, C. S. Kim, and S. Oh, Chin. Phys. **C41**, 113102 (2017), 1702.04335.

[75] BaBar, J. P. Lees *et al.*, Phys. Rev. Lett. **109**, 101802 (2012), 1205.5442.

[76] BaBar, J. P. Lees *et al.*, Phys. Rev. **D88**, 072012 (2013), 1303.0571.

[77] Belle, M. Huschle *et al.*, Phys. Rev. **D92**, 072014 (2015), 1507.03233.

[78] Belle, Y. Sato *et al.*, Phys. Rev. **D94**, 072007 (2016), 1607.07923.

[79] LHCb, R. Aaij *et al.*, Phys. Rev. Lett. **115**, 111803 (2015), 1506.08614, [Erratum: Phys. Rev. Lett.115,no.15,159901(2015)].

[80] Belle, S. Hirose *et al.*, Phys. Rev. Lett. **118**, 211801 (2017), 1612.00529.

[81] Belle, S. Hirose *et al.*, Phys. Rev. **D97**, 012004 (2018), 1709.00129.

[82] LHCb, R. Aaij *et al.*, Phys. Rev. **D97**, 072013 (2018), 1711.02505.

[83] D. Bigi and P. Gambino, Phys. Rev. **D94**, 094008 (2016), 1606.08030.

[84] F. U. Bernlochner, Z. Ligeti, M. Papucci, and D. J. Robinson, Phys. Rev. **D95**, 115008 (2017), 1703.05330, [erratum: Phys. Rev.D97,no.5,059902(2018)].

[85] D. Bigi, P. Gambino, and S. Schacht, JHEP **11**, 061 (2017), 1707.09509.

[86] S. Jaiswal, S. Nandi, and S. K. Patra, JHEP **12**, 060 (2017), 1707.09977.

[87] R. Alonso, B. Grinstein, and J. Martin Camalich, Phys. Rev. Lett. **118**, 081802 (2017), 1611.06676.

[88] D. A. Faroughy, A. Greljo, and J. F. Kamenik, Phys. Lett. **B764**, 126 (2017), 1609.07138.

[89] A. Greljo, J. Martin Camalich, and J. D. Ruiz-Álvarez, (2018), 1811.07920.

[90] M. Freytsis, Z. Ligeti, and J. T. Ruderman, Phys. Rev. **D92**, 054018 (2015), 1506.08896.

[91] D. J. Robinson, B. Shakya, and J. Zupan, JHEP **02**, 119 (2019), 1807.04753.

[92] D. Buttazzo, A. Greljo, G. Isidori, and D. Marzocca, JHEP **11**, 044 (2017), 1706.07808.

[93] A. Crivellin, D. Müller, and T. Ota, JHEP **09**, 040 (2017), 1703.09226.

[94] M. Misiak *et al.*, Phys. Rev. Lett. **114**, 221801 (2015), 1503.01789.

[95] A. J. Buras, D. Buttazzo, J. Girrbach-Noe, and R. Knegjens, JHEP **11**, 033 (2015), 1503.02693.

[96] NA62, E. Cortina Gil *et al.*, (2018), 1811.08508.

[97] KOTO, J. K. Ahn *et al.*, Phys. Rev. Lett. **122**, 021802 (2019), 1810.09655.

[98] Y. Nir, Flavour Physics and CP Violation, in *Proceedings, 7th CERN Latin-American School of High-Energy Physics (CLASHEP2013): Arequipa, Peru, March 6-19, 2013*, pp. 123–156, 2015, 1605.00433.

[99] ATLAS, M. Aaboud *et al.*, Phys. Rev. Lett. **119**, 051802 (2017), 1705.04582.

[100] CMS, V. Khachatryan *et al.*, Phys. Lett. **B744**, 184 (2015), 1410.6679.

[101] A. L. Kagan *et al.*, Phys. Rev. Lett. **114**, 101802 (2015), 1406.1722.

[102] W. Altmannshofer, J. Brod, and M. Schmaltz, JHEP **05**, 125 (2015), 1503.04830.

[103] Physics of the HL-LHC Working Group, M. Cepeda *et al.*, (2019), 1902.00134.

[104] W. Altmannshofer, S. Gori, A. L. Kagan, L. Silvestrini, and J. Zupan, Phys. Rev. **D93**, 031301 (2016), 1507.07927.

[105] CMS, A. M. Sirunyan *et al.*, JHEP **06**, 001 (2018), 1712.07173.

[106] R. Harnik, J. Kopp, and J. Zupan, JHEP **03**, 026 (2013), 1209.1397.

[107] G. Blankenburg, J. Ellis, and G. Isidori, Phys. Lett. **B712**, 386 (2012), 1202.5704.

[108] G. W. S. Hou, Springer Tracts Mod. Phys. **233**, pp. (2019).

[109] Z. Maki, M. Nakagawa, and S. Sakata, Prog. Theor. Phys. **28**, 870 (1962), [,34(1962)].

[110] B. Pontecorvo, Sov. Phys. JETP **7**, 172 (1958), [Zh. Eksp. Teor. Fiz.34,247(1957)].

[111] M. C. Gonzalez-Garcia and Y. Nir, Rev. Mod. Phys. **75**, 345 (2003), hep-ph/0202058.

# Neutrino physics

*S. Pascoli*
Institute for Particle Physics Phenomenology, Department of Physics, Durham University, Durham, United Kingdom

**Abstract**
The discovery of neutrino oscillations just over 20 years ago has opened a new page in particle physics. It implies that neutrinos have masses and mix and, consequently, that the Standard Model of particle physics is incomplete. The key question we need to answer is: what is the origin of neutrino masses and of leptonic mixing? An impressive effort has being able to paint a precise picture of neutrino mixing. The first hints of CP violation have been reported and hunts for the nature of neutrinos are ongoing. This information guides us in extending the Standard Model to a full theory, advocating new particles and interactions. We will present a concise discussion of these issues, focusing mainly on the theoretical and phenomenological aspects and we will briefly discuss the role of neutrinos in the early Universe and in astrophysical objects.

**Keywords**
Neutrinos, neutrino oscillations, neutrino masses, leptonic mixing, see-saw mechanism, leptogenesis, lectures.

## 1 Introduction

Neutrinos are all around us but they remain the most elusive of the known fermionic particles. Their properties might hold the key to unveiling the physics beyond the Standard Model (SM) of particle physics and indeed, together with dark matter, so far they are the only evidence we have that a new theory is required.

We are not far away from the centenary of the idea itself of the neutrino, proposed in December 1930 by W. Pauli. In the 20's physicists were puzzled by the continuous spectrum of $\beta$-decays. In this process a nucleus transforms itself into another one with the emission of an electron

$$ {}^{A}_{Z}X \to {}^{A}_{Z+1}X' + e^{-}(+...) . \tag{1} $$

Only the parent and daughter nuclei and the electron could be seen. Based on these observations, according to energy-momentum conservation, the electron should carry away an energy corresponding to the difference in mass between parent and daughter nuclei. Therefore there should be a monochromatic line in the $\beta$-spectrum, in disagreement with observations. It was Pauli who suggested a possible solution to save the principle of energy-momentum conservation. In a famous open letter sent to the Gauverein meeting in Tubingen, addressed to "Dear Radioactive Ladies and Gentlemen", Pauli suggests: "I have hit upon a desperate remedy to save... the law of conservation of energy. Namely, the possibility that there could exist in the nuclei electrically neutral particles, that I wish to call neutrons, which have spin 1/2... The continuous beta spectrum would then become understandable by the assumption that in beta decay a neutron is emitted in addition to the electron such that the sum of the energies of the neutron and the electron is constant..." A copy of the letter with English translation is available at [1]. Soon after, J. Chadwick discovered a new heavy neutral particle and named it also the neutron [2]. A new name was needed for Pauli's hypothetical particle, and E. Amaldi playfully called it the "neutrino" in a conversation with E. Fermi, in contrast to his bigger synonymous, the neutrone[1]. E. Fermi adopted

---

[1]In Italian, the suffix -one indicates something big while -ino is a diminutive.

the name at the Paris Solvay Conference in 1932 and later in 1933. Moreover, Fermi, taking seriously Pauli's idea, constructed the theory of $\beta$-decay [3] in 1934, explaining it in terms of a 4-fermion interaction $n \to p e^- \bar{\nu}$ with strength $G_F$. This interaction would also predict the scattering of neutrinos off matter, via the inverse $\beta$ process $\bar{\nu} p \to n e^+$. In 1934 Bethe and Peierls were able to estimate the cross section for this process [4], finding it smaller than $10^{-44}$ cm$^2$ for a neutrino energy of 2 MeV. The discouraging implication was that "it was absolutely impossible to observe processes of this kind". It was B. Pontecorvo who suggested that indeed one could use the large neutrino fluxes becoming available [5] due to the advances in nuclear energy at the time. After unfruitfully considering the use of a nuclear explosion, F. Reines and C. L. Cowan devised a method to detect antineutrinos coming from a nuclear reactor. This technique, still in use today, exploits the simultaneous emission of a neutron and a positron in inverse beta decays to significantly reduce backgrounds. Indeed, in 1956 at the Savannah River Plant in South Carolina, they were able to detect neutrinos [6] and soon wrote a telegram to Pauli to inform him that they had "definitely detected neutrinos from fission fragments". Reines received the Nobel Prize in Physics for this discovery in 1995.

Although it was assumed for a long time that parity was an obvious symmetry of nature, in the 50's the idea that it is not conserved in weak interactions started to emerge, mainly thanks to the work of T. D. Lee and C. N. Yang [7]. Soon after, in 1956, Madame Wu and collaborators were able to prove that parity is violated in the $\beta$-decay of polarised $^{60}$Co [8] and in 1958 M. Goldhaber, L. Grodzins and A. W. Sunyar [9] showed that neutrinos are polarised in the opposite direction to their motion, i.e. they are left-handed. Landau [10], Lee and Yang [11] and Salam [12] proposed that neutrinos can be described with a left-handed Weyl spinor. This property was embedded in the $V - A$ theory of weak interactions and ultimately in the Standard Model (SM) of particle physics by S. L. Glashow [13], S. Weinberg [14] and A. Salam [15], spectacularly confirmed by the discovery of the $W$ and $Z$ bosons in 1983 and of the Higgs boson in 2012.

The idea that neutrinos and antineutrinos could be indistinguishable was due to Majorana in 1937 [16]. This question turns out to be intrinsically linked to the conservation or not of lepton number. The latter symmetry was first introduced by E. J. Konopinski and H. M. Mahmoud in 1953 to explain some missing decay modes [17]. Leptons, i.e. the electron, muon and tau and the neutrinos are given lepton number 1 and their antiparticles lepton number $-1$. The Reines and Cowan experiment preserves lepton number while searches for solar electron antineutrinos, carried out by R. Davis soon after, did not lead to any positive result. Indeed they break lepton number by two units. Davis will go on to detect electron neutrinos from the sun with the Homestake experiment.

Another important chapter in the understanding of neutrinos concerns the concept of families or generations. The muon was discovered in 1937 by J. C. Street and E. C. Stevenson [18] and by S. H. Neddermeyer and C. D. Anderson [19]. Being a heavier version of the electron, it enters Fermi interactions accompanied by a neutrino. The question was if this neutrino was the same as the one in beta decays or a different type. Following a suggestion by Pontecorvo [20], in 1962 L. M. Lederman, M. Schwartz and J. Steinberger *et al.* created the first accelerator neutrino beam, from pion decays from a boosted proton beam hitting a target, and showed that the neutrinos produced in pion decays associated with a muon do not lead to electrons in scatterings off matter [21]. Indeed, this is the proof that there are two types of neutrinos, electron and muon neutrinos and that they participate separately in weak interactions with their corresponding charged leptons. This result earned Lederman, Schwartz and Steinberger the Nobel prize in 1988. The third type of neutrinos, the one associated with the $\tau$ lepton, was finally discovered in 2000 by the DONUT experiment [22].

Once different neutrino families were established the question of whether there could be mixing and transitions between them was open. The first idea of neutrino oscillations was put forward by B. Pontecorvo in 1957 [23]. In 1962 Z. Maki, M. Nakagawa and S. Sakata introduced the concept of mixing between mass and flavour states [24]. In 1967 Pontecorvo gave a first intuitive link between two neutrino mixing and oscillations [25], subsequently completed by him with V. N. Gribov two years later [26]. In

the subsequent decade the theory was fully developed [27]. On the experimental side, first indications of a transition between flavours emerged in solar neutrino experiments. Since the 60's the Homestake experiment led by R. Davies detected solar neutrinos using a radiochemical technique with chlorine [28]. In 2002 R. Davies Jr. and M. Koshiba were awarded the Nobel prize in Physics "for pioneering contributions to astrophysics, in particular for the detection of cosmic neutrinos". Davis' experiment observed a flux smaller than predicted by J. Bahcall and collaborators [29] in the solar neutrino model. Other radiochemical experiments Gallex/GNO [30] and Sage [31] confirmed these results. These experiments can measure neutrinos at low energies with a sub-MeV threshold but cannot reconstruct the energy or the direction of the neutrinos. Water-Cherenkov detectors, starting with Kamiokande [32], could detect solar neutrinos via elastic scattering $\nu_e\,e^- \to \nu_e\,e^-$. Their threshold is much higher making them sensitive only to the $^8$B component of the flux but they can measure the energy and direction of the incoming neutrino. Super-Kamiokande was able to measure the flux deficit with great precision [33]. The solar neutrino problem remained open for a long time: whether neutrinos oscillate into flavour which cannot be detected in the experiments or the flux predictions were badly flawed. The definitive answer came in 2001 thanks to the SNO experiment [61]. It exploited two interactions in addition to elastic scattering: the charged current (CC) interaction $\nu_e + d \to p + p + e^-$ and the neutral current (NC) one $\nu_\alpha + d \to p + n + \nu_\alpha$, $\alpha = e, \mu, \tau$. The latter is particularly important as it is sensitive to all the three neutrino flavours. By comparing the $\nu_e$ and $\nu_\alpha$ fluxes deduced from the data, the SNO experiment was able to demonstrate that $\nu_e$ constitute only roughly a third of the overall solar neutrino flux at these energies and that the observed total flux is in good agreement with the theoretical predictions. The parameters required to explain the solar neutrino transitions were confirmed by the reactor neutrino experiment KamLAND soon after in 2002 [66].

Neutrino oscillations were observed also in atmospheric neutrinos. These are produced in the atmosphere when cosmic rays interact with nuclei in the atmosphere sourcing pions and kaons, and subsequently muons, which decay producing muon and electron neutrinos. Atmospheric neutrinos were first detected in 1965 deep underground at the Kolar Gold Field Mine in India [36] and soon after in the East Rand Proprietary Gold Mine in South Africa [37] looking for upgoing muon events signaling a muon neutrino interaction. First indications of a deficit of muon neutrinos were reported by Kamiokande, IMB, Soudan2, and by MACRO [38]. In 1998, the Super-Kamiokande experiment discovered neutrino oscillations [39] showing that the muon neutrino depletion is $L/E$ dependent in agreement with an oscillatory behaviour. We now know that muon neutrinos oscillate into tau neutrinos, which cannot be efficiently detected in the experiment. In 2015 T. Kajita for the Super-Kamiokande collaboration and A. B. McDonald for the SNO collaboration received the Nobel Prize in Physics for "*the discovery of neutrino oscillations, which shows that neutrinos have a mass*". As we will discuss later, this is the first particle physics evidence that the SM is incomplete, see Sec. 6. Neutrino oscillations have been studied since with great precision in solar, atmospheric, accelerator, reactor neutrino experiments and a rich programme is planned for the future, see Sec. 3.3.

## 2 Neutrinos in the Standard Model of Particle Physics and beyond

The Standard Model of particle physics [13–15] is based on the gauge symmetry $SU(3) \times SU(2)_L \times U(1)_Y$ and categorises all known fermions via the corresponding quantum numbers. They are given in Table 1.

Neutrinos are singlets of $SU(3)$ but belong to $SU(2)_L$ doublets together with their corresponding charged leptons. They have hypercharge $-1/2$ and do not carry electric charge, as $Q = T_3 + Y$. In the SM, neutrinos are Weyl fermions with left chirality, $\nu_{\alpha L} \equiv P_L \nu_\alpha$, $\alpha = e, \mu, \tau$. The chiral projectors are $P_L = (1 - \gamma_5)/2$ and $P_R = (1 + \gamma_5)/2$. For massless neutrinos, chirality and helicity match as the chiral projectors and the projectors on helicity components are the same up to corrections of order $m/E$. Left-handed neutrinos are accompanied by right-handed antineutrinos as required by the invariance of the theory under CPT (charge conjugation, parity, time reversal). Parity, the transformation of left into

**Table 1:** SM fermionic content and its irreducible representations with respect to the groups $SU(3)$, $SU(2)_L$ and $U(1)_Y$. **3** indicates a triplet of $SU(3)$, **2** a doublet of $SU(2)_L$ and **1** a singlet with respect to either group. $Y$ is the hypercharge of the fields.

| Particles | $SU(3)$ | $SU(2)_L$ | $U(1)_Y$ |
|---|---|---|---|
| Leptons | | | |
| $\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$ | **1** | **2** | $-1/2$ |
| $e_R, \quad \mu_R, \quad \tau_R$ | **1** | **1** | $-1$ |
| Quarks | | | |
| $\begin{pmatrix} u \\ d \end{pmatrix}_L, \quad \begin{pmatrix} c \\ s \end{pmatrix}_L, \quad \begin{pmatrix} t \\ b \end{pmatrix}_L$ | **3** | **2** | $1/6$ |
| $u_R, \quad c_R, \quad t_R$ | **3** | **1** | $2/3$ |
| $d_R, \quad s_R, \quad b_R$ | **3** | **1** | $-1/3$ |

right and viceversa, is maximally violated in the SM as there are no right-handed neutrinos.

Left-handed neutrinos interact via the weak force according to the charged current and neutral current terms in the SM Lagrangian:

$$\mathcal{L}_{\text{SM}} = -\frac{g}{\sqrt{2}} \sum_{\alpha=e,\mu,\tau} \overline{\nu}_{\alpha L} \gamma^\mu \ell_{\alpha L} W_\mu - \frac{g}{2\cos\theta_W} \sum_{\alpha=e,\mu,\tau} \overline{\nu}_{\alpha L} \gamma^\mu \nu_{\alpha L} Z_\mu + \text{h.c.} \,, \tag{2}$$

where $g$ is the $SU(2)_L$ coupling, $\theta_W$ is the Weinberg angle, and all other symbols have the common meaning. We notice that the structure of the SM weak interaction is of the $V - A$ type.

As discussed in the Introduction, neutrinos come in three families. A fourth active neutrino is not allowed by the invisible width of the $Z$ boson to which it would contribute as much as one active neutrino, $Z \to \nu_\alpha \bar{\nu}_\alpha$. The invisible width has been measured with great accuracy at LEP and leads to the following constraint on the active number of neutrinos [40]:

$$N_\nu = \frac{\Gamma_{inv}}{\Gamma_{\bar{\nu}\nu}} = 2.984 \pm 0.008. \tag{3}$$

Additional neutrinos could be present, as we will discuss later, but they need not partake in SM interactions, and therefore are called sterile neutrinos.

## 2.1 Leptonic mixing

Since neutrinos have masses, there are two bases that can be used to describe them: the flavour basis, $\nu_\alpha$, $\alpha = e, \mu, \tau$, depicted in Table 1, in which each neutrino is associated to the corresponding charged lepton, and the mass basis, $\nu_i$, $i = 1, 2, 3$, in which each neutrino has a definite mass. The two bases, as required by probability conservation, are related by a unitary matrix $U$, the so-called Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [23, 24]:

$$\nu_{\alpha L} = \sum_{i=1}^{3} U_{\alpha i} \, \nu_{iL} \,. \tag{4}$$

The PMNS matrix then enters the CC Lagrangian when we express it in terms of mass fields (in the basis in which the charged lepton mass matrix is diagonal):

$$\mathcal{L}_{\text{SM}} = -\frac{g}{\sqrt{2}} \sum_{\alpha, i} \overline{\nu}_i U^*_{\alpha i} \gamma^\mu P_L \ell_\alpha W_\mu + \text{h.c.} \,, \tag{5}$$

with $\alpha = e, \mu, \tau$ and $i = 1, 2, 3$[23]. From here it plays a role in neutrino oscillations as we will discuss later.

In general, a $3 \times 3$ unitary matrix can be parameterized in terms of 3 angles and 6 phases. Several of the phases are unphysical. In fact, we have the freedom to phase-rotate the fields as $\psi \rightarrow e^{i\phi}\psi$. If we do so for the charged leptons, we can eliminate three phases from the PMNS matrix and these disappear from the Lagrangian as they do not affect the kinetic terms, the NC one and the mass term for the leptons as far as both left and right-handed component undergo the same rephasing. If neutrinos are Dirac particles, as the charged leptons, the same rephasing can be applied to them as well, eliminating two further phases. There remains only one physical phase, called the Dirac phase, as in the Cabibbo-Kobayashi-Maskawa mixing matrix in the quark sector. If neutrinos are Majorana, such rephasing does not eliminate two phases which will reappear in the Majorana condition and in the Majorana mass term. Therefore, for Majorana neutrinos, there are three physical phases, two of which enter only in lepton number violating processes. Since for antineutrinos we need to use the conjugate of $U$, any physical phase represents a violation of the CP symmetry and will be called a CP violating (CPV) phase.

The PMNS matrix can be parameterized as [41,42]

$$U_{\alpha i} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \cdot \mathcal{P} , \tag{6}$$

where we define $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$, with $\theta_{ij} \in [0, 90°]$. In this notation, $\delta$ is the Dirac CPV phase $\delta \in [0, 360°]$ and $\mathcal{P}$ is a diagonal phase matrix $\mathcal{P} \equiv \mathrm{diag}(1, e^{i\frac{\alpha_{21}}{2}}, e^{i\frac{\alpha_{31}}{2}})$ which embeds the two Majorana CPV phases $\alpha_{21}, \alpha_{31}$.

It is interesting to express the CP-violating effects due to the Dirac phase in a rephasing-invariant manner. This can be done using the Jarlskog invariant [43]

$$J \equiv \Im[U_{\mu 3}U_{e2}U_{\mu 2}^*U_{e3}^*] = \frac{1}{8}\sin 2\theta_{12}\sin 2\theta_{23}\sin 2\theta_{13}\cos\theta_{13}\sin\delta . \tag{7}$$

This formulation makes apparent that Dirac CP violation is a genuine 3-neutrino mixing effect whose physical impact depends on all of the three mixing angles, including the relatively small $\theta_{13}$.

## 3 Neutrino oscillations

In presence of leptonic mixing and non-degenerate neutrino masses, the phenomenon of neutrino oscillations takes place. This is a beautiful manifestation of quantum mechanics on macroscopic distances. The basic picture is the following. In production and detection neutrinos are described by flavour states. Let's assume that a muon neutrino is produced. This is a coherent superposition of massive states which have slightly different masses. The *coherence* is a key condition which needs to be satisfied to have neutrino oscillations. It is satisfied thanks to the uncertainty in the neutrino momentum at production[4]. The massive components of the initial state propagate over long distances with slightly different phases. This amounts to a change in the state over distance. It is then possible that at detection, when projecting the flavour components out, a different flavour is found compared to the initial one. In order for the oscillatory behaviour to hold, coherence is needed also during propagation and this is possible because of the very weakly interacting nature of neutrinos[5].

---

[2]Unless otherwise indicated, we will use Greek indexes for flavour fields/states and Roman ones for mass fields/states.

[3]The flavour states are related to mass states as $|\nu_\alpha\rangle = \sum_i U_{\alpha i}^* |\nu_i\rangle$.

[4]If the momentum uncertainty is small compared to the mass differences, for instance if there exists a very heavy nearly-sterile neutrino, such coherence is lost and oscillations do not develop. At production in a specific event either the light states will be produced coherently or the heavy one.

[5]Over astronomical distances the massive components of neutrinos can separate in the wave function due to the slightly different velocities, effectively destroying coherence.

### 3.1 Oscillation probability in vacuum

The oscillation probability can be derived in different ways but here we will limit the discussion to the commonly used plane-wave approximation. This approximation does not capture the momentum uncertainty necessary for coherence. We will assume by hand that the initial state is a coherent superposition of massive states with a definite spatial momentum $p \equiv |\mathbf{p}|$.

Let's consider a $\nu_\alpha$ produced at $t = 0$ in a charged current interaction. We describe the initial state as a superposition of mass eigenstates, which we take as plane waves with momentum $p$,

$$|\nu, t = 0\rangle = |\nu_\alpha\rangle = \sum_i U_{\alpha i}^* |\nu_i\rangle \ . \tag{8}$$

The mass states $|\nu_i\rangle$ are eigenstates of the free Hamiltonian $\hat{H}$ with eigenvalues $E_i = \sqrt{\mathbf{p}^2 + m_i^2}$. The evolution of the neutrino state can be obtained by solving the evolution equation and is expressed as[6]

$$|\nu, t\rangle = \exp(-i\hat{H}t)|\nu_a\rangle = \sum_i U_{\alpha i}^* \exp(-iE_i t)|\nu_i\rangle \ . \tag{9}$$

The probability of transition from $\nu_\alpha$ to $\nu_\beta$ at time $t$ is obtained projecting the state $|\nu, t\rangle$ in the $\nu_\beta$ direction as

$$P(\nu_\alpha \to \nu_\beta, t) = |\langle \nu_\beta |\nu, t\rangle|^2 = \left| \sum_i U_{\beta i} U_{\alpha i}^* \exp(-iE_i t) \right|^2 \ , \tag{10}$$

where we have used the fact that $\langle \nu_j |\nu_i\rangle = \delta_{ij}$.

In all experimentally relevant situations, neutrinos are highly relativistic and one can approximate, for common momentum $p$,

$$E_i - E_j \simeq \frac{m_i^2 - m_j^2}{2p} \ , \tag{11}$$

and moreover one can take $L = t$.

Finally, one obtains the general formula for neutrino oscillations in vacuum

$$P(\nu_\alpha \to \nu_\beta, t) = |\langle \nu_\beta |\nu, t\rangle|^2 = \left| \sum_i U_{\beta i} U_{\alpha i}^* \exp\left(-i\frac{\Delta m_{i1}^2 t}{2E}\right) \right|^2 \ , \tag{12}$$

where we have defined $\Delta m_{i1}^2 \equiv m_i^2 - m_1^2$ and we have approximated $E \simeq p$. It is apparent from this formula that oscillations between one flavour and another are possible only if there is leptonic mixing, $U \neq 1$, and neutrinos have masses. This is the reason why the discovery of neutrino oscillations in 1998 has had such a groundbreaking effect in our understanding of neutrinos and more broadly of particle physics.

We notice that neutrino oscillations conserve lepton number, i.e. if a neutrino is produced, the state will continue being a neutrino, but does not respect leptonic flavour as the neutrino can change from one to the other over distances. We furthermore notice that Majorana phases do not enter in the oscillation formula as expected since this is a lepton number conserving process. Moreover, the overall mass scale does not play a role in it.

The case $\alpha = \beta$ is usually referred to as a survival probability or disappearance channel, the opposite one $\alpha \neq \beta$ is the transition probability or appearance channel. Conservation of probability is satisfied as $\sum_\beta P(\nu_\alpha \to \nu_\beta, t) = 1$. For antineutrinos, one substitutes $U$ with its complex conjugate $U^*$.

---

[6]We use natural units throughout: $c = 1, \hbar = 1$.

It is sometimes useful to separate the real and imaginary parts of the leptonic mixing terms as

$$P(\nu_\alpha \to \nu_\beta) = \delta_{\alpha\beta} - 4\sum_{i>j} \Re[U^*_{\alpha i} U_{\alpha j} U^*_{\beta j} U_{\beta i}] \sin^2\left(\frac{\Delta m^2_{ij} L}{4E}\right) + 2\sum_{i>j} \Im[U^*_{\alpha i} U_{\alpha j} U^*_{\beta j} U_{\beta i}] \sin\left(\frac{\Delta m^2_{ij} L}{2E}\right) .$$

(13)

The plane-wave derivation we have discussed above has the advantage of simplicity but cannot account for the momentum uncertainty necessary for coherence and the spatial size of the neutrino wave function resulting from the fact that production and detection are localised processes. A more precise treatment of this problem has been achieved using wave packets so that the initial state is the superposition of the wave packets which describe each massive neutrino [44]. This derivation allows to incorporate decoherence and momentum uncertainty effects. The approximation $L = t$ is also problematic in the plane-wave description as plane waves extend all over the space with the same amplitude. The wave packets can describe localised particles and solve this apparent paradox as well. In all experimentally relevant cases, it will result in the same formula as Eq. (12) as far as coherence is maintained and the momentum uncertainty is sufficiently large. For further details about this derivation and a discussion of neutrino oscillations in the context of QFT, see e.g. [45].

### 3.1.1   2-neutrino oscillations

We now study the probability in Eq. (12) more in detail [26, 27]. Let's consider first the oscillation probability in the two neutrino case. The massive basis $\nu_1$, $\nu_2$ is related to the flavour basis $\nu_\alpha$, $\nu_\beta$ as

$$\begin{pmatrix} \nu_\alpha \\ \nu_\beta \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} ,$$

(14)

where $\theta$ is the mixing angle in vacuum which parameterizes the $2 \times 2$ mixing matrix.

The oscillation appearance probability is given by

$$P(\nu_\alpha \to \nu_\beta) = \sin^2 2\theta \sin^2\left(\frac{\Delta m^2 L}{4E}\right) .$$

(15)

We schematically show $P(\nu_\mu \to \nu_\tau)$ in Fig. 1.

As expected, we see that this probability is different from zero only in presence of mixing and of neutrino masses. Oscillations do not develop if the distance travelled by the neutrinos is too short and they reach a maximum when $\frac{\Delta m^2 L}{4E} = \pi/2$. For a given baseline, set by the distance between the source and the detector, the energy of the first oscillation maximum is controlled by $\Delta m^2$. It is useful to express the argument in terms of experimentally relevant units as

$$\frac{\Delta m^2 L}{4E} = 1.27 \frac{\Delta m^2}{\text{eV}^2} \frac{L}{\text{km}} \frac{\text{GeV}}{E} .$$

(16)

The disappearance probability $P(\nu_\alpha \to \nu_\alpha)$ is simply $1 - P(\nu_\alpha \to \nu_\beta)$. We also notice that CPT invariance guarantees that $P(\nu_\alpha \to \nu_\beta) = P(\overline{\nu}_\beta \to \overline{\nu}_\alpha)$ and that the disappearance probability is the same for neutrinos and antineutrinos $P(\nu_\alpha \to \nu_\alpha) = P(\overline{\nu}_\alpha \to \overline{\nu}_\alpha)$. Moreover, we have that the probability is invariant under a T and CP transformations, since $P(\nu_\alpha \to \nu_\beta) = P(\nu_\beta \to \nu_\alpha) = P(\overline{\nu}_\alpha \to \overline{\nu}_\beta)$. This implies that 2-neutrino oscillations in vacuum are not sensitive to leptonic CP violation and therefore to hunt for the Dirac CPV phase it is necessary to exploit setups for which 3-neutrino oscillation effects are relevant.

### 3.1.2   3-neutrino oscillations

In the case of 3-neutrino mixing, the probability of $\nu_\alpha \to \nu_\beta$ oscillations in vacuum has a more complex form in terms of the two mass squared differences $\Delta m^2_{21}$ and $\Delta m^2_{31}$ and of the mixing parameters. In experimental situations, two limits are particularly interesting.
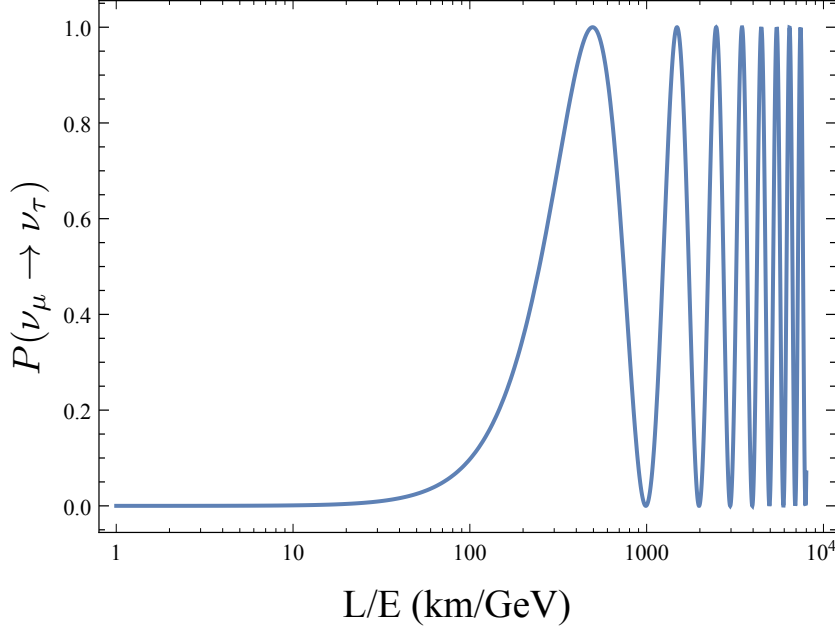
**Fig. 1:** The transition probability $P(\nu_\mu \to \nu_\tau)$ in the 2-neutrino mixing approximation as a function of $L/E$ for $\sin^2 2\theta = 1$ and $\Delta m^2 = 2.5 \times 10^{-3}$ eV$^2$.

– Case A: $\frac{\Delta m^2_{21} L}{2E} \ll 1$. This is the case relevant for accelerator, atmospheric, and medium baseline reactor neutrino experiments as far as subdominant 3-neutrino mixing effects can be neglected. The oscillations due to $\Delta m^2_{21}$ do not develop and the probability reduces to:

$$P(\nu_\alpha \to \nu_\beta) \simeq 4|U_{\alpha 3}|^2 |U_{\beta 3}|^2 \sin^2 \left( \frac{\Delta m^2_{31} L}{4E} \right) . \tag{17}$$

This formula resembles the one for the 2-neutrino oscillation case and indeed has the same properties, in that it is not sensitive to CP-violating effects.

Accelerator neutrino experiments such as T2K and NOvA, and in the past MINOS, K2K, and atmospheric neutrino experiments, specifically Super-Kamiokande, exploit a muon neutrino beam, mainly from pion decays and can measure quite precisely its disappearance probability given by

$$P(\nu_\mu \to \nu_\mu) \simeq 1 - 4|U_{\mu 3}|^2 (1 - |U_{\mu 3}|^2) \sin^2 \left( \frac{\Delta m^2_{31} L}{4E} \right) \simeq 1 - \sin^2 2\theta_{23} \sin^2 \left( \frac{\Delta m^2_{31} L}{4E} \right) + \mathcal{O}(s^2_{13}) . \tag{18}$$

Consequently, they provide the dominant information on $\Delta m^2_{31}$ and on the $\theta_{23}$ angle. The current generation experiments T2K and NOvA are also designed to detect electron neutrinos from the $\nu_\mu \to \nu_e$ oscillations at long distance. The probability at leading order is given by

$$P(\nu_\mu \to \nu_e) \simeq 4|U_{e3}|^2 |U_{\mu 3}|^2 \sin^2 \left( \frac{\Delta m^2_{31} L}{4E} \right) \simeq s^2_{23} \sin^2 2\theta_{13} \sin^2 \left( \frac{\Delta m^2_{31} L}{4E} \right) . \tag{19}$$

We note that this probability is suppressed by the small mixing angle $\theta_{13}$. Subdominant terms arise due to matter effects, see Sec. 3.2, and Dirac CP violation. This is the channel of choice to determine the mass ordering and discover CP violation in long baseline neutrino oscillation experiments.

Finally, the probability which is relevant for medium baselines $L \sim 1$ km in reactor neutrino experiments is

$$P(\overline{\nu}_e \to \overline{\nu}_e) \simeq 1 - \sin^2 2\theta_{13} \sin^2 \left( \frac{\Delta m^2_{31} L}{4E} \right) . \tag{20}$$

This probability is controlled again by $\theta_{13}$ and, thanks to this, the Daya Bay experiment [46], as well as RENO [47] and Double CHOOZ [48], discovered $\theta_{13} \neq 0$ in 2012.

– Case B: $\frac{\Delta m_{31}^2 L}{2E} \gg 1$. Long baseline reactor neutrino experiments such as KamLAND exploit this approximation. In this case, the oscillations controlled by $\Delta m_{31}^2$ are effectively averaged out. The probability can be well approximated by

$$P(\overline{\nu}_e \to \overline{\nu}_e) \simeq c_{13}^2 \left[ 1 - \sin^2 2\theta_{12} \sin^2 \left( \frac{\Delta m_{21}^2 L}{4E} \right) \right] + s_{13}^2 \ . \tag{21}$$

It follows that KamLAND can measure very precisely the value of $\Delta m_{21}^2$ and is sensitive to $\theta_{12}$ and at some level also to $\theta_{13}$ [49].

If 3-neutrino mixing effects are at play, the neutrino probability becomes sensitive to Dirac CP violation. This can be seen computing the CP asymmetry $\mathcal{A}(\nu_\alpha \to \nu_\beta) \equiv P(\nu_\alpha \to \nu_\beta) - P(\overline{\nu}_\alpha \to \overline{\nu}_\beta)$. Using Eq. (13) it follows that

$$\mathcal{A}(\nu_\alpha \to \nu_\beta) = 4 s_{12} c_{12} s_{13} c_{13}^2 s_{23} c_{23} \sin \delta \left[ \sin \left( \frac{\Delta m_{21}^2 L}{4E} \right) + \sin \left( \frac{\Delta m_{13}^2 L}{4E} \right) + \sin \left( \frac{\Delta m_{32}^2 L}{4E} \right) \right] \ . \tag{22}$$

We notice that CPV effects can be parameterised in terms of the Jarlskog invariant and depend on the Dirac CPV phase. Moreover, they are different from zero only in presence of 3-neutrino oscillation effects, i.e. if $\Delta m_{21}^2$ can be neglected $\mathcal{A}(\nu_\alpha \to \nu_\beta)$ goes to zero. This implies that CPV effects are suppressed by the small mass squared difference $\Delta m_{21}^2$ and are controlled by the small mixing angle $\theta_{13}$ making their search challenging. Current long baseline neutrino oscillation experiments have started being sensitive to these effects which are the main focus of next generation experiments.

## 3.2 Matter effects in neutrino oscillations

Neutrinos are affected by the medium in which they travel. They can incoherently scatter off its components, e.g. electrons, neutrons, protons, but typically these interactions can be neglected. Using dimensional arguments, a crude estimate of the interaction cross section for a neutrino of energy $E$ is

$$\sigma_\nu \sim G_F^2 E M \sim 10^{-38} \ \text{cm}^2 \frac{EM}{\text{GeV}^2} \ ,$$

where we have used a mass $M$ for the target in the medium, typically being nucleons. Considering for instance the Earth density this cross section leads to a mean free path of $10^{14}$ cm at 1 GeV, well above the Earth diameter. Indeed the Earth becomes opaque to neutrinos only at energies above $10^2$ TeV. Only in extremely dense environments, such as supernovae cores and the Early Universe, these interactions are sufficiently frequent to trap the neutrinos.

In the low density situations of interest, e.g. the Earth and the Sun, the medium affects neutrinos nevertheless by modifying their effective masses. In matter, neutrinos interact with the background particles, e.g. electrons, protons and neutrons, via forward elastic scattering [50]. Let's consider a neutral unpolarised medium at rest. We consider centre-of-mass energies well below the $W$ and $Z$ masses, for which we can use the Fermi approximation. The effective Hamiltonian density for the CC interaction is given by

$$\text{H}_{CC} = 2\sqrt{2} G_F [\overline{e} \gamma_\mu P_L \nu_e][\overline{\nu}_e \gamma^\mu P_L e] = 2\sqrt{2} G_F [\overline{\nu}_e \gamma^\mu P_L \nu_e][\overline{e} \gamma_\mu P_L e] \ , \tag{23}$$

where we have used a Fierz transformation to separate the neutrino part from the background electron one. Averaging the electron component over the background at rest gives

$$\langle \overline{e} \gamma_\mu P_L e \rangle = \delta_{\mu 0} \frac{N_e}{2} \ , \tag{24}$$
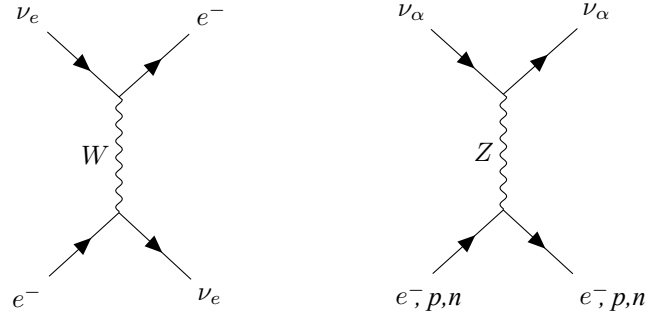
**Fig. 2:** Feynman diagrams for the CC and NC neutrino interactions with a medium, such as the Earth or the Sun.

where $N_e$ is the electron density. Similarly, one can compute the contribution due to the NC interactions. In this case, in a neutral background, the electron and proton contributions cancel out and only the neutron density $N_n$ is relevant.

By considering the modified dispersion relations in matter, one can see that an effective potential is induced in the Hamiltonian:

$$V_e = \sqrt{2}G_F\left(N_e - \tfrac{N_n}{2}\right) , \tag{25}$$

$$V_{\mu,\tau} = \sqrt{2}G_F\left(-\tfrac{N_n}{2}\right) . \tag{26}$$

For antineutrinos the potential changes sign. This indicates a violation of CP and CPT symmetries, which is due to the fact that the background is itself CP and CPT violating as it contains only particles and not antiparticles[7].

Notice that these terms are diagonal in the flavour basis as there are no SM processes which change one flavour into another and that the NC terms are the same for all three flavours as NC interactions are flavour blind in the SM.

The effective Hamiltonian, describing the neutrino propagation in the medium, is given by the vacuum one $H^0$ augmented by the potential terms as

$$\mathrm{H}^m = \mathrm{H}^0 + \mathrm{diag}(V_e, V_\mu, V_\tau) , \tag{27}$$

in the flavour basis.

Let's consider the simplest case of 2-neutrino oscillations and choose the $\nu_e$–$\nu_\mu$ flavours[8]. In the flavour basis the neutrino propagation can be described by

$$i\frac{d}{dt}\begin{pmatrix}\nu_e \\ \nu_\mu\end{pmatrix} = \begin{pmatrix}-\frac{\Delta m^2}{4E}\cos 2\theta_o + \sqrt{2}G_F N_e(t) & \frac{\Delta m^2}{4E}\sin 2\theta_o \\ \frac{\Delta m^2}{4E}\sin 2\theta_o & \frac{\Delta m^2}{4E}\cos 2\theta_o\end{pmatrix}\begin{pmatrix}\nu_e \\ \nu_\mu\end{pmatrix} . \tag{28}$$

For clarity we have indicated the mixing angle in vacuum as $\theta_o$ and it corresponds to $\theta$ in Eq. 14. To derive this expression we have eliminated any common term in the diagonal as only relative phases between the states are relevant in the probabilities. Computing the resulting evolution can be highly non trivial, especially in the full 3-neutrino mixing picture, and one may have to resort to numerical tools. In some case, analytical approximations can be applied. We consider two particularly relevant ones: the constant density case and the case of varying density with adiabaticity.

---

[7]This is not the case in the Early Universe in which both types are typically present with a very similar density.

[8]As the matter potential is the same for $\nu_\mu$ and $\nu_\tau$, we do not expect matter effects to arise in the oscillations between these two flavours, at least at leading order.

### 3.2.1 Constant density case

For constant density, the evolution of the two eigenstates in matter can be decoupled. The mixing between the flavour states and the eigenstates in matter is

$$\tan 2\theta_m = \frac{\frac{\Delta m^2}{2E} \sin 2\theta_o}{\frac{\Delta m^2}{2E} \cos 2\theta_o - \sqrt{2} G_F N_e} \ . \tag{29}$$

The probability of oscillation can be computed in analogy to the vacuum case and is given by

$$P(\nu_e \rightarrow \nu_\mu; t) = \sin^2 2\theta_m \sin^2 \left( \frac{L}{2} (E_A - E_B) \right) \ , \tag{30}$$

with

$$|E_A - E_B| = \sqrt{\left( \frac{\Delta m^2}{2E} \cos 2\theta_o - \sqrt{2} G_F N_e \right)^2 + \left( \frac{\Delta m^2}{2E} \sin 2\theta_o \right)^2} \ . \tag{31}$$

We notice that the behaviour in matter can be significantly different to the vacuum case [51].

– *Vacuum limit.* If $\sqrt{2} G_F N_e \ll \frac{\Delta m^2}{2E} \cos 2\theta_o$, $\theta_m \simeq \theta_o$ and the vacuum solution is recovered.

– *Matter domination.* If $\sqrt{2} G_F N_e \gg \frac{\Delta m^2}{2E} \cos 2\theta_o$, matter effects dominate and the transition probability is very suppressed. This can be understood as matter effects are flavour diagonal and therefore tend to realign the evolving state onto the initial flavour direction, suppressing flavour transitions.

– *Resonance.* The remaining option is particularly interesting and happens when $\sqrt{2} G_F N_e = \frac{\Delta m^2}{2E} \cos 2\theta_o$. In this case the mixing angle in matter is maximal, $\theta_m = \pi/4$, independently from the value of $\theta_o$. This case is called "resonance" and can happen for neutrinos (antineutrinos) if $\Delta m^2 > 0$ ($\Delta m^2 < 0$), for $\cos 2\theta_o > 0$. Once the resonant condition is satisfied, the oscillation length is controlled by $|E_A - E_B| = \frac{|\Delta m^2|}{2E} \sin 2\theta_o$ requiring very long distances for the oscillations to develop if $\theta_o$ is very small.

The case of constant density is relevant for long baseline neutrino oscillations experiments, e.g. NOvA, DUNE. By searching for an enhancement of the oscillation probability in neutrinos or antineutrinos due to matter effects, these experiments are sensitive to the sign of $\Delta m_{31}^2$, or the mass ordering. These are typically small effects and are partly degenerate with intrinsic CPV effects due to the Dirac phase, which are also opposite for neutrinos and antineutrinos. Focusing on the $\nu_\mu \rightarrow \nu_e$ transition, an approximate form of the oscillation probability allows to study the dependence on the various effects [52]

$$P(\nu_\mu \rightarrow \nu_e) \simeq s_{23}^2 \sin^2 2\theta_{13} \left( \frac{\Delta_{31}}{\Delta_{31} \mp A} \right)^2 \sin^2 \frac{(\Delta_{31} \mp A)L}{2}$$

$$+ c_{13} \sin 2\theta_{13} \sin 2\theta_{12} \sin \theta_{23} \frac{\Delta_{21}}{A} \sin \frac{AL}{2} \frac{\Delta_{31}}{|\Delta_{31} \mp A|} \sin \frac{|\Delta_{31} \pm A|L}{2} \cos \left( \frac{\Delta_{31}L}{2} \mp \delta \right) + \mathcal{O}(\Delta_{12}^2 L), \tag{32}$$

with $\Delta_{ij} \equiv \Delta m_{ij}^2 / 2E_\nu$ and $A \equiv V_e - V_\mu = \sqrt{2} G_F N_e$. Due to the fact that $\theta_{13}$ is not too small, see Sec. 3.4, the first term dominates and provides sensitivity to matter effects and the mass ordering. The second term in this expression depends on the CPV phase $\delta$. As expected, it arises from the interference of the oscillations due to both $\Delta m_{31}^2$ and $\Delta m_{21}^2$ and increases at lower energies. From this expression, we can understand that determining the mass ordering and CPV in long baseline neutrino oscillation experiments is possible but presents some challenges. Having information at different energies and both for neutrinos and antineutrinos alleviates the degeneracy between the various parameters and enhances the physics reach.

### 3.2.2 *Varying density*

In many situations, the neutrinos travel through a medium of varying density. This is the case for neutrinos produced in the inner parts of the Sun or for atmospheric neutrinos going through the Earth.

At any given time, it is possible to diagonalise the Hamiltonian and find the corresponding instantaneous propagation states, $\nu_A$, $\nu_B$. The mixing angle in matter is time dependent as its expression depends on the local density. As a result, in this basis, the Hamiltonian acquires off-diagonal terms which depend on the time derivative of the potential as

$$i\frac{d}{dt}\begin{pmatrix} \nu_A \\ \nu_B \end{pmatrix} = \begin{pmatrix} E_A & -i\dot{\theta}_m(t) \\ i\dot{\theta}_m(t) & E_B \end{pmatrix}\begin{pmatrix} \nu_A \\ \nu_B \end{pmatrix}. \tag{33}$$

For constant density, the off-diagonal terms are zero and the two states $\nu_A$ and $\nu_B$ evolve independently. For varying density, the off-diagonal terms indicate the possibility of a transition from one state to the other. An analytical solution is typically very difficult to obtain and numerical tools need to be employed to compute the transition probabilities. If the off-diagonal terms are subdominant, as it happens for a slowly varying density, then some approximate solution can be found. This is the adiabatic case for which the adiabatic condition is satisfied

$$|E_A - E_B| \gg \left|\frac{d}{dt}\theta_m\right|. \tag{34}$$

The case of varying density is realised in the Sun, in which the neutrinos see a slowly varying density as they travel towards the surface. Let's consider the 2-neutrino mixing approximation. At any time $t$, it is possible to relate the flavour states to the instantaneous propagation states. Explicitly we have

$$\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_m(t) & \sin\theta_m(t) \\ -\sin\theta_m(t) & \cos\theta_m(t) \end{pmatrix}\begin{pmatrix} \nu_A \\ \nu_B \end{pmatrix}. \tag{35}$$

As neutrinos originate from close to the centre of the Sun, at sufficiently high energies, matter effects dominate and $\theta_m \sim \pi/2$ implying that electron neutrinos are mainly in the heavy state $\nu_B$. This can be seen from the form of the Hamiltonian in Eq. (28) which is nearly diagonal with the dominant term in the $ee$ position. As neutrinos reach the surface, they have remained in the same state $\nu_B$ as far as the adiabaticity condition is satisfied. At this position, the vacuum case applies so that $\theta_m = \theta_o$ and

$$|\nu_B\rangle = \sin\theta_o|\nu_e\rangle + \cos\theta_o|\nu_\mu\rangle, \tag{36}$$

implying that the survival probability is

$$P(\nu_e \to \nu_e, \text{surface}) = \sin^2\theta_o. \tag{37}$$

If the mixing angle in vacuum is small, this corresponds to a nearly total transition to $\nu_\mu$. This is the so-called MSW effect [50, 51] and explains neutrino transitions in the Sun for energies of few MeV. It should be noted that it is improper to speak about oscillations for these transitions. In fact the survival probability does not result from the coherent evolution of the mass eigenstates produced at the source but by the independent evolution of the propagation states, so that, at these energies, most solar neutrinos we observe on the Earth are mass eigenstates $\nu_2$.

At low energies, matter effects are always negligible and vacuum oscillations take place, averaged over the long distances. This produces a very typical transition behaviour, with the probability at low energies given by $P(\nu_e \to \nu_e, \text{surface}) = 1 - 1/2\sin^2 2\theta_o$, a transition region around the few MeV and $P(\nu_e \to \nu_e) = \sin^2\theta_o$ at high energies. The dependence on neutrino masses arises from the energy at which the resonant condition is satisfied. In Fig. 3 we schematically show the behaviour of the solar neutrino survival probability.
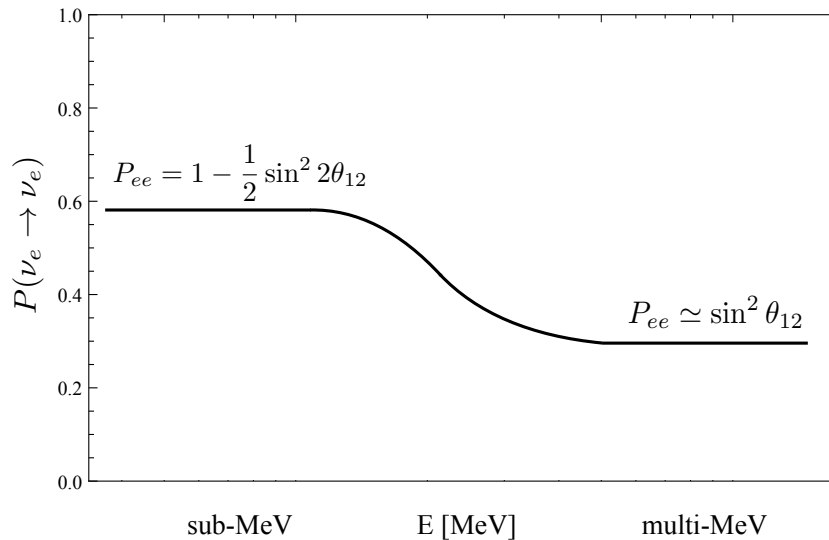
**Fig. 3:** A schematic representation of the electron neutrino survival probability $P(\nu_e \to \nu_e)$ for solar neutrinos, as a function of the energy.

## 3.3 Experimental knowledge on neutrino oscillations

Since the discovery of neutrino oscillations, a rather precise picture of neutrino oscillation properties has been painted by a very rich experimental programme. Neutrino oscillations have been observed in atmospheric, accelerator, solar, reactor neutrino experiments. Here, we provide a very concise summary, referring the reader to more updated and broad discussions available in the literature and in conferences.

### 3.3.1 Atmospheric neutrinos

Neutrinos are produced in the atmosphere by pion and kaon decays, and subsequent muon decays, produced by cosmic rays hitting the atmosphere. The flux is mainly made of muon neutrinos and electron neutrinos with a ratio of two since there are two muon neutrinos, one coming from pion decay and one from muon decay, per electron neutrino[9]. The spectrum is very broad going from sub-GeV to multi-TeV energies. For neutrino oscillation purposes the range of interest spans from hundrends of MeV to a few GeV.

Since the discovery of neutrino oscillations in atmospheric neutrinos by the Super-Kamiokande experiment, several experiments have studied these oscillations in greater detail. Super-Kamiokande 1-4 has collected more data [53], MINOS [54] has been able to distinguish neutrinos from antineutrinos, thanks to its magnetisation, and IceCube/DeepCore have also provided relevant information [55].

Atmospheric neutrinos contribute to our current knowledge of neutrino parameters mainly by observing the muon neutrino disappearance channel. This gives information on $\Delta m_{31}^2$ and the angle $\theta_{23}$, see Eq. (18). Thanks to the strong matter effects, some information can also be obtained on the mass ordering, although the lack of magnetisation of the Super-Kamiokande detector and the limited number of events do not allow to reach a high statistical significance.

### 3.3.2 Accelerator neutrinos

Accelerator neutrinos are produced in the similar manner as atmospheric neutrinos, by focusing a pion beam down a decay pipe. This allows to have a controlled beam, in which the electron neutrino compo-

---

[9]At high energy this ratio becomes much bigger as muons hit the Earth before decaying, so that the electron component is suppressed.

nent is suppressed, the spectrum can be predicted with good accuracy, the intensity is enhanced and the average energy tuned to match the first oscillation maximum.

After K2K, several experiments took place including MINOS [54, 56], and the currently running T2K [57] and NOvA [58]. MINOS [54, 56] used the NuMI (Neutrinos at Main Injector) beam sourced at Fermilab and two iron magnetised detectors made of alternating planes of steel and plastic scintillators. The near detector, with a mass of 980 tons, was close to the beam target. The far detector was located in the Soudan mine in Northern Minnesota, 735 km from Fermilab, and had a mass of 5.4 ktons. Being made of magnetised steel these detectors had very good muon reconstruction capabilities and could distinguish neutrinos from antineutrino events. T2K (Tokai-to-Kamioka) experiment [57] exploits a beam produced at the J-PARC facility and the Super-Kamiokande detector located 295 km away. Due to its position, the beam reaching the detector is off-axis, resulting in a beam peaked at lower energies compared with the on-axis one. This is useful as it reduces the backgrounds due to the high energy tail of the neutrinos and increases the number of events at the first oscillation maximum. The Water-Cherenkov detector allows to reconstruct both muon and electron neutrino events. The NOvA experiment [58] also aims at detecting electron neutrinos, as well as muon ones, from the NuMI beam. The 14 kton detector is at Ash River in Minnesota 810 km from Fermilab and is made of cells of plastic PVC filled with liquid scintillator. Also in this experiment, the location is off-axis.

All these experiments provide key information on $\Delta m^2_{31}$ and the angle $\theta_{23}$, thanks to their ability to measure the muon neutrino survival probability. T2K and NOvA are also aimed at detecting the $\nu_\mu \to \nu_e$ transition channel whose probability depends on $\theta_{13}$ and subdominantly on the CPV Dirac phase $\delta$. NOvA, thanks to its rather long baseline of $\sim 810$ km, has also some sensitivity to the mass ordering via matter effects.

We should also mention the OPERA experiment, which detected $\nu_\tau$ from oscillations of the CNGS beam sourced at CERN [59]. It was located at the Gran Sasso Laboratories 735 km away. The detector was made of lead bricks and nuclear emulsions, to search for the characteristic tau tracks. It was able to observe 5 tau neutrino events, confirming the hypothesis of $\nu_\mu \to \nu_\tau$ oscillations.

### 3.3.3 Solar neutrinos

Solar electron neutrinos are produced in the nuclear reactions that burn hydrogen into helium in the Sun, and all other stars:

$$4p \to {}^4\text{He} + 2e^+ + 2\,\nu_e \,. \tag{38}$$

A multicomponent flux is generated: $pp$ neutrinos dominate but have rather low energies, $E_\nu \lesssim 0.4\,\text{MeV}$, $^8\text{B}$ neutrinos have much higher energies reaching above 10 MeV although with a much lower flux, $^7\text{Be}$ and $pep$ neutrinos have monochromatic lines at intermediate energies.

In recent years solar neutrinos have been studied mainly with the Super-Kamiokande detector [60], via the elastic scattering process $\nu_e e \to \nu_e e$, with SNO [61], sensitive to both electron neutrino and the overall flux as discussed earlier, and more recently with Borexino [62], which thanks to its low threshold can detect $^7\text{Be}$ and $pep$ neutrinos. These experiments provide information on the mixing angle $\theta_{12}$ and, by reconstructing the transition probability above and below the resonance, on $\Delta m^2_{21}$. As discussed in Sec. 3.2, the resonance condition depends on $\Delta m^2_{21}$ whose value can be then extracted with some level of precision. By observing the resonant behaviour for neutrinos, we can also deduce that $\Delta m^2_{21}$ is positive, for $\cos 2\theta_{12} > 0$, establishing the hierarchy between $m_1$ and $m_2$.

### 3.3.4 Reactor neutrinos

After the Reines and Cowan experiment, many other reactor neutrino experiments have taken place using reactor electron anti-neutrinos. Depending on the distance, we can classify them as short[10], for

---

[10]We will refer to reactor neutrino experiments searching for sterile neutrinos using baselines of tens of meters as "very-short" baseline experiments.

$L \sim 1$ km, intermediate, for $L \sim 50$–$60$ km, or long baseline, for $L > 100$ km, ones. Short-baseline experiments have a detector located typically around 1 km from the nuclear core. This is the case for currently running Daya Bay [46,63], RENO [47,64] and Double CHOOZ [48,65]. By exploiting inverse beta decays, they search for electron antineutrino disappearance. As shown in Eq. (20), this depends on $\theta_{13}$ which has been found to be non-zero in 2012 by these experiments, after previous hints and some clear indications by T2K.

The long baseline KamLAND experiment [66], a 1 kton liquid scintillator detector, observed neutrinos coming from all nuclear reactors in Japan, with an average distance of 175 km. At these baselines, the oscillation expression in Eq. (21) applies showing that this experiment provides the most precise information on $\Delta m_{21}^2$ as well as a measurement of $\theta_{12}$ [49,66]. Remarkably, the solar neutrino oscillations and KamLAND results pointed to the same values of these parameters.

### 3.3.5 *Short baseline neutrino oscillations*

Although we focus on the 3-neutrino mixing scenario, we mention here that neutrino oscillation experiments at short baselines have reported some hints which can be interpreted as being due to light sterile neutrinos. In the 90s the LSND experiment found evidence of $\bar{\nu}_\mu \to \bar{\nu}_e$ transitions using muon antineutrinos from pion decays [67]. This result could be explained with neutrino oscillations at very short baseline induced by a large mass squared difference $\Delta m^2 \sim 1$ eV$^2$. In order to accommodate such large value together with the measured $\Delta m_{31}^2 \simeq 2.5 \times 10^{-3}$ eV$^2$ and $\Delta m_{21}^2 \simeq 8 \times 10^{-5}$ eV$^2$, it is necessary to introduce 4 massive neutrinos. The fourth flavour state needs not to have SM interactions as implied by the $Z$ invisible width, see Eq. (3), hence the name of sterile. The MiniBooNE experiment, designed to test this result, observed some anomaly as well, namely an excess of electron neutrino events at low energies [68]. The LSND and MiniBooNE results could be interpreted in terms of neutrino oscillations, see also Ref. [70][11]. These results are somewhat in tension with disappearance experiments, driven mainly by MINOS+ and IceCube, which put very stringent constraints on the mixing between the muon and the fourth massive neutrinos [70], disfavouring this explanation. Regarding the mixing with electron neutrinos, there are some additional indications in favour of sterile neutrinos. Very short baseline reactor neutrino experiments, with $L \sim$ few m, have measured a flux which is lower than predicted by $\sim 3\%$ [71]. Although there are significant uncertainties on the reactor neutrino flux computations, these results could be regarded in favour of sterile neutrino oscillations with $|U_{e4}|^2 \sim 0.01$. The Gallium anomaly refers to a deficit of measured electron neutrinos from radioactive sources at the GALLEX and SAGE solar neutrino experiments [72], calling for a somewhat larger value of the mixing angle $|U_{e4}|^2 \sim 0.1$.

A coherent picture is still missing and several experiments are taking data. The SBN (Short-Baseline Neutrino) program [73] at Fermilab exploits the Booster neutrino beam and 3 detectors: ICARUS, a 500 ton LAr TPC at a distance of 600 m, MicroBooNE which is located 470 meters away and has 80 tons of liquid argon, and Short-Baseline Near Detector, or SBND, at 110 meters with 112 tons fiducial mass. It aims at testing the sterile neutrino explanation for the LSND and MiniBooNE anomalies by looking for $\nu_\mu \to \nu_e$ oscillations at short distances. Thanks to the excellent detector capabilities, it has also a rich programme of exotic physics, e.g. heavy sterile neutrinos, dark matter searches. Experiments using reactor and radioactive source neutrinos with detectors at a distance of few meters from the source, such as DANSS, NEOS, SOLiD, PROSPECT, Neutrino-4, are ongoing and will be able to clarify the presence of a reactor neutrino anomaly by looking for electron antineutrino disappearance at different distances.

---

[11]This explanation does not provide a particularly good fit to the MiniBooNE energy spectrum and alternative explanations for the latter have been put forward (see e.g. [69]).

### 3.4 Current knowledge of neutrino oscillation parameters and plans for the future

Thanks to the impressive programme discussed above, we have now a quite precise picture of neutrino properties, although some key questions remain unanswered.

The $\Delta m_{21}^2$ mass squared splitting is determined with very good accuracy to be $7.39 \times 10^{-5}$ eV$^2$ with a $3\sigma$ range of $6.79$–$8.01 \times 10^{-5}$ eV$^2$ [74]. The sign of this mass squared difference is positive. $\Delta m_{31}^2$ is known slightly less precisely and its sign is not yet established, leaving open two possibilities, normal (NO) or inverted (IO) ordering, see later. The measured values slightly differ between the orderings due to subleading effects in the oscillation probabilities. For NO one has $\Delta m_{31}^2 = 2.525$ ($2.431$–$2.622$) $\times 10^{-3}$ eV$^2$, for the best fit ($3\sigma$ range) and similarly for IO $\Delta m_{32}^2 = -2.512$ ($-2.413$–$2.606$) $\times 10^{-3}$ eV$^2$ [74].

There are three mixing angles and they control the flavour content of the three mass eigenstates, given by $|U_{\alpha i}|^2$. Their values are known with quite good accuracy [74]:

$$\theta_{12} = 33.82 \, (31.61 - 36.27) \qquad \text{for both mass orderings,} \tag{39}$$

$$\theta_{23} = 49.7 \, (40.9 - 52.2) \, (\text{NO}) \qquad \theta_{23} = 49.7 \, (41.2 - 52.1) \, (\text{IO}), \tag{40}$$

$$\theta_{13} = 8.61 \, (8.22 - 8.98) \, (\text{NO}) \qquad \theta_{13} = 8.65 \, (8.27 - 9.03) \, (\text{IO}), \tag{41}$$

in degrees. We notice that all three angles are sizable and $\theta_{23}$ could even be maximal. The first hints of leptonic CPV have been reported, thanks to the combination of results from long-baseline experiments and of $\theta_{13}$ measurement by reactor neutrino experiments. Currently, there is a preference for large CP violation with $\delta = 217$ ($135$–$366$) (NO) and $\delta = 280$ ($196$–$351$) (IO), in degrees, although at $3\sigma$ the CP-conserving values $\delta = 0, 180°$ for NO are still allowed. More data is required to confirm whether CP is violated in the lepton sector. No information on the Majorana phases is currently available. In Fig. 4 we report the two-dimensional projections in the neutrino mixing parameter space after marginalization with respect to the parameters not shown. The figure is taken from Ref. [74].

#### 3.4.1 Future experiments

Future neutrino oscillation experiments aim at answering the questions related to the neutrino mass ordering, CP violation and the precise determination of the oscillation parameters, as well as providing important information on the Sun and supernovae. We provide here a concise review focusing on the main efforts currently planned.

**DUNE.** The DUNE experiment [75], exploiting the LBNF facility at Fermilab, will use a beam sourced at the Main Injector with 1.2 MW of power, upgraded to 2.4 MW after 6 years. The far detector will be constituted by 4 10 kton modules of LAr TPC, located at the Sanford Underground Research Facility site, at a distance of 1300 km from Fermilab. The first module is planned for 2024 and two technologies are being developed, the single phase and the dual phase LAr TPC ones. A near detector will be located at $\sim 500$ m from the target. Its design is being finalised. The flux has a broad spectrum with a peak around 3 GeV and a significant component at lower energies to optimise the sensitivity to CP violation. Thanks to the long baseline this experiment will see strong matter effects and will be able to determine the mass ordering at $5\sigma$ irrespective of the value of $\delta$. The main drive for the experiment is the discovery of CP violation: DUNE will reach $3\sigma$ for 75% of the values of $\delta$ after an exposure of 1320 (850) kton MW years using the CDR (optimised) beam, and $5\sigma$ for 50% of the values of $\delta$ after 810 (550) kton MW years [75]. The LAr far detector is also an ideal target for SN neutrinos and will see atmospheric and solar neutrinos with a rich programme for non-accelerator neutrino physics and for proton decay.

**Hyper-Kamiokande and T2HK long-baseline experiment.** The Water-Cherenkov Hyper-Kamio-
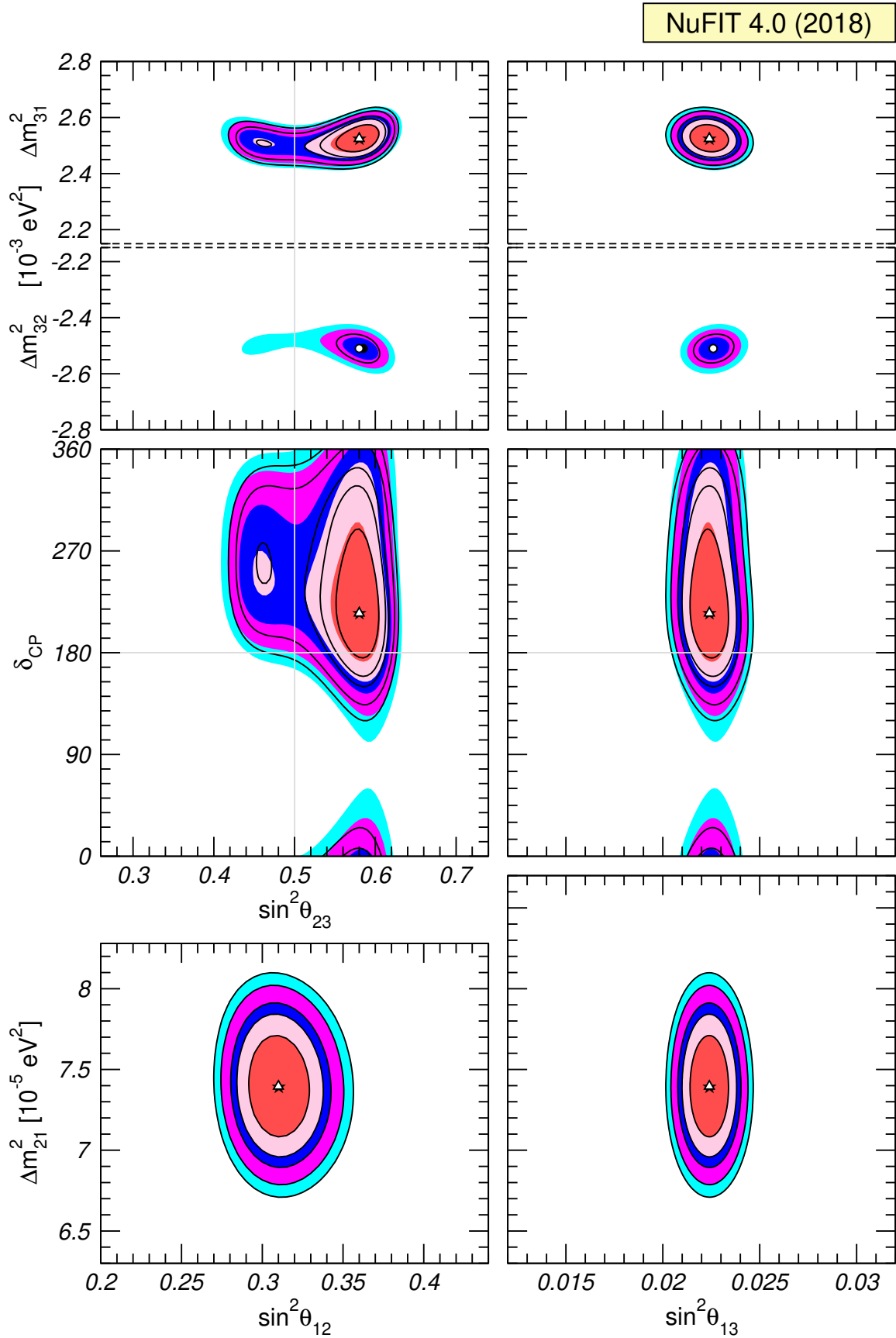
**Fig. 4:** Two-dimensional allowed regions in the neutrino parameters space at 1$\sigma$, 90%, 2$\sigma$, 99%, 3$\sigma$ C.L.. The coloured regions (black contour curves) are obtained without (with) including the SK-atmospheric $\chi^2$ data. Figure from Ref. [74].

kande detector [76] is the successor of Super-Kamiokande and will have a 187 kton fiducial mass per 1 tank, with improved detector capabilities. The optimal tank design will comprise two cylindrical detectors, 60 m in height and 74 m in diameter, with 40% photocoverage. Its main goals are the search for proton decay, the study of astrophysical neutrinos and to act as the target for a MW beam sourced at the J-PARC accelerator to discover CP violation. In regards to the beam, its location will be $2.5^o$ off-axis at 295 km in the Tochibora mine. Its physics reach is due to excellent energy resolution for neutrino-nucleus quasi-elastic interactions, the large number of events and the low intrinsic background. With a total exposure of $1.3$ MW $\times 10^8$ s, it can establish leptonic CP violation at $3\sigma$ for 76% of the values of $\delta$ and discover it at $5\sigma$ for 58% of them, and will achieve an error on $\delta$ smaller than $22°$ for any value of $\delta$[12]. The possibility to locate a second detector in South Korea is currently being investigated [77]. The longer baseline and higher energy, for a smaller off-axis angle, would allow to improve the sensitivity to the mass ordering as well as to CP violation and its precise determination.

**Other accelerator neutrino experiments.** ESS$\nu$SB [78] would exploit a 10 MW beam, sourced at the European Spallation Source. For a far detector distance between 300 and 550 km, its spectrum is peaked around the second oscillation maximum in order to maximise the sensitivity to CP violation. With 500 kton Water-Cherenkov detector, this setup has the ability to discover CP violation at $5\sigma$ for up to 50% of the values of $\delta$. Additional studies are currently ongoing in order to further optimise this facility. A neutrino factory [79] would constitute the ultimate neutrino oscillation experiment with unsurpassed physics reach. Neutrinos are produced by the decays of high energy muons in a decay ring, which source a collimated beam of muon neutrinos and electron antineutrinos. Magnetisation is necessary at the far detector to distinguish between muon neutrinos from the beam from muon antineutrinos from $\bar{\nu}_e \to \bar{\nu}_\mu$ oscillations. In the baseline design, a 100 kton magnetised iron MIND is used, 10 GeV muons and a source-detector distance of 2000 km. Thanks to the high number of events, very low backgrounds and the wide and well known energy spectrum, this setup would achieve a superior performance. For the sake of completeness, we mention that there are non-long-baseline strategies to search for leptonic CP violation. DAE$\delta$ALUS (Decay-At-rest Experiment for $\delta_{CP}$ studies At the Laboratory for Underground Science) [80] uses a cyclotron-driven muon antineutrino beam aimed at a very large detector optimised for low energies at different distances in the few km range.

**Atmospheric neutrinos.** Very large detectors for atmospheric neutrinos are being planned or constructed and will have a very good sensitivity to the mass ordering and possibly to CP violation [81][13]. Typically, these searches are performed as part of the high energy neutrino programme of IceCube2, KM3Net, in highly instrumented regions of the detector so that a lower energy threshold can be achieved. ORCA (Oscillation Research with Cosmics in the Abyss) [82] is part of KM3Net 2.0, with an intra-distance of 9 m between the digital optical modules, and could achieve a mass ordering discovery by 2024/25 if current hints of NO are confirmed. IceCube plans a near future upgrade with 7 additional strings in the Deep Core area, densely instrumented to study GeV neutrinos, and in a second phase IceCube Gen2 with a high-density core for low-energy neutrinos (PINGU) [83].

**JUNO.** The JUNO (Jiangmen Underground Neutrino Observatory) experiment [84], due to start data taking in 2021, will have a very rich experimental programme, both in astrophysical, terrestrial and reactor neutrinos. It will use a 20 kton LSc (Liquid Scintillator) detector placed at a distance of 53 km from the 26.6 GW Yangjiang and Taishan Nuclear reactors and 700 m underground. It will have an unprecedented 3% energy resolution (at 1 MeV), necessary to study neutrino oscillations with great accuracy and in particular the interference between the solar and atmospheric amplitude contributions at $L/E \sim 10^4$ km/GeV [85, 86]. This effect is sensitive to the mass ordering and could lead to its discovery within few years in a way complementary to the exploitation of matter effects in accelerator and atmospheric neutrino experiments. Moreover, it will greatly increase the precision on the solar mixing

---

[12]The precision on $\delta$ is best for CP-conserving values while it is worse for maximal CPV, in vacuum.

[13]The latter aim is quite challenging and would require an effective volume of 5-10 Mton with an energy threshold of 0.5-1 GeV and excellent detector performance.

parameters with an expected 0.7% $1\sigma$ error on $\theta_{12}$ and 0.6% on $\Delta m_{21}^2$. The accurate determination of $\theta_{12}$ is of great importance in distinguishing flavour models as it is a typical prediction for these models and enter sum-rules and other relations between the oscillation parameters. For instance, its value for tri-bimaximal mixing is $\sin^2 \theta_{12} = 1/3$, see Sec. 6.3.

**Solar neutrinos.** In addition to the dedicated Borexino experiment, several multipurpose detectors planned for the future will be able to provide information on solar neutrinos. Liquid scintillator detectors SNO+ and JUNO, LAr DUNE and Water-Cherenkov HK, as well as future ideas for water-based LSc Theia and the Jinping Slow LSc detector, could be used for this purpose with different energy thresholds, directionality and energy resolution capabilities. They will mainly focus on the regeneration of solar electron neutrinos while traversing the Earth, the so-called day/night effect, the precise shape of the probability from low energy to high energy, and, from the astrophysics point of view, getting information on the Sun. If their energy threshold and background reduction allow, they will also aim at the observation on the CNO and hep neutrinos.

## 4  Majorana and Dirac neutrinos

Neutral fermions could be either of Dirac or Majorana type, as first suggested by E. Majorana in 1937 [16]. In the first case, the particles and antiparticles are different as it is the case, e.g., for electrons and positrons. In the latter case, there is no distinction between particle and antiparticles. In the SM only neutrinos could be of Majorana type, as they are the only known neutral fermions. Their nature is strictly related to the conservation of lepton number and therefore offers an important window on the properties of the ultimate theory of particles. Processes which violate lepton number by two units can provide information on this important question, the most sensitive of which is neutrinoless double beta decay. A rich experimental programme is ongoing and planned for the future.

### 4.1  Charge conjugation

We start by defining the charge conjugate of a fermion field as

$$(\psi)^c(x) \equiv \xi_c \, C \overline{\psi}^T(x) \,, \tag{42}$$

where $\xi_c$ is the charge parity of the field. $C$ is the charge conjugation matrix which satisfies the following properties

$$C \gamma_\mu^T C^{-1} = -\gamma_\mu \,, \tag{43}$$

$$C^\dagger = C^{-1} \,, \tag{44}$$

$$C^T = -C \,. \tag{45}$$

In the Dirac representation of the $\gamma_\mu$ matrices, one has $C = i\gamma^2\gamma^0$. Since two charge conjugation transformations must bring back the field to its initial value

$$\psi \xrightarrow{c} \xi_c \, C \overline{\psi}^T \xrightarrow{c} |\xi_c|^2 \psi \,, \tag{46}$$

we find $|\xi_c|^2 = 1$. The parameter $\xi_c$ is a phase which represents the intrinsic charge parity of the field. From here onwards we take $\xi_c = 1$ for left-handed neutrinos[14]. Under a charge conjugation transformation, we have $\mathcal{U}_c \psi(x) \mathcal{U}_c^\dagger = \psi^c(x)$.

We notice an important property, that is, the charge conjugate of a left-handed field is right-handed and viceversa. In fact, using the left-handed projector $P_L$, one can show

$$P_L C \overline{\psi_L}^T = C(\overline{\psi_L} P_L)^T = C \left( (P_R \psi_L)^\dagger \gamma^0 \right)^T = 0 \,. \tag{47}$$

---

[14]Since weak interactions violate maximally the charge conjugation symmetry, the charge parity of neutrinos is arbitrary.

Under a charge conjugation transformation, we find that $\mathcal{U}_c \psi_L(x) \mathcal{U}_c^\dagger = (\psi_R(x))^c$. Therefore, a Lagrangian which contains only left-handed fields, such as the charge and neutral current terms in the Standard Model one, cannot preserve charge conjugation as a symmetry.

## 4.2 Majorana fields

A Majorana field is defined as

$$\psi = \psi^c \, . \tag{48}$$

This condition means that particle and antiparticle are indistinguishable and therefore can only apply to neutral fields. Majorana fields have several specific properties.

 – Majorana fields satisfy the Dirac equation for particles and antiparticles

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 \, . \tag{49}$$

 – In terms of its chiral components the Majorana condition implies that the field can be written as

$$\psi = \psi_L + \psi_L^c \, , \tag{50}$$

 where we have used the fact that $(\psi_L)^c = (\psi^c)_R$.
 – Majorana fields have only 2 degrees of freedom, differently from Dirac ones which have 4.
 – Majorana fields are quantised in terms of only one type of creation operator. The Fourier expansion is given by

$$\psi(x) = \int \frac{\mathrm{d}^3 p}{(2\pi)^3 \sqrt{2E}} \sum_{h=\pm 1} \left[ a_h(p) u_h(p) e^{-ip \cdot x} + a_h^\dagger(p) v_h(p) e^{ip \cdot x} \right] \, , \tag{51}$$

 where the Majorana condition has imposed $a_h = b_h$ compared to a Dirac field. So there is only one type of operator and there is no distinction between particle and antiparticle.
 – Their electromagnetic current $j^\mu = q \bar{\psi} \gamma^\mu \psi$ vanishes exactly. Moreover, Majorana particles cannot carry any $U(1)$ quantum number.

In the SM only neutrinos are neutral fermions and can be Majorana particles. As they cannot carry any charge, this implies that lepton number is not a conserved symmetry if neutrinos are of Majorana type. This is evident from the fact that the Majorana condition is not invariant under a $U(1)_\mathscr{L}$ transformation and will become apparent considering Majorana mass terms. Therefore, the question of the nature of neutrinos is directly related to the fundamental symmetries of nature. Lepton number is an accidental symmetry of the SM, meaning that it is conserved at the Lagrangian level because of the gauge symmetry and particle content of the SM. Is the ultimate theory of particles lepton-number conserving or not? This question is intrinsically bound to neutrinos.

Commonly one still uses the notion of neutrino and antineutrino for Majorana fields, as far as neutrinos are ultrarelativistic (UR). An UR Majorana neutrino of negative helicity interacts as a Dirac neutrino with the same helicity and for this reason it is common to call this particle a neutrino. Conversely, an UR Majorana neutrino of positive helicity will behave as a Dirac antineutrino of the same helicity and will be called an antineutrino. We also stress that from the kinematic point of view Dirac and Majorana neutrinos are equivalent as they satisfy the same energy-momentum dispersion relation $E = \sqrt{\mathbf{p}^2 + m^2}$.

## 4.3 Neutrinoless double beta decay

Neutrino oscillations do not distinguish between Majorana and Dirac particles, as they conserve lepton number. To test this symmetry and establish the nature of neutrinos, it is necessary to search for processes which break lepton number. The most sensitive of these is neutrinoless double beta decay (DBD0$\nu$).

This process takes place in nuclei when two neutrons simultaneously decay into two protons and two electrons, with no neutrino emission. Its SM counterpart is the two-neutrino double beta decay (DBD2$\nu$), first proposed by M. Goeppert-Mayer in 1935 [87], in which two electron antineutrinos are produced:

$$\mathcal{N}(A, Z) \quad \rightarrow \mathcal{N}(A, Z+2) + 2e^- + 2\overline{\nu}_e \qquad \text{for DBD2}\nu\,, \tag{52}$$

$$\mathcal{N}(A, Z) \quad \rightarrow \mathcal{N}(A, Z+2) + 2e^- \qquad\qquad \text{for DBD0}\nu\,. \tag{53}$$

In 1939, W. H. Furry [88] proposed that this process could proceed without emission of neutrinos, if the latter are Majorana particles. Differently from DBD2$\nu$, neutrinoless double beta decay violates lepton number by two units and is not allowed by the SM. For this reason, its discovery would be of paramount importance and would imply that neutrinos are of Majorana type.

We restrict the discussion to the simplest and most studied case in which we add to the SM massive neutrinos, as required by the oscillation data. We assume that neutrinos are Majorana particles. The inverse of the half-life is given by

$$T_{\beta\beta_{0\nu}}^{-1} \simeq \frac{G_{0\nu}}{m_e}\, |m_{\beta\beta}|^2\, M_{\text{NUCL}}^2\,, \tag{54}$$

where $G_{0\nu}$ is a known phase-space factor, $m_e$ is the electron mass, $M_{\text{NUCL}}$ is the nuclear matrix element for the nucleus of the process (NME). $m_{\beta\beta} \equiv m_{ee} \equiv \langle m \rangle$ is the *effective Majorana mass parameter* which embeds all the dependence on neutrino quantities as

$$|m_{\beta\beta}| \equiv \left| m_1 |U_{e1}|^2 + m_2 |U_{e2}|^2\, e^{i\alpha_{21}} + m_3 |U_{e3}|^2\, e^{i(\alpha_{31} - 2\delta)} \right|\,. \tag{55}$$

Here, $m_i$, $i = 1, 2, 3$, indicate the three light neutrino masses, and $U_{ei}$ are the elements of the first row of the PMNS lepton mixing matrix.

### 4.3.1 *Predictions for the effective Majorana mass parameter*

From Eq. (55) we see that the predicted value of $|m_{\beta\beta}|$ depends critically on the neutrino mass spectrum and on the values of the two Majorana phases in the PMNS matrix, $\alpha_{21}$ and $\alpha_{31}$ (see, e.g. Refs. [89, 90] and also Ref. [91]). We can consider the three limiting neutrino mass spectra discussed in Sec. 5, and we find that

$$|m_{\beta\beta}|^{\text{NH}} \quad\simeq\quad \left| \sqrt{\Delta m_{21}^2} \sin^2\theta_{12} \cos^2\theta_{13} + \sqrt{\Delta m_{31}^2} \sin^2\theta_{13} e^{i(\alpha_{32} - 2\delta)} \right|\,, \tag{56}$$

$$|m_{\beta\beta}|^{\text{IH}} \quad\simeq\quad \sqrt{|\Delta m_{32}^2|}\, \cos^2\theta_{13} \sqrt{1 - \sin^2 2\theta_{12} \sin^2\left(\frac{\alpha_{21}}{2}\right)}\,, \tag{57}$$

$$|m_{\beta\beta}|^{\text{QD}} \quad\simeq\quad m_0 \left| (\cos^2\theta_{12} + \sin^2\theta_{12} e^{i\alpha_{21}}) \cos^2\theta_{13} + e^{i(\alpha_{31} - 2\delta)} \sin^2\theta_{13} \right|\,. \tag{58}$$

We can obtain the predictions for $|m_{\beta\beta}|$ and consequently the decay rates, by assuming a specific mass spectrum, substituting the measured values for $\Delta m_{21}^2$, $\Delta m_{31}^2$, $\theta_{12}$, $\theta_{13}$, and varying the CPV phases in their allowed ranges. We get, including a $3\sigma$ error on the oscillation parameters,

$$|m_{\beta\beta}|^{\text{NH}} \quad\simeq\quad 1.1 - 4.2\ \text{meV} \tag{59}$$

$$|m_{\beta\beta}|^{\text{IH}} \quad\simeq\quad 15 - 50\ \text{meV} \tag{60}$$

$$|m_{\beta\beta}|^{\text{QD}} \quad\simeq\quad (0.29 - 1)m_0\,. \tag{61}$$

In the most general case, varying the minimal value of neutrino masses, we show in Fig. 5 the current predictions for $|m_{\beta\beta}|$ for the two mass orderings.
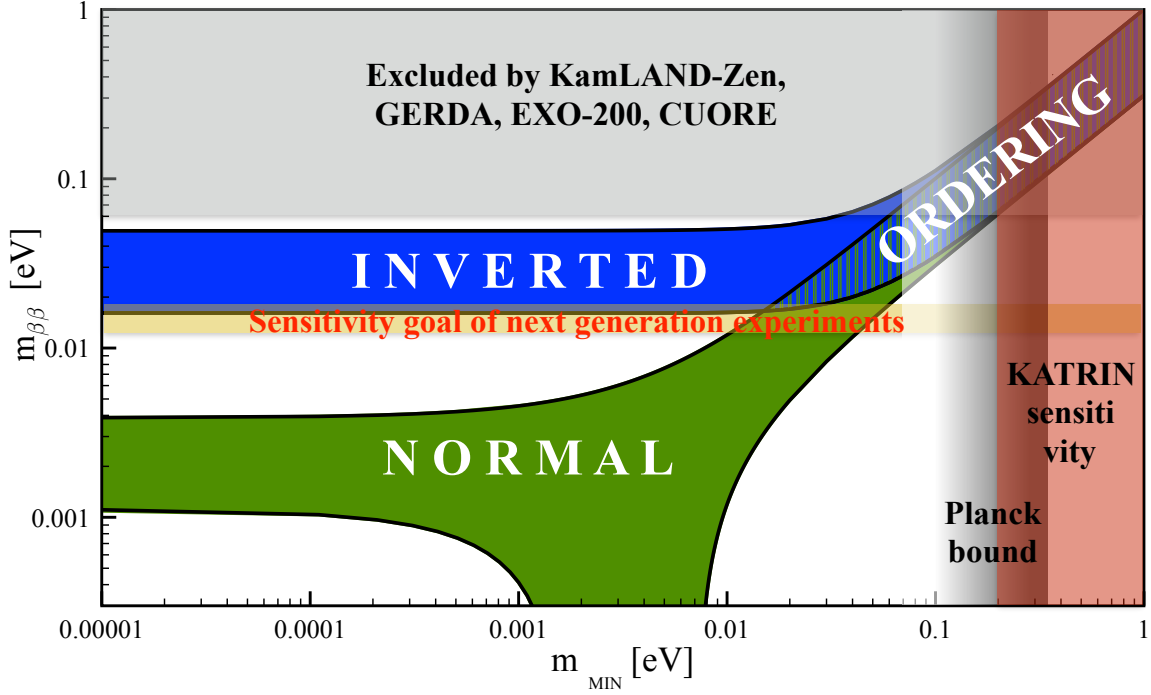
**Fig. 5:** The effective Majorana mass $|m_{\beta\beta}|$ at $2\sigma$ as a function of the smallest neutrino mass $m_{\mathrm{MIN}}$. The Majorana phases $\alpha_{21}$ and $\alpha_{31}$, and $\delta$, are varied within their allowed intervals $[0, 180°]$.

As it was noticed in Ref. [89] (see also Refs. [90]), in the case of *large but non-maximal solar mixing angle*, there is significant lower bound on $|m_{\beta\beta}|$ for IO given by

$$|m_{\beta\beta}|^{\mathrm{IO}} \geq \sqrt{|\Delta m_{32}^2|} \cos 2\theta_{12} \simeq 15 \text{ meV} . \tag{62}$$

In the case of NO the effective Majorana mass can be zero even if neutrinos are Majorana particles due a cancellation for values of $m_{\mathrm{MIN}} \sim 0.005$ eV, as shown in Fig. 5.

It follows that neutrinoless double beta decay can provide information on the neutrino mass spectrum [89, 92, 93]. In the ideal case of perfectly known NME, a measurement of $|m_{\beta\beta}| > 0.1$ eV would imply that the spectrum is QD. For values of $|m_{\beta\beta}| < 15$ meV the ordering would necessarily be normal if neutrinos are Majorana particles. For values in between, both orderings are possible, but with constraints on the masses. For instance, for $15 \text{ meV} \leq |m_{\beta\beta}| \leq 50 \text{ meV}$, the neutrino mass spectrum could be inverted hierarchical or with NO and partial hierarchy with $m_1 > 15$ meV. Similar, although somewhat weaker, conclusions can be obtained once the uncertainties on the NME and the experimental error on $|m_{\beta\beta}|$ are taken into account.

In principle, neutrinoless double beta decay could also tell us something about CP violation due to Majorana phases [89, 92, 94]. A very precise measurement of $|m_{\beta\beta}|$ and an accurate determination of the neutrino masses would open this possibility. However, this search is extremely challenging as it would require to know the NME with a very small error, at most at the few 10% level, which at present seems difficult to achieve.

### 4.3.2 Experimental status

Searches for 2-neutrino double beta decay started in the 40s and in 1950 the double beta decay half-life of $^{130}$Te was measured with geochemical methods [95]. The first observation in a laboratory experiment

was carried out in 1987 using $^{82}$Se [96]. Since then it has been observed in many other nuclei while neutrinoless DBD remains elusive. Double beta decay can be searched for in nuclei for which single beta decay is not kinematically allowed. These include e.g. $^{48}$Ca, $^{76}$Ge, $^{82}$Se, $^{100}$Mo, $^{130}$Te, $^{136}$Xe. In most experiments, the observation of double beta decay relies on the observation of the energy of the two electrons emitted. For neutrinoless double beta decay, the energy sits at the end point of the 2-electron spectrum as there is no neutrino emission and therefore the electrons carry away all the energy available. For DBD2$\nu$ this is not the case and the spectrum is continuum reaching the end point. It follows that DBD2$\nu$ is a background for DBD0$\nu$ searches and excellent energy resolution is required to distinguish between the two[15]. These experiments are very challenging as they also require very low backgrounds, hence they are located deep underground and use ultra-pure components, and need large masses because of the very slow decay rate. After several years of development and construction, a new generation of experiments has recently started and is giving new results.

We provide a concise summary of the current bounds and the future prospects. Current limits on $m_{\beta\beta}$ correspond to the quasi-degenerate neutrino mass spectrum or with partial hierarchy for either ordering. The goal of the next generation of experiments is to test the values of $m_{\beta\beta}$ predicted for the inverted ordering and initial plans to go beyond are being discussed.

– Loaded liquid scintillator detectors. KamLAND-Zen uses $^{136}$Xe which has the advantage of being available in large quantities. With an exposure of 126 kg yrs, KamLAND-Zen provides the current best bound on neutrinoless double beta decay: $T_{\beta\beta_{0\nu}} > 1.07 \times 10^{26}$ yrs and $|m_{\beta\beta}| < 61 - 165$ meV at 90% C.L. [97]. The detector mass is being increased to 750 kg of Xe in KamLAND-Zen 800 and subsequently there are plans to bring it to 1 ton in KamLAND2-Zen which, with an improved energy resolution, aims at $|m_{\beta\beta}| \sim 20$ meV close to the lower value predicted for inverted ordering. The SNO+ experiment is a multi-purpose detector using liquid scintillator and $^{130}$Te to target $T_{\beta\beta_{0\nu}} > 10^{27}$ yrs [98]. It has not yet started data taking in this configuration. The possibility to use water-based scintillators might open the option of going to much bigger scales, such as in the 50 kton THEIA proposal, which with a 3% natural Te could reach $T_{\beta\beta_{0\nu}} > 10^{28}$ yrs and go as low as 5 meV in the effective Majorana mass parameter.

– Xe-based TPCs. EXO-200, with 110 kg active mass, has reached a bound of $T_{\beta\beta_{0\nu}} > 1.8 \times 10^{25}$ yrs and $|m_{\beta\beta}| < 147 - 398$ meV at 90% C.L. [99]. Plans for nEXO with 5 tons of Xe are being considered with a reach of $T_{\beta\beta_{0\nu}} > 9.2 \times 10^{27}$ yrs going below 10 meV for $|m_{\beta\beta}|$ in ten years [100]. NEXT [101] uses high pressure enriched Xe TPC and will reach $\sim 10^{26}$ yrs in a first phase and $1.5 \times 10^{27}$ yrs in a second phase for the decay lifetime. The possibility to reduce the DBD2$\nu$ backgrounds by tagging the daughter particle Ba$^{++}$ seems promising. A similar effort is currently ongoing in China with the PANDAX-III experiment which ultimately aims at reaching the 1 ton scale [102].

– Germanium diodes. GERDA [103] uses high-purity Ge detectors, exploiting the decay of $^{76}$Ge. This type of detectors have excellent energy resolution with no intrinsic backgrounds. Combining all results, for a total of 35 kg, a limit on $T_{\beta\beta_{0\nu}} > 0.8 \times 10^{26}$ yrs at 90% C.L. has been obtained. Weaker bound has been found by the Majorana collaboration, $T_{\beta\beta_{0\nu}} > 2.7 \times 10^{25}$ yrs at 90% C.L. [104]. Majorana and GERDA have combined their plans for the future in the LEGEND detector which aims at reaching the $10^{28}$ yrs sensitivity, corresponding to $|m_{\beta\beta}| < 11 - 23$ meV [105].

– Bolometers. CUORE searches for neutrinoless double beta decay in $^{130}$Te using tellurium oxide bolometers. Combining the data from its demonstrator, CUORE-0, and Cuoricino, a bound of $T_{\beta\beta_{0\nu}} > 1.5 \times 10^{25}$ yrs at 90% C.L., corresponding to $|m_{\beta\beta}| < 110 - 520$ meV has been obtained [106]. There are plans to increase the mass and use $^{100}$Mo in $Li_2MoO_4$ crystals to reach

---

[15]Ideas about the identification of the daughter nuclei, specifically in Xe, have been put forward and might become feasible in the future.
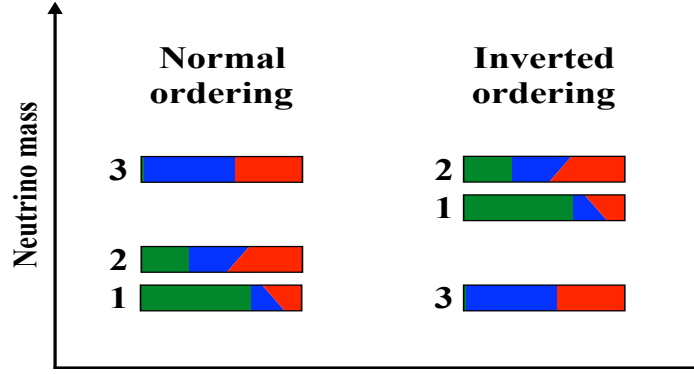
**Fig. 6:** Fractional flavour content, $|U_{\alpha i}|^2$ ($\alpha = e, \mu, \tau$), of the three mass eigenstates $\nu_i$, based on the current best-fit values of the mixing angles. $\delta$ is varied from 0 (bottom of each coloured band) to $180°$ (top of coloured band), for normal and inverted mass ordering on the left and right, respectively. The different colours correspond to the $\nu_e$ fraction (green), $\nu_\mu$ (blue) and $\nu_\tau$ (red).

$T_{\beta\beta_{0\nu}} > 2.1 \times 10^{27}$ yrs, and $|m_{\beta\beta}| < 6 - 17$ meV. This effort called CUPID [107] will exploit scintillator bolometers to reduce backgrounds.

– Other efforts. The SuperNEMO experiment exploits a unique approach to track the individual electrons emitted in the decay. It uses a thin foil of $\beta\beta$ emitter surrounded by a low-density tracker and a fast calorimeter. In this way it can provide a full reconstruction of the event topology which could provide important information in terms of angular distribution, in case of a positive signal. The first demonstrator will start taking data this year using $^{82}$Se. The COBRA (Cadmium-zinc-telluride 0-neutrino double Beta Research Apparatus) experiment uses the semiconductor CdZnTe detector technology that contains nine double beta decay isotopes: five decays $\beta^-\beta^-$ and four $\beta^+\beta^+$.

It should be noted that the extraction of $|m_{\beta\beta}|$ from a limit or future measurement on $T_{\beta\beta_{0\nu}}$ is affected by the theoretical evaluation of the NME. At present there are still large uncertainties in their computation and a strong theoretical effort is needed. Limits on $|m_{\beta\beta}|$ are given as a range which accounts for the uncertainty on the NME in the literature for a given nucleus.

## 5 Neutrino properties and open questions

The information on the mass squared differences from neutrino oscillation experiments indicates that there are three massive neutrinos and that we can order them in two ways[16]:

– normal ordering (NO): $m_1 < m_2 < m_3$, i.e. $\Delta m_{31}^2 > 0$,
– inverted ordering (IO): $m_3 < m_1 < m_2$, i.e. $\Delta m_{32}^2 < 0$.

In Fig. 6 we show the flavour content of each massive neutrino $\nu_i$ corresponding to $|U_{\alpha i}|^2$.

For each ordering[17] the three neutrino masses can be expressed in term of just one unknown parameter, the lightest neutrino mass, $m_{\text{MIN}}$, see Fig. 7. We have

$$m_1 = m_{\text{MIN}}, \qquad m_2 = \sqrt{m_{\text{MIN}}^2 + \Delta m_{21}^2}, \qquad m_3 = \sqrt{m_{\text{MIN}}^2 + \Delta m_{31}^2}, \quad \text{for NO;} \qquad (63)$$

---

[16] The convention of ordering the masses depends on the definition of the mixing angles, e.g. the correspondence between the solar mixing angle and $\theta_{12}$. We adopt here the most widely used convention for which the meaning of the mixing angles does not change between the NO and IO.

[17] We prefer the use of "ordering" rather than hierarchy for neutrino masses, as it has not yet been established that they are indeed hierarchical.
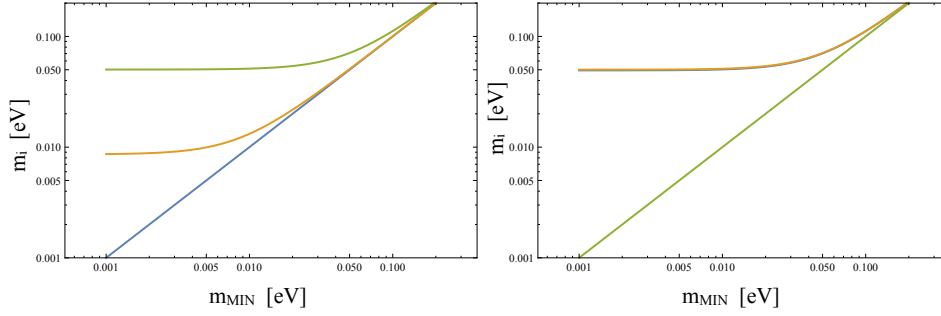
**Fig. 7:** Values of neutrino masses $m_i$ ($m_1$ in blue, $m_2$ in orange and $m_3$ in green) for normal ordering (left) and inverted ordering (right) as a function of $m_{\text{MIN}}$. The current best-fit values of the mass squared-differences have been used [74].

$$m_3 = m_{\text{MIN}}, \quad m_1 = \sqrt{m_{\text{MIN}}^2 + |\Delta m_{32}^2| - \Delta m_{21}^2}, \quad m_2 = \sqrt{m_{\text{MIN}}^2 + |\Delta m_{32}^2|}, \quad \text{for IO.} \quad (64)$$

Therefore, determining the value of neutrino masses requires to establish the neutrino mass ordering and the absolute mass scale. Three different limiting cases can be identified:

- *Normal Hierarchical Spectrum (NH).* For $m_{\text{MIN}} \to 0$, for NO we have $m_1 \ll m_2 \ll m_3$, with $m_1 \equiv m_{\text{MIN}}, m_2 \cong \sqrt{\Delta m_{21}^2}$ and $m_3 \cong \sqrt{\Delta m_{31}^2}$.
- *Inverted Hierarchical Spectrum (IH).* In the limit $m_{\text{MIN}} \to 0$, for IO we have $m_3 \ll m_1 < m_2$, with $m_{1,2} \cong \sqrt{|\Delta m_{32}^2|}$ and $m_3 \equiv m_{\text{MIN}}$.
- *Quasi-Degenerate Spectrum (QD).* For large values of $m_{\text{MIN}}$ ($m_{\text{MIN}} \gg \sqrt{|\Delta m_{31}^2|}$) the three mass eigenstates are almost degenerate, $m_i^2 \simeq m_{\text{MIN}} \equiv m_0, i = 1, 2, 3$.

On the mixing, it is interesting to note that leptonic mixing shows a very different pattern compared to quark mixing, which is rather small. In the leptonic mixing matrix, indeed two angles are very large, with $\theta_{23}$ which could be even maximal and $\theta_{12}$ which instead is far from maximality. The third mixing angle is significantly different from zero opening the possibility to search for matter effects and leptonic CP violation. The latter is one of the key open questions in this field as, in some models, it can be related to the baryon asymmetry of the Universe, see Sec. 6.4. The precise measurement of the oscillation parameters, including $\delta$, is critical to hunt for the origin of the mixing structure and solve the leptonic flavour problem.

We note that we restrict our discussion to the 3-neutrino mixing scenario. As discussed earlier, controversial hints in favour of deviations from it, in terms of sterile neutrinos possibly with non-standard properties, have been found by experiments. A consistent picture has not emerged and tension is present with disappearance experiments and cosmology. Upcoming data will shed further light on these issues and we do not discuss them further.

## 5.1 Open questions regarding neutrino properties

The key open questions in current neutrino phenomenology can be summarised as follows:

- What is the nature of neutrinos? Are neutrinos Dirac or Majorana particles?
- What are the absolute values of the masses? In order to answer this question is necessary to establish the mass ordering and the overall mass scale.
- Is there leptonic CP violation? And if so, what is the precise value of the $\delta$ phase?
- What are the precise values of the mixing angles? Do they point towards an underlying flavour principle?

– Is the standard 3-neutrino picture correct or are there other effects, such as sterile neutrinos, non-standard interactions or even more exotic ones, e.g. Lorentz-violation?

We discussed in some detail the searches for neutrinoless double beta decay which can answer the first question. As discussed previously, the neutrino mass ordering can be determined in accelerator neutrino experiments, as well as atmospheric ones, exploiting matter effects, and in JUNO. Neutrinoless double beta decay can provide information on the neutrino mass spectrum as well, if neutrinos are Majorana particles. Cosmology can test the sum of neutrino masses, see Sec. 7.3. Direct neutrino mass searches aim at measuring the neutrino masses in a model independent way. They exploit the fact that in a beta decay the electron spectrum is affected by neutrino masses around the end point, as suggested initially by Fermi [108] and Perrin [109]. More specifically, for the values of the masses to which current experiments are sensitive, that is in the QD spectrum, the differential decay rate can be expressed as

$$
\frac{d\Gamma_e}{dE_e} = \frac{G_F^2}{2\pi^3} m_e^5 \cos^2\theta_C p_e (E_e + m_e)(E_0 - E_e)\sqrt{(E_0 - E)^2 - m_0} F(E_e)|\mathrm{NME}_\beta|^2 \,, \qquad (65)
$$

where $E_e$, $p_e$ and $m_e$ are the electron kinetic energy, momentum and mass, respectively. $\theta_C$ is the Cabibbo angle, $F(E_e)$ is the Fermi function arising from the Coulomb interactions of the final particles, $\mathrm{NME}_\beta$ is the nuclear matrix element. The nucleus of choice is currently tritium as it has several advantages. Its decay is superallowed so that the nuclear matrix element is constant and the beta spectrum is uniquely determined by phase space. The energy release $Q$ is small, $Q = 18.6$ keV, which is beneficial since the high end part of the spectrum scales as $Q^{-3}$. The lifetime is not too long, $T_{1/2} = 12.3$ yrs. After the Troitzk and Mainz experiments set bounds in the eV range, $m_0 < 2.2$ eV [40], a new experimental effort using tritium is ongoing, the KATRIN project [110]. It will reach a sensitivity of $0.2$ eV after three years of beam time. Ideas about using Cyclotron Radiation Emission Spectroscopy with atomic tritium are being explored by the Project 8 collaboration, with the ultimate goal of $0.04$ eV. Other efforts, ECHo and HOLMES, exploit $^{163}$Ho and, although are not currently competitive with KATRIN, will aim at obtaining sub-eV sensitivities in the future.

The hunt for leptonic CP violation is ongoing in long baseline neutrino oscillation experiments by observing the $\nu_\mu \to \nu_e$ transition channel. As discussed, first hints have been reported by T2K and NOvA in combination with Daya Bay and future experiments, in particular DUNE and T2HK, will be able to discover it for a large fraction of the parameter space. The same experiments can also provide a precise determination of the mixing angle $\theta_{23}$, while JUNO will give the most accurate measurement of $\theta_{12}$. Although accessible in principle in neutrinoless double beta decay experiments, it will be difficult to obtain information on Majorana CPV in the near future.

Regarding the last question, we just mention that there are several searches for sterile neutrinos both in neutrino oscillation experiments, e.g. the SBN programme at Fermilab and many reactor and radioactive source experiments, as well as in beta decay and other experiments. Present and future oscillation experiments can also search for non-standard neutrino properties and new interactions. The hunt is on.

## 5.2 Complementarity and synergy between different experiments

With the vast experimental programme currently ongoing or planned for the future, a strong complementarity and synergy between different strategies is present. We mention here some relevant examples.

– The determination of the mass ordering can be achieved exploiting matter effects in long baseline neutrino oscillation experiments in a very controlled manner in DUNE, as well as in atmospheric neutrinos and by looking for the $\Delta m_{21}^2 - \Delta m_{31}^2$ interference in reactor neutrino oscillations in vacuum. Given the importance of this issue, it is critical to have multiple experiments establishing the ordering. Once this is achieved, subdominant effects can be studied e.g. in atmospheric neutrinos

and important information can be deduced on supernovae (SN) evolution, once SN neutrinos are detected.

– Neutrinoless double beta decay and neutrino oscillation experiments. As shown in Sec. 4.3.1, the predictions for $|m_{\beta\beta}|$ depend on the neutrino mass ordering. If neutrino oscillation experiments determine that the neutrino mass ordering is inverted, $|m_{\beta\beta}|$ is predicted to be bigger than 15 meV providing a clear target for the neutrinoless double beta decay experiments. Further conclusions could be obtained depending on the experimental results. Let's first assume that the ordering is established to be inverted. (i) If $|m_{\beta\beta}| \geq 15$ meV, one can conclude that neutrinos are Majorana particles. Moreover, if $|m_{\beta\beta}| > 50$ meV both upper and lower bounds on $m_3$ can be deduced, given approximately by $|m_{\beta\beta}| \leq m_3 \leq |m_{\beta\beta}|/\cos 2\theta_{12}$. Consequently, a predicted range for the sum of neutrino masses relevant in cosmology could be found. For $15$ meV $\leq |m_{\beta\beta}| \leq 50$ meV, the spectrum would need to be inverted hierarchical. In principle, if a precise measurement of the masses is obtained from cosmological observations, CPV due to Majorana phases could be hunted for but a very precise determination of $|m_{\beta\beta}|$ would be needed. (ii) If $|m_{\beta\beta}| < 15$ meV is measured, neutrinos are also established to be Majorana particles but there must be some cancellation between the standard light neutrino mass contribution and new physics. The simplest example is the case of a light see-saw mechanism in which some of the heavy neutrinos have masses below 100 MeV. (iii) If only an upper bound below 15 meV is found on $|m_{\beta\beta}|$, then the simplest conclusion would be that neutrinos are Dirac particles, although a cancellation between the three-neutrino contribution and new physics could still be at work, for instance in the case of a light see-saw. It would be of particular importance to test this second hypothesis by looking for new particles and interactions which can give a sizable contribution to neutrinoless double beta decay. Let's now consider the case in which neutrino oscillation experiments determine that the ordering is normal, as first hints seem to indicate. We have seen that the predictions for $|m_{\beta\beta}|$ go from current bounds to a complete cancellation (see Fig. 5). A measurement of $|m_{\beta\beta}|$ would establish that neutrinos are Majorana particles and would restrict their masses to a specific range. We consider values up to few meV which may be at reach in a next-to-next generation of experiments. (i) If $|m_{\beta\beta}| \gg 4$ meV, the neutrino mass spectrum has a partial hierarchy, with $|m_{\beta\beta}| \leq m_1 \leq |m_{\beta\beta}|/\cos 2\theta_{12}$ with a predicted range for $\Sigma_i m_i$ in cosmology. (ii) If $|m_{\beta\beta}| \ll 4$ meV, and the process has not been observed, no conclusion can be drawn on the nature of neutrinos.

– Cosmology and terrestrial experiments. If terrestrial experiments establish that the ordering is inverted, $\Sigma_i m_i \geq 0.1$ eV. A precise measurement of its value would lead to an accurate determination of the values of neutrino masses, with implications for neutrinoless double beta decay as discussed above. If it is found that $\Sigma_i m_i < 0.1$ eV from cosmological observations, necessarily there are new cosmological or particle physics effects which reduce the impact of neutrino masses in the formation of large scale structures or which counter them.

# 6 Neutrino masses beyond the Standard Model

As we know that neutrinos have mass, it is necessary to augment the SM Lagrangian including the neutrino mass terms and then to explain the origin of these terms in a gauge invariant manner, hunting for the ultimate theory of particles and their interactions.

## 6.1 Dirac and Majorana mass terms

Being neutrinos Dirac or Majorana, different options are available to describe their masses.

– *Dirac masses.* The Lagrangian contains a mass term

$$-\mathcal{L}_{Dirac} = \bar{\nu} m_{\rm D} \nu = \bar{\nu}_L m_{\rm D} \nu_R + \text{h.c.} \ . \tag{66}$$

We notice that such mass term requires both $\nu_L$ and $\nu_R$ and is analogous to the mass terms for the SM charged fermions. This term conserves lepton number as it is possible to give both chiral components the same lepton number, so that under a $U(1)_{\mathscr{L}}$ transformation $\nu_{L,R} \to e^{i\eta}\nu_{L,R}$

$$\mathcal{L}_{Dirac} \xrightarrow{U(1)_{\mathscr{L}}} e^{i\eta}e^{-i\eta}\mathcal{L}_{Dirac} = \mathcal{L}_{Dirac} \tag{67}$$

the mass term remains invariant. Generically, there will be several neutrinos and the mass term contains a mass matrix $M_D$. In order to find the masses, it is necessary to diagonalise this mass matrix via a biunitary transformation $V_{\nu L}^\dagger M_D V_{\nu R} = \text{diag}(m_i)$: the eigenvalues correspond to the neutrino masses and the mass states are related to the initial states as $\nu_{iL} = V_{\nu L}^\dagger \nu_L$. The matrix $V_{\nu L}$ will then enter the CC Lagrangian, together with one coming from the diagonalisation of the charged lepton mass matrix, and from there neutrino oscillations.

– *Majorana masses.* Using only one Weyl spinor $\nu_L$ it is still possible to construct a mass term using the fact that $(\nu_L)^c$ is a right-handed field. It reads

$$-\mathcal{L}_{Majorana} = \frac{1}{2}\overline{\nu^c}m_{\mathrm{M}}\nu = -\frac{1}{2}\nu_L^T C^\dagger m_{\mathrm{M}}\nu_L + \text{h.c.} . \tag{68}$$

This term is Lorentz invariant as both $\nu$ and $\nu^c$ behave in the same way under a Lorentz transformation. This term breaks lepton number by two units

$$\mathcal{L}_{Majorana} \xrightarrow{U(1)_{\mathscr{L}}} e^{2i\eta}\mathcal{L}_{Majorana} . \tag{69}$$

For multiple $\nu_L$, the mass $m_M$ is promoted to a matrix $M_{\mathrm{M}}$ which needs to be diagonalised to find the values of the masses and the corresponding eigenstates. It can be shown that this matrix is symmetric, $M_{\mathrm{M}} = M_{\mathrm{M}}^T$, and can be diagonalised using one unitary matrix

$$V_\nu^T M_{\mathrm{M}} V_\nu = m_{diag} , \tag{70}$$

where $m_{diag}$ contains the real and positive masses $m_i$. The massive fields $\nu_{i,L}$ will be related to the initial states $\nu_L$ as $\nu_{iL} = V_\nu^\dagger \nu_L$. If one defines the Majorana fields $\nu_i \equiv \nu_{i,L} + \nu_{i,L}^c$, this term can be rewritten as $-\mathcal{L}_{Majorana} = \frac{1}{2}m_i\overline{\nu}_i\nu_i$, showing that the resulting massive fields are of Majorana type.

– *Dirac plus Majorana masses.* In presence of both $\nu_L$ and $\nu_R$ fields, generically both Dirac and Majorana mass terms will be present

$$\mathcal{L}_{Dirac+Majorana} = -\overline{\nu}_L M_{\mathrm{D}}\nu_R + \frac{1}{2}\nu_L^T C^\dagger M_{\mathrm{M},L}\nu_L + \frac{1}{2}\nu_R^T C^\dagger M_{\mathrm{M},R}\nu_R + \text{h.c.} . \tag{71}$$

Defining the left-handed field

$$N_L \equiv \begin{pmatrix} \nu_L \\ \nu_R^c \end{pmatrix} , \tag{72}$$

one can rewrite these Lagrangian terms as

$$\mathcal{L}_{Dirac+Majorana} = \frac{1}{2}N_L^T C^\dagger \mathscr{M} N_L + \text{h.c.} , \tag{73}$$

where the mass matrix $\mathscr{M}$ is given by

$$\mathscr{M} = \begin{pmatrix} M_{\mathrm{M},L} & M_{\mathrm{D}}^* \\ M_{\mathrm{D}}^\dagger & M_{\mathrm{M},R} \end{pmatrix} . \tag{74}$$

Upon diagonalisation of this mass matrix, the mass eigenvalues can be found and the mass eigenstates $\nu_{i,L} = V_{D+M}^\dagger \begin{pmatrix} \nu_L \\ \nu_R^c \end{pmatrix}$. The resulting fields are Majorana, as expected since this Lagrangian breaks lepton number.

## 6.2 Neutrino masses beyond the SM

The mass terms discussed above are forbidden in the SM. There are no right-handed neutrinos and a Dirac mass term cannot be included. A Majorana mass term can be constructed using the $\nu_L$ fields only, but breaks the SM gauge invariance. Consequently, the SM, in its minimal form, does not allow for neutrino masses. In this sense, neutrino masses and mixing constitute the first particle physics evidence that the SM is incomplete. How can one extend it in order to account for neutrino masses in a consistent framework? A vast number of models has been proposed. We review here the key features of Dirac and Majorana mass models and we discuss more in detail the see-saw type I mechanism.

### 6.2.1 Dirac masses

The simplest extension which can be made to the SM involves adding new SM gauge singlets, called sterile neutrinos. We indicate them as $\nu_R$. The following Yukawa coupling is allowed by the gauge symmetries

$$-\mathcal{L}_y = \overline{L}y_\nu \cdot \widetilde{H}\nu_R + \text{h.c.} , \tag{75}$$

where $L \equiv (\nu_L^T, \ell^T)^T$ is the leptonic doublet, $\widetilde{H} = i\sigma_2 H^*$ and $H$ is the Higgs field. Once the neutral component of the Higgs field acquires a vacuum expectation value $\langle \widetilde{H} \rangle = (v_H/\sqrt{2}, 0)^T$, this term generates a Dirac mass for the light neutrinos

$$-\mathcal{L}_y \xrightarrow{\langle \widetilde{H} \rangle \neq 0} -\mathcal{L}_{Dirac} = \frac{v_H}{\sqrt{2}}\overline{\nu}_L y_\nu \nu_R + \text{h.c.} . \tag{76}$$

This Yukawa coupling and the resulting Dirac mass conserve lepton number. Indeed, as a Majorana mass term for $\nu_R$ is not forbidden by gauge invariance, its absence must be imposed by requiring lepton number conservation. In this case, this symmetry needs to be promoted from being an accidental symmetry of the SM to a fundamental ingredient of the theory of particle interactions. In this sense, this is a major departure from the Standard Model.

We can estimate the order of magnitude of the coupling $y_\nu$. Working in a one generation case, taking $m_\nu = y_\nu v_H/\sqrt{2}$ to be sub-eV, we get that $y_\nu \sim 10^{-12}$. This is a very small number and in this minimal model there is no explanation for the very strong hierarchy of masses between the charged leptons and the neutrinos. Moreover, one would naively expect a similar hierarchy between the neutrino masses and a similar mixing structure to the quark sector, contradicting the observations. For these reasons, other explanations for neutrino masses have also been considered.

### 6.2.2 Majorana masses and the Weinberg operator

Among all SM fermions, neutrinos are the only ones that can have a Majorana mass term. Noticing that the term $\overline{L} \cdot \widetilde{H}$ is gauge invariant, it is possible to construct a singlet combination [111]

$$\mathcal{L}_{M,BSM} = \frac{\lambda}{\Lambda}L^T \cdot \widetilde{H}^* C^\dagger \widetilde{H}^\dagger \cdot L + \text{h.c.} . \tag{77}$$

This term, called the Weinberg operator, has dimension 5 and requires a mass scale $\Lambda$ in the denominator. It should be pointed out that this operator is the only $D = 5$ admitted by the SM, with other effective operators being of higher dimension. Its presence is of great importance. It suggests that there is a new theory at a high scale $\Lambda$ which is integrated out at low energies. This is in analogy to the Fermi theory being the low energy realisation of the weak interactions mediated by the $W$ boson. The hunt for the new particles and interactions involved is at the centre of much of current research in theoretical neutrino physics.

The Weinberg operator breaks lepton number by two units and leads to a Majorana mass term once the Higgs boson gets a vacuum expectation value

$$\mathcal{L}_{M,BSM} \xrightarrow{\langle \widetilde{H} \rangle \neq 0} \frac{\lambda v_H^2}{2\Lambda}\nu_L^T C^\dagger \nu_L + \text{h.c.} . \tag{78}$$
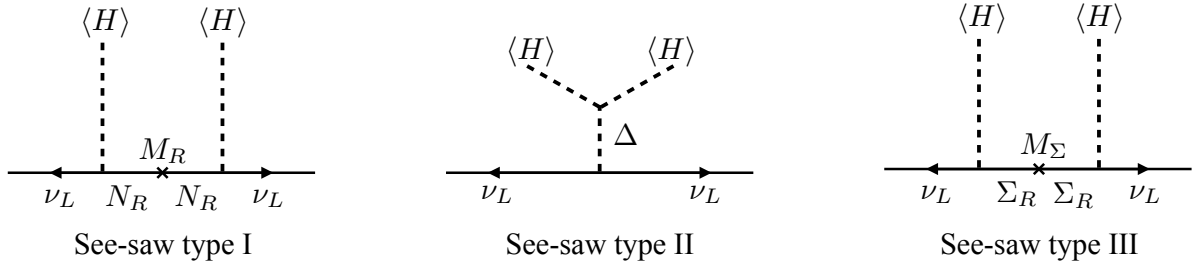
**Fig. 8:** Diagrams contributing to light neutrino masses in the three see-saw cases. $\langle H \rangle$ indicates the vev of the neutral component of the Higgs field. $\Delta$ is a scalar triplet and $\Sigma$ is the neutral component of a fermion triplet, with mass $M_\Sigma$.

The resulting massive neutrinos are of Majorana type.

### 6.2.3  *Hunting for the origin of the Weinberg operator: the seesaw mechanism*

As the Weinberg operator is a low energy effective term, the key question revolves around the physics responsible for it. We can proceed by analogy with the Fermi theory. In the latter case, the four-fermion interaction is mediated at high energy by the exchange of a massive boson at tree level. For energies much smaller than the boson mass, its presence is felt in the $m_W^2$ term in the propagator, which enters $G_F$. Also in the case of the Weinberg operator one can consider the exchange of virtual massive particles at tree level. Their mass $M$ corresponds to the scale $\Lambda$. There are three main options which have been classified based on the exchanged particle as

  – see-saw type I [112] for a singlet fermion;
  – see-saw type II [113] using heavy triplet scalars;
  – see-saw type III [114] for triplet fermions.

In Fig. 8 we schematically show the contribution to neutrino masses in the three cases. In many extensions, the scale $\Lambda$ is taken to be at or close to the Grand Unification one, $\sim 10^{14}$ GeV. The advantage of this formulation is that large couplings are allowed and the suppression of neutrino masses is due to the heavy masses. Moreover, see-saw mechanisms can be embedded in GUT theories, for instance see-saw type I naturally emerges in $SO(10)$ models. This choice remains very popular and has the added advantage that leptogenesis can be readily embedded in this framework providing an explanation for the baryon asymmetry of the Universe. The drawback is that it would be impossible to test such models, apart from a rather indirect indication coming from proton decay. Moreover, if a stabilising mechanism for the electroweak breaking scale, e.g. supersymmetry, is absent, the new physics will generically induce a correction to the Higgs mass and this suggests a scale lower than $10^7$ GeV [115]. In order to lower the scale and make the models directly testable, one can consider smaller couplings. A lot of attention has been devoted to the TeV scale as this is accessible at the LHC. New particles, such as scalar and fermion triplets as well as sterile neutrinos would leave characteristic signatures. For instance sterile neutrinos would induce same-sign dileptons plus jets with no missing energy and lepton flavour violating signals [116, 117]. Searches of this kind have been made at LHCb [118], ATLAS [119], CMS [120], Belle [121]. Allowing for even smaller couplings would lower the scale further, with heavy neutrinos with GeV, MeV and even eV masses. These low energy see-saw models have very interesting signatures which depend on their mass and flavour mixing. For eV masses, they will induce neutrino oscillations at short distances. In the keV range, their emission in beta decays distorts the electron spectrum with kinks and for very small mixing angles they could be stable on cosmological timescales providing a candidate for dark matter [122]. For MeV-GeV masses they induce peaks in the spectrum of electrons and muons emitted in meson decays [123]. These so-called peak searches are rather model independent and

provide very strong constraints on the mixing angles [124, 125] with new results being recently provided by the NA62 experiment [126]. Another strategy to search for them is to look for their decays once they are produced in meson or lepton decays [127]. Past experiments such as PS191 [128] set some of the strongest bounds for 100s of MeV masses and present/future experiments used in beam dump mode, e.g. T2K [129], NA62, DUNE, T2HK, as well the purpose made SHiP one will be able to improve on current bounds [130]. It should be noted that these bounds could be significantly modified and weakened if additional interactions lead to fast invisible decays. If the neavy neutral leptons are Majorana particles they will induce lepton number violating processes, such as neutrinoless double beta decay and LNV meson and tau decays [117, 131]. The GeV mass range is of particular interest as these sterile neutrinos could be at the origin of the baryon asymmetry of the Universe via the ARS mechanism [132, 133].

Apart from making the couplings smaller and smaller, there are other ways to enhance the testability of neutrino mass models. Research has been done in models in which neutrino masses are forbidden at tree-level and arise at the loop one [134–136]. Another possibility is to extend the see-saw framework imposing a quasi-preserved lepton number symmetry. This is the case for inverse and linear, as well as extended, see-saws [137]. For instance in the inverse see-saw mechanism, one introduces multiple sterile neutrinos and imposes a quasi-preserved lepton number symmetry. The smallness of neutrino masses is explained in terms of small lepton number violating parameters and the heavy states are pseudo-Dirac [18] particles which can have sizable mixing with the active neutrinos.

### 6.2.4 See-saw type I model

The see-saw type I mechanism is the simplest extension of the SM which can explain not only neutrino masses but also their smallness. It breaks lepton number by two units and predicts Majorana neutrinos.

Let's introduce 2 or more[19] sterile neutrinos $N_{j,R}$, $j > 1$. These are fermions with no SM gauge numbers. The most general Lagrangian which respects the SM gauge group reads

$$\mathcal{L}_{seesaw} = \mathcal{L}_{\text{SM}} - \sum_{j,\alpha} \overline{L}_\alpha y_{\alpha j} \cdot \widetilde{H} N_{j,R} + \sum_{j,k} \frac{1}{2} N_{j,R}^T C^\dagger M_{N,jk} N_{k,R} + \text{h.c.} , \tag{79}$$

where $y$ is a $3 \times j$ matrix and $M_N$ is a $j \times j$ symmetric Majorana mass matrix. Without loss of generality one can choose to work in the basis in which $M_N$ is real and diagonal with heavy masses $M_j$. Once the neutral component of the Higgs boson gets a vev, the Yukawa term will induce a Dirac mass for the neutrinos $m_D \equiv y v_H / \sqrt{2}$. In the $\nu_{\alpha,L}$–$N_{j,R}$ basis the mass terms read

$$\mathcal{L}_{seesaw,mass} = \frac{1}{2} \begin{pmatrix} (\nu_L^c)^T & N_R^T \end{pmatrix} C^\dagger \begin{pmatrix} 0 & m_D \\ m_D^T & M_N \end{pmatrix} \begin{pmatrix} \nu_L^c \\ N_R \end{pmatrix} + \text{h.c.} . \tag{80}$$

This Lagrangian is of the Dirac+Majorana form discussed in Eq. (74).

We consider the limit[20] in which $m_D \ll M_N$. Upon diagonalisation of the mass matrix, we find that the heavy neutrinos remain mainly in the sterile neutrino direction, hence their name of nearly-sterile neutrinos, and have mass $\sim M_j$. The light neutrinos, mainly in the active component, acquire a small mass

$$m_\nu \simeq -m_D \frac{1}{M_N} m_D^T . \tag{81}$$

The larger $M_N$ the smaller the neutrino masses, hence the name of *see-saw*. The smallness of neutrino masses is due to the large hierarchy between the $m_D$ and $M_N$ scales. If we take $M_N$ at the GUT scale and $m_D$ of the same order of the charged fermions, one can easily obtain neutrino masses in the right

---

[18]Pseudo-Dirac particles are Majorana particles which have nearly degenerate masses with opposite CP-parity. If the masses were exactly equal, they would form a Dirac pair.

[19]At least two heavy neutrinos are required to reproduce two mass squared differences.

[20]In the opposite case, $M \to 0$, we recover lepton number conservation and Dirac masses for the neutrinos.
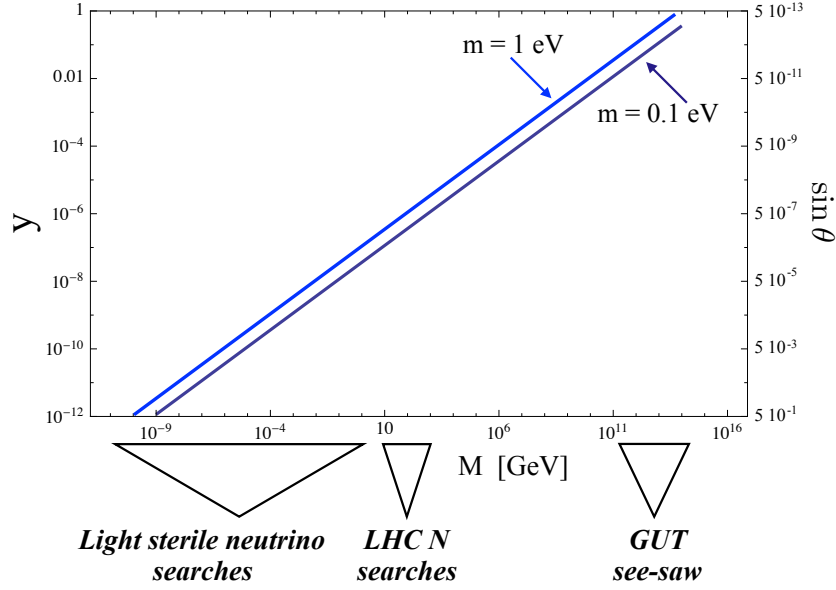
**Fig. 9:** Yukawa coupling versus nearly-sterile neutrino mass needed to generate a light neutrino mass $m = 1$–$0.1$ eV. The corresponding value of $\sin\theta$ is also indicated. Experimental strategies to search for the nearly-sterile neutrinos in the mass ranges of interest are schematically reported.

ballpark. In Fig. 9 we show the values of the Yukawa couplings required to give a light neutrino mass $m_\nu$ in a simplified one-generation case. The mixing angle between the heavy neutrino and the active one is $\sin^2\theta = m_\nu/M_N$ and therefore typically very small.

It is useful to consider the number of free parameters. For 3 sterile neutrinos, a see-saw type I model has 3 heavy masses, 9 real parameters in the Yukawa coupling and 6 phases. Of these only 3 light neutrino masses, 3 mixing angles and 3 phases can be measured at low energy[21]. If the scale is sufficiently low, some additional parameters, e.g. the heavy neutrino masses, might be at reach. A useful parameterisation allows to separate the high energy parameters from the low energy observable ones. This is the so-called Casas-Ibarra parameterisation [138]

$$y = \frac{i\sqrt{2}}{v_H} V^* \operatorname{diag}(m_\nu)^{1/2} R\, M_R^{1/2} \;, \tag{82}$$

where $R$ is a complex orthogonal matrix so that $RR^T = 1$.

### 6.3 Leptonic mixing

Apart from neutrino masses, the other key question in neutrino theory is the origin of the observed leptonic flavour structure. It is significantly different from the one in the quark sector [40, 74], which is close to the identity with correction of order $\lambda \sim 0.2$ and its powers $\lambda^2$, $\lambda^3$. We have for the $U_{\rm PMNS}$ and the CKM mixing matrices

$$|U_{\rm PMNS}|_{3\sigma} = \begin{pmatrix} 0.797 - 0.842 & 0.518 - 0.585 & 0.143 - 0.156 \\ 0.233 - 0.495 & 0.448 - 0.679 & 0.639 - 0.783 \\ 0.287 - 0.532 & 0.486 - 0.706 & 0.604 - 0.754 \end{pmatrix} \;, \tag{83}$$

$$|V_{\rm CKM}| = \begin{pmatrix} 0.97427(14) & 0.22536(61) & 0.00355(15) \\ 0.22522(61) & 0.97343(15) & 0.0414(12) \\ 0.00886(33) & 0.0405(12) & 0.99914(5) \end{pmatrix} \;. \tag{84}$$

---

[21]Realistically only one CPV phase can be measured in long baseline neutrino oscillation experiments and the determination of the two Majorana phases will remain most likely out of reach.

Motivated by the measured values of the mixing angles, in particular $\theta_{23}$ being (nearly) maximal and $\theta_{13}$ being smaller than the other two angles, a lot of attention has been devoted to leading-order patterns. They typically have $\sin^2 2\theta_{23} = 1, \sin^2 2\theta_{13} = 0$ and various values for $\theta_{12}$. Among these, the most popular patterns include the tri-bimaximal (TBM) one [139], with $\sin^2 2\theta_{23} = 1$, $\sin^2 \theta_{12} = \frac{1}{3}$, $\sin^2 \theta_{13} = 0$ and the bimaximal (BM) one [140], with $\sin^2 2\theta_{23} = 1$, $\sin^2 2\theta_{12} = 1$, $\sin^2 \theta_{13} = 0$. Other cases are the golden ratio, trimaximal or hexagonal mixing, see e.g. Ref. [141], and it is also possible to have non-zero $\theta_{13}$ from the start. The basic idea is to invoke a principle, e.g. a flavour symmetry, which leads to such a pattern and then introduce corrections to reproduce $\sin \theta_{13} \neq 0$, if needed, and deviations from the canonical values of the angles. In order to achieve this goal a specific form of the neutrino and charged lepton mass matrices is imposed.

We recall that the neutrino mass matrix and its eigenvalues (i.e., the neutrino masses) are related by a unitary matrix $V_\nu$[22] as

$$\text{diag}(m_1, m_2, m_3) = V_\nu^T \mathcal{M}_\nu V_\nu . \tag{85}$$

In a similar manner, the charged lepton mass matrix can be diagonalised using two unitary matrices $V_\ell$ and $V'_\ell$, the first acting on the chiral fields $\ell_L$ and the latter on $\ell_R$. The mass eigenstates, with definite non-negative masses, are what we commonly refer to as "electron", "muon" and "tau" leptons, which are differentiated only by their masses. The PMNS mixing matrix, which enters the CC interactions and whose elements can be measured in neutrino oscillation experiments, emerges from the product of the two transformations

$$U_{\text{PMNS}} = V_\ell^\dagger V_\nu , \tag{86}$$

which inherits the flavour structure from both the neutrino and charged lepton sectors.

Various theoretical approaches have been invoked to explain the observed structure of the PMNS matrix. As the neutrino mixing angles do not exhibit a very strong hierarchy, the notion of anarchy [142] was put forward in which the values of all the entries of $\mathcal{M}_\nu$ are of the same order. Anarchy models can reproduce the observed values of the neutrino parameters but offer no further insight into the leptonic flavour problem.

An alternative approach is to invoke a guiding principle which dictates the values of the mixing angles, given the fact that $\theta_{23}$ is close to maximal and $\theta_{12}$ is not far from special numerical values such as $1/\sqrt{3}$. The most popular strategy is to employ flavour symmetries and control their breaking. Significant work has been done using non-abelian discrete groups $S_4$, $A_4$, $A_5$, $D_N$, $\Sigma(2N^2)$, $\Sigma(3N^3)$, $\Delta(3N^2)$, $\Delta(6N^2)$ (for reviews see e.g. Refs. [143]) and the continuous groups $SU(3)$ [144] and $SO(3)$ [145]. The leptonic doublets are charged under a given flavour symmetry $G_f$ which needs to be subsequently broken, as the charged lepton masses are non-degenerate [146]. The breaking needs to lead to different structures in the charged lepton and the neutrino sectors, typically Abelian residual symmetries $G_\ell$ and $G_\nu$. These groups constrain the form of the neutrino and charged lepton mass matrices, leading to $V_\ell$ and $V_\nu$ and ultimately to $U_{\text{PMNS}}$.

We do not enter into a detailed discussion of leptonic flavour models as the literature is very vast. Instead, we give a simple illustration of how a symmetry can induce a specific value for a mixing angle. Let's assume a $L_\mu - L_\tau$ symmetry in a 2-neutrino mixing scenario. The neutrino mass matrix is forced to have the following structure:

$$\mathcal{M}_\nu = \begin{pmatrix} a & b \\ b & a \end{pmatrix} , \tag{87}$$

where we take $a$ and $b$ to be two arbitrary parameters which satisfy the hierarchy $a \ll b$. The diagonalisation of this matrix leads to two quasi-degenerate neutrino masses $b \pm a$[23] and to maximal mixing

---

[22]We refer here to the neutrino mixing matrix as $V_\nu$, in a generic basis for the charged leptons. Previously, we used the PMNS mixing matrix in the basis in which the charged leptons are in the mass eigenstate basis.

[23]A negative sign for the neutrino mass can be absorbed into a redefinition of the neutrino field via its Majorana phase.

$\theta = 45^o$. A structure of this kind could be employed to explain a maximal atmospheric mixing angle and an inverted hierarchical spectrum. This reasoning can be extended to three-neutrino mixing.

It is very common for models based on flavour symmetries to have correlations between mixing parameters, usually the mixing angles and the CPV phases [147–149]. To start with, the leading-order mass matrices have very few parameters once the flavour symmetry is imposed. The corrections introduced to shift the angles also typically depend on very few quantities, leading to correlations between the mixing parameters commonly known as *sum rules*. They can be typically divided into two main classes:

- atmospheric sum rules [148], which are of the type $\sin^2\theta_{23} = 1/2 + \zeta \sin\theta_{13}\cos\delta$, with $\zeta$ a real parameter predicted by the model. An important example is trimaximal mixing, which can be obtained from TBM pattern using a rotation in the 2-3 or 1-3 plane: $\sin^2\theta_{23} = 1/2 - \sqrt{2}\sin\theta_{13}\cos\delta$ and $\sin^2\theta_{23} = 1/2 + 1/\sqrt{2}\sin\theta_{13}\cos\delta$, respectively.
- solar sum rules [149–151] for which $\sin^2\theta_{12} = 1/3 + \zeta'\sin^2\theta_{13}\cos\delta$, with $\zeta'$ a real parameter given by the model. They often arise in models in which the leading-order mixing matrix receives corrections from the charged lepton sector. In models motivated by Grand Unified Theories (GUTs), the charged lepton mass matrix is related to the down-type quark one and $V_\ell$ gets corrections of the order of the Cabibbo angle.

It is interesting to note that, as the values of the mixing angles are known with good accuracy, the sum rules amount to predictions for the $\delta$ phase. For instance, focusing on solar sum rules, assuming that there are no 1-3 rotations in $V_\nu$ or $V_\ell$ at leading order, $\cos\delta$ is predicted to be [152]

$$\cos\delta = \frac{t_{23}\sin^2\theta_{12} + \sin^2\theta_{13}\cos^2\theta_{12}/t_{23} - \sin^2\theta_{12}^\nu(t_{23} + \sin^2\theta_{13}/t_{23})}{\sin 2\theta_{12}\sin\theta_{13}}, \tag{88}$$

where $t_{23} \equiv \tan\theta_{23}$. $\theta_{12}^\nu$ is the value in $V_\nu$ predicted by the flavour symmetry at leading order, e.g. $\sin^2\theta_{12}^\nu = 1/3$ for TBM and $\sin^2\theta_{12}^\nu = 1/2$ for BM. Present and/or future oscillation experiments can test these relations and provide useful insight on the origin of the observed leptonic flavour structure.

The possibility to impose a symmetry associated with CP at the Lagrangian level has also been investigated. It requires the introduction of a generalised CP symmetry [153–155] which is a combination of charge conjugation and parity and acts in a non-trivial manner on the flavour indices of a field $\psi$:

$$\psi(x) \to X\psi^c(x'). \tag{89}$$

Here, $x' = (t, -\mathbf{x})$ and $X$ is a symmetric unitary matrix ($XX^* = X^*X = 1$) which guarantees that the original state is recovered after applying the transformation twice. The basic idea is to combine this symmetry with a discrete flavour one. Typically, definite predictions for the CPV phases, including the Majorana ones, emerge in this scheme, leading to testable signatures in neutrino oscillation and neutrinoless double beta decay experiments.

## 6.4 Leptogenesis and the baryon asymmetry

The origin of the matter-antimatter asymmetry in the Universe is one of the most compelling questions in cosmology. Its value has been well measured using the cosmic microwave background (CMB) radiation by Planck [156]

$$Y_B^{\rm CMB} \simeq (8.67 \pm 0.09) \times 10^{-10}, \tag{90}$$

where $Y_B$ is the baryon to photon ratio at recombination and is in good agreement with the one derived from Big Bang Nucleosynthesis (BBN).

Assuming that the Universe started with the same amount of baryons and antibaryons, as suggested by inflationary models, the baryon asymmetry can be dynamically generated if the Sakharov conditions [157] are satisfied:

- *Baryon number violation* (or Lepton number violation, for the leptogenesis mechanism). In the SM $B$ and $L$ are accidental symmetries which are conserved at the perturbative level. However, $B + L$ is anomalous and transitions which violate lepton and baryon number can happen at sufficiently high temperatures via thermal excitations with topological charge called sphalerons. This implies that a violation of lepton number can induce a baryon asymmetry.
- *C and CP violation.* If CP is conserved, every reaction which produces a particle will be accompanied by the opposite one creating an antiparticle, so that there is no net creation of a baryon number.
- *Departure from thermal equilibrium.* In equilibrium, the production and destruction of a baryon asymmetry proceed with the same rate. This condition is readily satisfied by the expansion of the Universe.

Several mechanisms have been proposed to explain the origin of the baryon asymmetry. Here we focus on one of the most popular and successful: leptogenesis [158]. The basic idea is that a lepton number is created in the Early Universe and this is then converted into a baryon number. This conversion happens in the SM itself at the non-perturbative level by sphaleron effects [159].

Leptogenesis is particularly appealing because it can naturally take place in see-saw models [112]. For simplicity, we focus only on leptogenesis in see-saw type I models, but it is possible to embed this mechanism also in other neutrino mass models. A see-saw type I model can satisfy the three Sakharov conditions: (i) lepton number is violated by the Majorana $M_N$ term; (ii) CP violation can be present in the $y$ matrix if it is complex; (iii) the departure from equilibrium is due to the $N_j$ decays in the expanding Universe, once the temperature drops below their mass.

At very high temperatures, the heavy neutrinos $N_j$ are in thermal equilibrium with the rest of the plasma thanks to their Yukawa interactions

$$N_j \leftrightarrow H^0 \nu_L, \qquad N_j \leftrightarrow H^+ \ell \,. \tag{91}$$

As the Universe cools, $T$ becomes smaller than $M_j$ implying the particles in the plasma do not have sufficient energy to produce back right handed neutrinos. Only their decays are allowed

$$N_j \to H^0 \nu_L, \qquad N_j \to H^+ \ell \,. \tag{92}$$

Being Majorana neutrinos, $N_j$ can decay both into one channel and its charge conjugate one

$$N_j \to H^- \ell^+, \qquad N_j \to H^+ \ell^- \,. \tag{93}$$

If the rates of these two processes are different, due to CP violation, then a net charge asymmetry $\varepsilon$ is generated. This process is not instantaneous and washout effects, due e.g. by inverse decays, will partly erase the asymmetry. The remaining lepton asymmetry can then be converted by sphaleron processes into a baryon asymmetry. The latter depends on

$$\eta_B \propto k c_s \varepsilon_1 \,, \tag{94}$$

where $k$ is the washout factor which takes into account the fact that the decoupling is not instantaneous, $c_s$ is the sphaleron constant which quantifies how much of the lepton asymmetry is converted to a baryon asymmetry, $\varepsilon_1$ is the CP-asymmetry given by $\varepsilon_1 \equiv \frac{\Gamma(N \to lH) - \Gamma(N \to l^c H^c)}{\Gamma(N \to lH) + \Gamma(N \to l^c H^c)}$.

At high temperatures, $T > 10^{12}$ GeV, different leptonic flavours cannot be distinguished as their Yukawa interactions are out of equilibrium. In this case, assuming $M_1 \ll M_2 \ll M_3$, it is useful to consider only the total CP-asymmetry [158, 160, 161] :

$$\varepsilon_1 = \frac{\Gamma(N_1 \to H^- \ell^+) - \Gamma(N_1 \to H^+ \ell^-)}{\Gamma(N_1 \to H^- \ell^+) + \Gamma(N_1 \to H^+ \ell^-)} \simeq \frac{3}{16\pi v_H^2} \sum_{j \neq 1} \frac{\Im(m_D \, m_D^\dagger)_{1j}^2}{(m_D \, m_D^\dagger)_{11}} \frac{M_1}{M_j} \,. \tag{95}$$

Once the interactions due to the charged lepton Yukawa couplings get into equilibrium, at $T \sim 10^{12}$ GeV for $\tau$ leptons and at $T \sim 10^9$ GeV for muons, different lepton flavours become distinguishable and the asymmetry and wash-out effects become flavour-dependent. The total lepton asymmetry can be obtained summing the separate contributions of each flavour CP asymmetry washed-out by the same-lepton number violating processes [162]. Each flavour asymmetry is given by

$$\varepsilon_l = \frac{3}{16\pi v_H^2} \frac{1}{(m_D m_D^\dagger)_{11}} \Im\Big( \sum_j \Big( (m_D)_{1l}(m_D m_D^\dagger)_{1j}(m_D^*)_{jl} \Big) \Big) \frac{M_1}{M_j} \, . \tag{96}$$

Similarly, one has to consider the "wash-out mass parameter" for each flavour $l$, $\widetilde{m_l} \propto |(m_D)_{1l}|^2$, which depends on the decay rate of $N_1$ to the leptons of flavour $l$.

Since leptogenesis requires CPV, one can ask whether the one necessary to explain the baryon asymmetry is related to the CP-violating phases observable at low energies (in the PMNS mixing matrix) [163, 164]. As discussed in Sec. 6.2.4, in general see-saw models contain a larger number of parameters than those measurable. Consequently, in a completely model-independent way it is not possible to draw a direct link between the two. However, we are interested in models which aim to explain the values of neutrino masses and of the mixing structure we observe. As we briefly reviewed in Sec. 6.2.4, these models have a reduced number of parameters which can control both the high energy and low energy ones, see Sec. 6.3. In these scenarios, it is common for the low energy phases to be directly connected to the baryon asymmetry.

We can even make some more general statement. Let's restrict the discussion to see-saw type I with three hierarchical heavy neutrino masses. In the one-flavour approximation, using the Casas-Ibarra approximation, one can show that the low energy $U$ cancels out in the CP-asymmetry. This implies that it is possible to have low energy CP violation and no leptogenesis. However, this is not true if one considers flavour effects, see Eq. (96). The PMNS matrix does not cancel out in the CP asymmetry and the low energy CP-violating phases do induce a baryon asymmetry. It can also be shown that even $\delta$ can generate enough CPV to reproduce the observed baryon asymmetry. This is a highly non-trivial statement since its CPV effects are suppressed by $\theta_{13}$. One can conclude that generically the observation of lepton number violation (e.g., neutrinoless double beta decay) and of CP violation in long-baseline neutrino oscillation experiments and/or possibly neutrinoless double beta decay would constitute a strong indication (even if not a proof) in favour of the leptogenesis mechanism as the origin of the baryon asymmetry of the Universe.

## 7 Neutrinos in the Early Universe and in Astrophysics

Thanks to their interactions, neutrinos were in thermal equilibrium with the rest of the plasma in the Early Universe. As the temperature $T$ dropped, neutrinos decoupled when the interaction rate became too slow compared to the expansion of the Universe. Since then they travelled unperturbed redshifting their momentum but, due to their large density, they nevertheless affected significantly the evolution of the Universe leaving an inprint in Big Bang Nucleosynthesis, the Cosmic Microwave background, large scale structure (LSS) formation and finally pervading the Universe with a cosmic neutrino background of around 100 neutrinos per cm$^3$.

### 7.1 Neutrino decoupling.

The SM interactions such as

$$\nu\bar{\nu} \leftrightarrow e^+ e^-, \qquad \nu e \leftrightarrow \nu e, ...$$

kept neutrinos in thermal equilibrium at sufficiently high densities. They decouple when the interaction rate $\Gamma$ becomes comparable with the expansion rate $\mathcal{H}$,

$$\Gamma \sim \mathcal{H} \, ,$$

with $\Gamma \sim G_F^2 T^5$ and $\mathcal{H} \sim T^2/M_{\rm Pl}$ in a radiation dominated era. This leads to a decoupling temperature $T_{\rm d} \sim 1$ MeV, which slightly differs for electron neutrinos, being lower, from that of $\mu$–$\tau$ neutrinos as the former have both CC and NC interactions. Since we know from beta-decay measurements that neutrino masses must be below the eV scale, relic neutrinos decoupled while they were still relativistic.

## 7.2  Neutrino temperature.

Soon after neutrino decoupling, electrons-positrons get out of equilibrium transferring their entropy to the photon plasma. The neutrinos do not partake in this transfer and consequently they remain colder than the photons. Using entropy conservation, one can find that

$$\frac{T_\nu}{T_\gamma} = \left(\frac{4}{11}\right)^{1/3} .$$

## 7.3  Neutrinos as hot dark matter.

After decoupling, neutrinos have their momentum redshifted by the expansion of the Universe and turn non-relativistic at late times[24]. An inprint of their existence is on the CMB and on LSS. At early times, massive neutrinos change the matter-radiation equality leading to a small shift and height increase of the peaks of the CMB power spectrum [156], due to the Sachs-Wolfe effect [165]. Later on, their biggest effect is that they behave as hot dark matter at the time of cosmological structure formation, suppressing their growth at small scales. These structures formed from initial seeds, i.e. perturbations in the dark matter density, under the gravitational pull. Cold dark matter falls into the gravitational wells which are created by the overdensities, making them grow further and leading to the formation of galaxies and clusters. Neutrinos were too fast to be trapped in the wells and free-streamed out of them, suppressing the growth of structures at sufficiently small scales. The neutrino free-streaming length $\lambda_{FS}$ depends on their thermal velocity $v_{th}$ as [166]

$$\lambda_{FS} = 2\pi \sqrt{\frac{2}{3}} \frac{v_{th}(t)}{\mathcal{H}(t)} ,$$

where $\mathcal{H}(t)$ is the Hubble rate at a given time $t$. We notice that the thermal velocity is inversely proportional to the neutrino mass. The suppression of the matter power spectrum at scales smaller than the free-streaming length is approximately given by

$$\frac{\Delta \mathscr{P}}{\mathscr{P}} \propto 8 \frac{\Omega_\nu}{\Omega_m} , \tag{97}$$

where $\Omega_m$ is the matter density fraction and $\Omega_\nu$ the neutrino density fraction[25]. The latter is related to neutrino masses as

$$\Omega_\nu = \sum_i m_i / (93 \text{ eV} h^2) , \tag{98}$$

where $h$ is defined as $h(t) \equiv \mathcal{H}(t)/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$. As these effects are purely gravitational, there is no distinction between the different mass states which all contribute in the same manner to a good approximation. As we know that at least two neutrinos are massive, we can set a lower limit on the sum of neutrino masses which depends on the neutrino mass ordering:

$$\sum_i m_i \; > \; \sqrt{\Delta m_{21}^2} + \sqrt{\Delta m_{31}^2} \simeq 0.06 \text{ eV} \quad \text{for NO,}$$

$$\sum_i m_i \; > \; \sqrt{|\Delta m_{31}^2|} + \sqrt{|\Delta m_{32}^2|} \simeq 0.10 \text{ eV} \quad \text{for IO.}$$

---

[24]As one of the neutrinos could be nearly massless, there is the possibility that one of the three massive neutrinos is still relativistic today.

[25]$\Omega_i$ is defined as the ratio of the density of particle $i$ over the critical density of the universe $\rho_c$.

The distribution of structures in the Early Universe is measured with good accuracy, leading to a strong upper bound on the sum of neutrino masses. There are two main tools to establish the distribution of matter at the relevant scales:

– observing tracers of the matter distribution, e.g. galaxy surveys, such as SDSS [167], BOSS [168], HETDEX [169] and DES [170], or low density intergalactic gas for Ly-$\alpha$ surveys [171]. This type of measurement is very powerful but is affected by the problem of bias, i.e. the relation between the matter and the tracer power spectra;

– looking directly at the dark matter distribution via gravitational lensing [172]. This is becoming more and more the tool of choice for constraining the matter power spectrum and consequently neutrino masses.

Although cosmology provides the most stringent constraints on neutrino masses, care should be given to the fact that the bounds can vary depending on the set of cosmological data included and one needs to exploit the complementarity between different probes, for instance high and low red-shift observables, to break the degeneracy with other cosmological parameters. We just note that there is currently some tension in the measured value of $H_0$ by Planck and SN surveys, with a possible impact on neutrino parameters. Importantly, the bounds on neutrino masses are obtained assuming the standard $\Lambda$CDM model. Different underlying cosmological models, e.g. invoked to explain the accelerated expansion of the Universe, could lead to significantly different constraints. Without entering in this complex discussion, we report the bound obtained by the combination of Planck CMB and BOSS BAO [156]

$$\sum_i m_i < 0.12 \,\text{eV} \quad (95\% \text{ C.L.}).$$

In the literature other recent bounds typically are around 0.15–0.3 eV depending on the datasets included. Therefore, as current and future cosmological measurements continue to improve their precision, it is expected that cosmology will be able to distinguish between the NO and IO ordering of masses. It will be particularly interesting to assess the reliability of such bounds and to combine them with terrestrial experiments, such as neutrinoless double beta decay and direct mass searches, see Sec. 5.2.

### 7.4 Sterile neutrinos as warm dark matter

Light nearly-sterile neutrinos $\nu_h$ with masses in the keV range have been advocated as dark matter candidates. If the mixing angle between the nearly-sterile neutrino and the active neutrino is large, these particles were in thermal equilibrium in the Early Universe and decoupled at a temperature slightly higher then that of standard neutrinos. For small mixing angles, they were produced via oscillations in neutrino processes without ever reaching thermal equilibrium [173].

These particles can decay into $3\nu$ via SM NC interactions, further suppressed by the mixing angle squared. If their mass is in the keV range, they are quasi-stable on cosmological timescales and they can constitute the dominant component of DM. The production of nearly-sterile neutrinos can be obtained by solving the Boltzmann equation for the nearly-sterile neutrino phase space density $f_h$ [173, 174]

$$\left(\frac{\partial f_h}{\partial T}\right)_{E/T} = \frac{1}{4}\frac{1}{\mathcal{H}T}\Gamma_\alpha \sin^2 2\theta_m (f_\alpha - f_h)\,,$$

where $\left(\frac{\partial f_h}{\partial T}\right)_{E/T}$ is the derivative taken at constant $E/T$, $\mathcal{H} = 5.4\,T^2/M_{Pl}$, with the Plank mass $M_{Pl} = 1.2 \times 10^{19}$ GeV, $\Gamma_\alpha$ is the $\nu_\alpha$ interaction rate $\Gamma_\alpha \sim G_F^2 T^4 E$ and $f_\alpha = (1 + \exp(E/T))^{-1}$ is the phase space density for active neutrinos. $\sin^2 2\theta_m$ denotes the mixing angle in the thermal bath which is related to the one in vacuum $\theta$ as

$$\sin^2 2\theta_m \simeq \frac{\left(\frac{m_4^2}{2E}\right)^2 \sin^2 2\theta}{\left(\frac{m_4^2}{2E} + 1.1 \times 10^{-8} T^4 E/\text{GeV}^4\right)^2}\,. \tag{99}$$

The contribution of the nearly-sterile neutrino to the energy density is given by $\rho_h = m_h n_h$, where $n_h$ is the number density obtained integrating $f_h$. Other mechanisms of production invoke a lepton asymmetry which can resonantly enhance the angle $\theta_m$ and lead to a colder spectrum or assume late decays of heavier particles giving a even colder distribution of sterile neutrinos. Depending on the spectrum, these nearly-sterile neutrinos can behave as cold, cool, warm dark matter leading to intermediate behaviour in structure formation [175]. Typically, the matter power spectrum is suppressed at small structures with a sharp cut-off in the linear matter power spectrum. Confronting these predictions with observations allows to set bounds on the nearly-sterile neutrino mass, given a specific production mechanism.

Apart from the decay into $3\nu$, $\nu_h$ can also subdominantly decay into a photon and a neutrino with a branching ratio of around 1%. This channel would give rise to a x-ray line with $E_\gamma = m_4/2$ which can be searched for by looking in DM-rich regions of the Universe, such as the centre of our Galaxy and DM-dominated dwarf galaxies.

## 7.5 Supernova neutrinos

During a supernova (SN) explosion, 99% of the energy is released in neutrinos as they can escape from deep regions thanks to their weak interactions [176]. One can identify three main phases of emission: the neutronization peak, the accretion phase and the cooling phase. The neutronization peak takes place in the first $\sim$ 25 ms. Electron neutrinos are produced by electron capture on protons and nuclei. In the second phase, lasting tens to hundreds of milliseconds, the shock stalls before the start of the SN explosion. Mainly electron neutrinos are emitted and the spectrum can be rather complex due to the standing accretion shock instability. The final phase can last few tens of seconds and it is when the neutron star, born from the explosion, cools. Most of the SN gravitational binding energy is released in this phase. An intense flux of neutrinos and antineutrinos of all flavours is emitted with energies of tens of MeV. Because electron neutrinos interact more strongly with the background than electron antineutrinos, they are emitted from a neutrinosphere located less deep in the SN, implying a smaller energy. Similarly, as muon and tau neutrinos have only NC interactions with the background of electrons, protons and neutrons, their energy will be even higher. The hierarchy of average energies expected is $\langle E_{\nu_\mu, \nu_\tau} \rangle > \langle E_{\bar{\nu}_e} \rangle > \langle E_{\nu_e} \rangle$.

As the SN neutrinos propagate from the inner parts of the SN to the outer layers, their evolution is affected by oscillations in matter. Deep in the SN core, the neutrino density is so high that neutrinos become a background to themselves and a large matter potential due to neutrino-neutrino interactions is generated. "Collective oscillations" can take place as the entire neutrino system evolves as a single collective mode. Due to the complex matter profiles and time evolution, neutrino transition can be quite complex and present a behaviour which can vary significantly with energy, flavour and time. As SN neutrinos leave the inner parts of the SN, at smaller densities, SN neutrinos will feel the effects of the matter resonance due to $\Delta m^2_{31}$ and $\Delta m^2_{21}$. Finally, they will travel through space reaching the Earth. Here, they may undergo additional transitions if they transverse a significant amount of matter.

SN neutrinos can be hunted for by several types of experiments [177]. Large Water-Cherenkov detectors, such as Super-Kamiokande, and liquid scintillators, e.g. KamLAND and the future SNO+ and JUNO, can detect mainly electron antineutrinos via inverse beta decay. The addition of gadolinium, already ongoing in Super-Kamiokande, significantly improves the neutron capture detection efficiency and reduces the backgrounds enhancing the sensitivity. Elastic scattering can also give an important, although much smaller, sample of neutrinos mainly of the electron flavour. NC and other interactions on nuclei give a small contribution to the number of events. LAr detectors, such as DUNE at SURF, have unique sensitivity to electron neutrinos via the reaction $\nu_e{}^{40}Ar \rightarrow e^-{}^{40}K^*$ and can detect antineutrinos via inverse beta decay and NC interactions.

On 24 February 1987 a bright supernova of type II was found in the Large Magellanic Cloud about 50 kpc away from the solar system [178]. It was named SN1987A. The neutrinos emitted by this supernova were observed at the neutrino detectors operating at the time: Kamiokande-II [179], IMB [180],

Baksan and LSD. They recorded few tens of events in a time window of 10 seconds few hours before the optical discovery of SN1987A [26]. Their observation allows to set bounds on neutrino properties. Their mass can be bound requiring that the spread in time of the events does not exceed few seconds. This translates into a bound of around 5-30 eV, depending on the neutrino emission assumptions. Moreover, they imply that a large fraction of the SN energy was emitted in neutrinos, limiting the amount which can be carried away by other invisible particles and therefore constraining neutrino interactions in extensions of the Standard Model, their decay time and their mixing with heavy neutrinos. The observation of the SN1987A neutrinos marked the start of extrasolar system neutrino astronomy. M. Koshiba received the Nobel Prize in Physics in 2012 for this discovery.

## 8  Conclusions and Outlook

Thanks to an impressive series of experiments, we have established that neutrino oscillate. This discovery is of ground breaking importance as it implies that neutrinos have mass and mix and constitute the first, and so far only, particle physics evidence that the SM is incomplete. The two key questions, possibly inter-related, which emerge concern the origin of neutrino masses and the principle behind the observed leptonic structure. In order to address them, we require a picture of neutrino properties as complete and as precise as possible. This means identifying the nature of neutrinos, establishing the absolute values of neutrino masses, by determining their ordering and the mass scale, measuring the mixing angles and the CP violating phase very precisely, and testing if the standard 3-neutrino mixing scenario is correct. Despite being challenging, an exciting experimental programme is ongoing and planned for the future and will be able to address these issues. In these lectures I have reviewed the phenomenology relevant in current and future neutrino experiments, the theory behind neutrino masses and mixing and the impact of neutrinos in the Early Universe, from leptogenesis to dark matter.

## 9  Acknowledgements

## References

[1] https://www.symmetrymagazine.org/article/march-2007/neutrino-invention.  See also *Cambridge Monogr. Part. Phys. Nucl. Phys. Cosmol.*, **14** (2000) 1.

[2] J. Chadwick, *Nature* **129** (1932) 312.

[3] E. Fermi, *Nuovo Cim.* **11** (1934) 1.

[4] H. Bethe and R. Peierls, *Nature* **133** (1934) 532.

[5] See, B. Pontecorvo, *Cambridge Monogr. Part. Phys. Nucl. Phys. Cosmol.*, **1** (1991) 25.

[6] F. Reines and C. L. Cowan, *Nature* **178** (1956) 446; C. L. Cowan *et al.*, *Science* **124** (1956) 103.

[7] T. D. Lee and C.-N. Yang, *Phys. Rev.* **104** (1956) 254.

[8] C. S. Wu *et al.*, *Phys. Rev.* **105** (1957) 1413.

---

[26]The LSD events were recorded 5 hours before those of the other three detectors and their origin remains controversial. The Baksan Underground Scintillation Telescope could not provide an independent observation due to the low statistics but when supplemented with information from the other two detectors they could identify around 5 events in the relevant time window.

[9] M. Goldhaber, L. Grodzins and A. W. Sunyar, *Phys. Rev.* **109** (1958) 1015.

[10] L. Landau, *Nucl. Phys.* **3** (1957) 127.

[11] T. D. Lee and C. N. Yang, *Phys. Rev.* **105** (1957) 1671.

[12] A. Salam, *Nuovo Cim.* **5** (1957) 299.

[13] S. L. Glashow, *Nucl. Phys.* **22** (1961) 579.

[14] S. Weinberg, *Phys. Rev. Lett.* **19** (1967) 1264.

[15] A. Salam, Proc. of the 8th Nobel Symposium on "Elementary Particle Theory, Relativistic Groups and Analyticity", Stockholm, Sweden, 1968, edited by N. Svartholm, p. 367.

[16] E. Majorana, *Nuovo Cim.* **14** (1937) 171.

[17] E. J. Konopinski and H. M. Mahmoud, *Phys. Rev.* **92** (1953) 1045.

[18] J. C. Street and E. C. Stevenson, *Phys. Rev.* **52** (1937) 1003.

[19] S. H. Neddermeyer and C. D. Anderson, *Phys. Rev.* **51** (1937) 884.

[20] B. Pontecorvo, *Sov. Phys. JETP* **10** (1960) 1236.

[21] G. Danby *et al.*, *Phys. Rev. Lett.* **9** (1962) 36.

[22] K. Kodama *et al.* (DONUT Collaboration), *Phys. Lett.* **B 504** (2001) 218.

[23] B. Pontecorvo, *Sov. Phys. JETP* **6** (1957) 429; B. Pontecorvo, *Sov. Phys. JETP* **7** (1958) 172.

[24] Z. Maki, M. Nakagawa and S. Sakata, *Prog. Theor. Phys.* **28** (1962) 870.

[25] B. Pontecorvo, *Sov. Phys. JETP* **26** (1968) 984.

[26] V. N. Gribov and B. Pontecorvo, *Phys. Lett.* **B28** (1969) 493.

[27] S. Eliezer and A. R. Swift, *Nucl. Phys.* **B105** (1976) 45; H. Fritzsch and P. Minkowski, *Phys. Lett.* **B62** (1976) 72; S. M. Bilenky and B. Pontecorvo, *Sov. J. Nucl. Phys.* **24** (1976) 316; S. M. Bilenky and B. Pontecorvo, *Nuovo Cim. Lett.* **17** (1976) 569.

[28] B. T. Cleveland *et al.*, and J. Ullman, *Astrophys. J.* **496** (1998) 505.

[29] J. N. Bahcall, M. H. Pinsonneault, and S. Basu, *Astrophys. J.* **555** (2001) 990.

[30] W. Hampel *et al.*, *Phys. Lett.* **B447** (1999) 127.

[31] J. N. Abdurashitov *et al.*, *J. Exp. Theor. Phys.* **95** (2002) 181 [*Zh. Eksp. Teor. Fiz.* **122** (2002) 211].

[32] Y. Fukuda *et al.*, *Phys. Rev. Lett.* **77** (1996) 1683.

[33] Y. Fukuda *et al.*, *Phys. Rev. Lett.* **82** (1999) 2430.

[34] Q. R. Ahmad *et al.*, *Phys. Rev. Lett.* **87** (2001) 071301; Q. R. Ahmad *et al.*, *Phys. Rev. Lett.* **89** (2002) 011301.

[35] K. Eguchi *et al.*, *Phys. Rev. Lett.* **90** (2003) 021802.

[36] C. V. Achar *et al.*, *Phys. Lett.* **18** (1965) 196; C. V. Achar *et al.*, *Phys. Lett.* **19** (1965) 78.

[37] F. Reines *et al.*, *Phys. Rev. Lett.* **15** (1965) 429.

[38] See e.g. T. J. Haines *et al.*, *Phys. Rev. Lett.* **57** (1986) 1986; K. S. Hirata *et al.*, *Phys. Lett.* **B205** (1988) 416; Y. Fukuda *et al.*, *Phys. Lett.* **B335** (1994) 237; W. W. M. Allison *et al.*, *Phys. Lett.* **B391** (1997) 491429; S. P. Ahlen *et al.*, *Phys. Lett.* **B357** (1995) 481; M. Ambrosio *et al.*, *Phys. Lett.* **B434** (1998) 451.

[39] Y. Fukuda *et al.*, *Phys. Rev. Lett.* **81** (1998) 1562.

[40] M. Tanabashi *et al.* (Particle Data Group), *Phys. Rev.* **D 98** (2018) 030001.

[41] S. M. Bilenky, J. Hosek and S. T. Petcov, *Phys. Lett.* **B94** (1980) 495.

[42] J. Schechter and J. W. F. Valle, *Phys. Rev.* **D22** (1980) 2227; M. Doi *et al.*, *Phys. Lett.* **B102** (1981) 323.

[43] C. Jarlskog, *Phys. Rev. Lett.* **55** (1985) 1039; C. Jarlskog, *Z. Phys.* **C29** (1985) 491; O. W. Greenberg, *Phys. Rev.* D **32** (1985) 1841; I. Dunietz, O. W. Greenberg and D. d. Wu, *Phys. Rev. Lett.* **55** (1985) 2935.

[44] For initial discussions of this issue, see S. Nussinov, *Phys. Lett.* **B63** (1976) 201; B. Kayser, *Phys. Rev.* D **24** (1981) 110; K. Kiers, S. Nussinov and N. Weiss, *Phys. Rev.* **D53** (1996) 537.

[45] E. K. Akhmedov and A. Yu. Smirnov, *Phys. Atom. Nucl.* **72** (2009) 1363; E. K. Akhmedov and J. Kopp, *JHEP* **04** (2010) 008 [Erratum: *JHEP* **10** (2013) 052].

[46] F. P. An *et al.*, *Phys. Rev. Lett.* **108** (2012) 171803.

[47] J. K. Ahn *et al.*, *Phys. Rev. Lett.* **108** (2012) 191802.

[48] Y. Abe *et al.*, *Phys. Rev. Lett.* **108** (2012) 131801.

[49] S. Abe *et al.*, *Phys. Rev. Lett.* **100** (2008) 221803.

[50] L. Wolfenstein, *Phys. Rev.* **D17** (1978) 2369.

[51] S. P. Mikheev and A. Yu. Smirnov, *Sov. J. Nucl. Phys.* **42** (1985) 913 [*Yad. Fiz.* **42** (1985) 1441].

[52] A. Cervera *et al.*, *Nucl. Phys.* B **579** (2000) 17 [Erratum: *Nucl. Phys.* B **593** (2001) 731]; M. Freund, *Phys. Rev.* D **64** (2001) 053003; E. K. Akhmedov *et al.*, *JHEP* **0404** (2004) 078.

[53] K. Abe *et al.* [Super-Kamiokande Collaboration], *Phys. Rev.* D **97** (2018) 072001.

[54] P. Adamson *et al.* [MINOS Collaboration], *Phys. Rev. Lett.* **110** (2013) 251801;

[55] M. G. Aartsen *et al.* [IceCube Collaboration], *Phys. Rev.* D **91** (2015) no.7, 072004

[56] P. Adamson *et al.* [MINOS Collaboration], *Phys. Rev. Lett.* **110** (2013) 171801.

[57] K. Abe *et al.* [T2K Collaboration], *Phys. Rev. Lett.* **107** (2011) 041801. K. Abe *et al.* [T2K Collaboration], *Phys. Rev. Lett.* **118** (2017) 151801; K. Abe *et al.* [T2K Collaboration], *Phys. Rev. Lett.* **121** (2018) 171802.

[58] D. S. Ayres *et al.* [NOvA Collaboration], hep-ex/0503053; P. Adamson *et al.* [NOvA Collaboration], *Phys. Rev. Lett.* **118** (2017) 151802; M. A. Acero *et al.* [NOvA Collaboration], *Phys. Rev.* D **98** (2018) 032012.

[59] N. Agafonova *et al.* [OPERA Collaboration], *Phys. Rev. Lett.* **120** (2018) 211801 [Erratum: *Phys. Rev. Lett.* **121** (2018) 139901].

[60] K. Abe *et al.* [Super-Kamiokande Collaboration], *Phys. Rev.* D **94** (2016) 052010.

[61] B. Aharmim *et al.* []SNO collaboration], *Phys. Rev.* **C88** (2013) 025501.

[62] G. Bellini *et al.* [Borexino Collaboration], *Phys. Rev.* D **82** (2010) 033006; G. Bellini *et al.*, *Phys. Rev. Lett.* **107** (2011) 141302; M. Agostini *et al.* [BOREXINO Collaboration], *Nature* **562** (2018) 505.

[63] D. Adey *et al.* [Daya Bay Collaboration], *Phys. Rev. Lett.* **121** (2018) 241805.

[64] G. Bak *et al.* [RENO Collaboration], *Phys. Rev. Lett.* **121** (2018) 201801.

[65] Y. Abe *et al.* [Double Chooz Collaboration], *JHEP* **1410** (2014) 086 [Erratum: *JHEP* **1502** (2015) 074].

[66] A. Gando *et al.* [KamLAND Collaboration], *Phys. Rev.* D **88** (2013) 033001.

[67] C. Athanassopoulos *et al.* (LSND Collaboration), *Phys.Rev. Lett.* **75** (1995) 2650; C. Athanassopoulos *et al.* (LSND Collaboration), *Phys.Rev. Lett.* **77** (1996) 3082.

[68] A. A. Aguilar-Arevalo *et al.* (MiniBooNE), *Phys.Rev. Lett.* **102** (2009) 101802; A. A. Aguilar-Arevalo *et al.* (MiniBooNE), *Phys. Rev. Lett.* **121** (2018) 221801.

[69] P. Ballett, S. Pascoli and M. Ross-Lonergan, *Phys. Rev.* D **99** (2019) 071701.

[70] J. Kopp, M. Maltoni, and T. Schwetz, *Phys.Rev. Lett.* **107** (2011) 091801; M. Dentler *et al.*, *JHEP* **1808** (2018) 010; S. Gariazzo *et al.*, *JHEP* **06** (2017) 135; G. H. Collin *et al.*, *Phys.Rev. Lett.* **117** (2016) 221801.

[71] T. Mueller *et al.*, *Phys. Rev.* **C83** (2011) 054615; P. Huber, *Phys. Rev.* **C84** (2011) 024617; G. Mention *et al.*, *Phys. Rev.* **D83** (2011) 073006.

[72] M. A. Acero, C. Giunti, and M. Laveder, *Phys. Rev.* **D78** (2008) 073009; C. Giunti and M. Laveder, *Phys. Rev.* **C83** (2011) 065504.

[73] M. Antonello *et al.* (LAr1-ND, ICARUS-WA104, MicroBooNE), arXiv:1503.01520 [physics.ins-det]; D. Cianci *et al.*, *Phys. Rev.* **D96**, 055001 (2017).

[74] I. Esteban *et al.*, *JHEP* **01** (2019) 106, NuFIT 4.1 (2019), www.nu-fit.org. See also, F. Capozzi, E. Lisi, A. Marrone and A. Palazzo, *Prog. Part. Nucl. Phys.* **102** (2018) 48 and 2019 update reported by E. Lisi at CERN Council Open Symposium on the Update of European Strategy for Particle Physics, 13-16 May 2019, Granada, Spain, https://indico.cern.ch/event/808335/

[75] C. Adams *et al.* [LBNE Collaboration], arXiv:1307.7335 [hep-ex]; R. Acciarri *et al.* [DUNE Collaboration], arXiv:1601.05471 [physics.ins-det]; R. Acciarri *et al.* [DUNE Collaboration], arXiv:1512.06148 [physics.ins-det].

[76] K. Abe *et al.*, arXiv:1109.3262 [hep-ex]; K. Abe *et al.* [Hyper-Kamiokande Proto-Coll.], *PTEP* **2015** (2015) 053C02.

[77] M. Ishitsuka, T. Kajita, H. Minakata and H. Nunokawa, *Phys. Rev.* D **72** (2005) 033003; K. Abe *et al.* [Hyper-Kamiokande proto-Collaboration], *PTEP* **2018** (2018) 063C01.

[78] E. Baussan *et al.* [ESSnuSB Collaboration], *Nucl. Phys.* B **885** (2014) 127; E. Baussan *et al.* [ESSnuSB Collaboration], arXiv:1212.5048 [hep-ex].

[79] S. Geer, *Phys. Rev.* D **57** (1998) 6989 [Erratum: *Phys. Rev.* D **59** (1999) 039903]; A. De Rujula, M. B. Gavela and P. Hernandez, *Nucl. Phys.* B **547** (1999) 21; S. Geer, O. Mena and S. Pascoli, *Phys. Rev.* D **75** (2007) 093001; S. Choubey *et al.* [IDS-NF Collaboration], arXiv:1112.2853 [hep-ex].

[80] C. Aberle *et al.*, arXiv:1307.2949 [physics.acc-ph].

[81] S. Razzaque and A. Y. Smirnov, *JHEP* **1505** (2015) 139.

[82] U. F. Katz [KM3NeT Collaboration], arXiv:1402.1022 [astro-ph.IM]; A. Domi *et al.* [KM3NeT Collaboration], *EPJ Web Conf.* **207** (2019) 04003.

[83] M. G. Aartsen *et al.* [IceCube Collaboration], *J. Phys. G* **44** (2017) 054006.

[84] F. An *et al.*, *J. Phys.* **G43** (2016) 030401.

[85] S. T. Petcov and M. Piai, *Phys. Lett.* **B 533** (2002) 94.

[86] S. Choubey, S. T. Petcov and M. Piai, *Phys. Rev.* D **68** (2003) 113006.

[87] M. Goeppert-Mayer, *Phys. Rev.* **48** (1935) 512.

[88] W. Furry, *Phys. Rev.* **56** (1939) 1184.

[89] S. M. Bilenky, S. Pascoli and S. T. Petcov, *Phys. Rev.* **D64** (2001) 053010; S. Pascoli and S. T. Petcov, *Phys. Lett.* **B544** (2002) 239.

[90] S. M. Bilenky *et al.*, *Phys. Rev.* **D54** (1996) 4432; S. M. Bilenky *et al.*, *Phys. Lett.* **B465** (1999) 193; V. Barger and K. Whisnant, *Phys. Lett.* **B456** (1999) 194.

[91] F. Vissani, *JHEP* **9906** (1999) 022; W. Rodejohann, *Phys. Rev.* D **62** (2000) 013011; H. V. Klapdor-Kleingrothaus, H. Pas and A. Y. Smirnov, *Phys. Rev.* D **63** (2001) 073005; F. Feruglio, A. Strumia and F. Vissani, *Nucl. Phys.* B **637** (2002) 345 [Addendum: *Nucl. Phys.* B **659** (2003) 359].

[92] H. Nunokawa, W. J. C. Teves and R. Zukanovich Funchal, *Phys. Rev.* D **66** (2002) 093010.

[93] S. Pascoli, S. T. Petcov and W. Rodejohann, *Phys. Lett.* B **558** (2003) 141.

[94] S. Pascoli, S. T. Petcov and L. Wolfenstein, *Phys. Lett.* B **524** (2002) 319; V. Barger, S. L. Glashow, P. Langacker and D. Marfatia, *Phys. Lett.* B **540** (2002) 247; H. Nunokawa, W. J. C. Teves and R. Zukanovich Funchal, *Phys. Rev.* D **66** (2002) 093010.

[95] M. G. Inghram and J. H. Reynolds, *Phys. Rev.* **78** (1950) 822.

[96] S. R. Elliott, A. A. Hahn and M. K. Moe, *Phys. Rev. Lett.* **59** (1987) 2020.

[97] A. Gando *et al.* [KamLAND-Zen Collaboration], *Phys. Rev. Lett.* **117** (2016) 082503 [Addendum: *Phys. Rev. Lett.* **117** (2016) 109903].

[98] V. Fischer [SNO+ Collaboration], arXiv:1809.05986 [physics.ins-det].

[99] J. B. Albert *et al.* [EXO-200 Collaboration], *Nature* **510** (2014) 229; J. B. Albert *et al.* [EXO Collaboration], *Phys. Rev. Lett.* **120** (2018) 072701.

[100] J. B. Albert *et al.* [nEXO Collaboration], *Phys. Rev.* C **97** (2018) 065503.

[101] D. Nygren, *Nucl. Instrum. Meth.* A **603** (2009) 337; J. J. Gomez-Cadenas *et al.* [NEXT Collaboration], *Adv. High Energy Phys.* **2014** (2014) 907067; J. Martín-Albo *et al.* [NEXT Collaboration], *JHEP* **1605** (2016) 159.

[102] X. Chen *et al.*, *Sci. China Phys. Mech. Astron.* **60** (2017) 061011; K. Han [PandaX-III Collaboration], arXiv:1710.08908 [physics.ins-det].

[103] M. Agostini *et al.* [GERDA Collaboration], *Phys. Rev. Lett.* **120** (2018) 132503; M. Agostini *et al.* [GERDA Collaboration], *Eur. Phys. J.* C **78** (2018) 388.

[104] S. I. Alvis *et al.* [Majorana Collaboration], arXiv:1902.02299 [nucl-ex].

[105] N. Abgrall *et al.*, *AIP Conf. Proc.* **1894** (2017) 020027.

[106] C. Alduino *et al.* [CUORE Collaboration], *Phys. Rev. Lett.* **120** (2018) 132501.

[107] [CUPID Interest Group], arXiv:1907.09376 [physics.ins-det].

[108] E. Fermi, *Ricerca Scientifica* **2** (1933) 12; E. Fermi, *Z. Phys.* **88** (1934) 161.

[109] F. Perrin, *Comptes Rendues* **197** (1933) 1625.

[110] J. Wolf [KATRIN Collaboration], *Nucl. Instrum. Meth.* A **623** (2010) 442; V. Sibille [Katrin Collaboration], *PoS NOW* **2018** (2019) 072.

[111] S. Weinberg, *Phys. Rev. Lett.* **43** (1979) 1566.

[112] P. Minkowski, *Phys. Lett.* B **67** (1977) 421; M. Gell–Mann, P. Ramond, and R. Slansky in *Supergravity*, p. 315, edited by F. Nieuwenhuizen and D. Friedman, North Holland, Amsterdam, 1979; T. Yanagida, Proc. of the *Workshop on Unified Theories and the Baryon Number of the Universe*, edited by O. Sawada and A. Sugamoto, KEK, Japan 1979; R.N. Mohapatra, G. Senjanovic, *Phys. Rev. Lett.* **44**, (1980) 912.

[113] M. Magg and C. Wetterich, *Phys. Lett.* **B94** (1980) 61; J. Schechter and J. W. F. Valle, *Phys. Rev.* **D22** (1980) 2227; C. Wetterich, *Nucl. Phys.* **B187**, (1981) 343; G. Lazarides, Q. Shafi, and C. Wetterich, *Nucl. Phys.* **B181** (1981) 287; R. N. Mohapatra and G. Senjanovic, *Phys. Rev.* **D23** (1981) 165.

[114] R. Foot, H. Lew, X. G. He, and G. C. Joshi, *Z. Phys.* **C44** (1989) 441; E. Ma, *Phys. Rev. Lett.* **81** (1998) 1171.

[115] F. Vissani, *Phys. Rev.* **D57** (1998) 7027; J. A. Casas, J. R. Espinosa, and I. Hidalgo, *JHEP* **11** (2004) 057, 2004; J. D. Clarke, R. Foot and R. R. Volkas, *Phys. Rev.* **D91** (2015) 073009; I. Brivio and M. Trott, *Phys. Rev. Lett.* **119** (2017) 141801.

[116] F. del Aguila and J. A. Aguilar-Saavedra, *Nucl. Phys.* B **813** (2009) 22; P. S. B. Dev, A. Pilaftsis, and U.-k. Yang, *Phys. Rev. Lett.* **112** (2014) 081801; Y. Cai, T. Han, T. Li and R. Ruiz, *Front. in Phys.* **6** (2018) 40.

[117] A. Atre, T. Han, S. Pascoli and B. Zhang, *JHEP* **0905** (2009) 030.

[118] R. Aaij *et al.* [LHCb Collaboration], *Phys. Rev. Lett.* **112** (2014) 131802.

[119] G. Aad *et al.* [ATLAS Collaboration], *JHEP* **1507** (2015) 162.

[120] V. Khachatryan *et al.* [CMS Collaboration], *Phys. Lett.* B **748** (2015) 144; A. M. Sirunyan *et al.* [CMS Collaboration], *Phys. Rev. Lett.* **120** (2018) 221801; A. M. Sirunyan et al. [CMS Collaboration], *JHEP* **1901** (2019) 122.

[121] D. Liventsev *et al.* [Belle Collaboration], *Phys. Rev.* D **87** (2013) 071102 [Erratum: *Phys. Rev.* D **95** (2017) 099903].

[122] For a comprehensive review see e.g. K. N. Abazajian *et al.*, arXiv:1204.5379 [hep-ph]; M. Drewes *et al.*, *JCAP* **1701** (2017) 025.

[123]  R. E. Shrock, *Nucl. Phys.* **B206** (1982) 359; R. E. Shrock, *Phys. Rev.* D **24** (1981) 1232; R. E. Shrock, *Phys. Rev.* D **24** (1981) 1275.

[124]  D. Britton *et al.*, *Phys. Rev.* **D46** (1992) 885; T. Yamazaki *et al.*, *Conf. Proc.* **C840719** (1984) 262; A. Aoki *et al.*, *Phys. Rev.* **D84** (2011) 052002; A.V. Artamonov *et al.* (BNL E949 Collaboration), *Phys. Rev.* **D91** (2015) 052001.

[125]  For a compilation of previous bounds see A. Kusenko, S. Pascoli and D. Semikoz, *JHEP* **0511** (2005) 028.

[126]  C. Lazzeroni *et al.* (NA62 Collaboration), *Phys. Lett.* **B772** (2017) 712; E. Cortina Gil *et al.* [NA62 Collaboration], *Phys. Lett.* B **778** (2018) 137; L. Peruzzo [NA48/2 and NA62 Collaborations], *EPJ Web Conf.* **182** (2018) 02095.

[127]  M. Gronau, C. N. Leung, and J. L. Rosner, *Phys. Rev.* **D29** (1984) 2539; D. Gorbunov and M. Shaposhnikov, *JHEP* **10** (2007) 015.

[128]  G. Bernardi *et al.*, *Phys. Lett.* **166B** (1986) 479; G. Bernardi *et al.*, *Phys. Lett.* **B 203** (1988) 332.

[129]  A. Izmaylov and S. Suvorov, *Phys. Part. Nucl.* **48** (2017) 984.

[130]  P. Ballett, T. Boschi and S. Pascoli, arXiv:1905.00284 [hep-ph]; S. Alekhin *et al.*, *Rept. Prog. Phys.* **79** (2016) 124201; C. Ahdida *et al.* [SHiP Collaboration], *JHEP* **1904** (2019) 077.

[131]  L. S. Littenberg and R. Shrock, *Phys. Lett.* **B 491** (2000) 285.

[132]  E. K. Akhmedov, V. A. Rubakov and A. Y. Smirnov, *Phys. Rev. Lett.* **81** (1998) 1359.

[133]  T. Asaka and M. Shaposhnikov, *Phys. Lett.* **B 620** (2005) 17; P. Hernández *et al.*, *JHEP* **1510** (2015) 067; T. Hambye and D. Teresi, *Phys. Rev. Lett.* **117** (2016) 091801; P. Hernández *et al.*, *JHEP* **1608** (2016) 157; T. Hambye and D. Teresi, *Phys. Rev.* D **96** (2017) 015031; M. Drewes *et al.*, *Int. J. Mod. Phys.* **A 33** (2018)1842002.

[134]  A. Zee, *Phys. Lett.* **93B** (1980) 389; T. P. Cheng and L. F. Li, *Phys. Rev.* **D22** (1980) 2860; A. Zee, *Nucl. Phys.* **B 264** (1986) 99; K. S. Babu, *Phys. Lett.* **B203** (1988) 132.

[135]  L. Hall and M. Suzuki, *Nuc. Phys.* **B231** (1984) 419; D. Chang and R. N. Mohapatra, *Phys. Rev. Lett.* **58** (1987) 1600; E. Ma, *Phys. Rev.* **D73** (2006) 077301.

[136]  P. S. B. Dev and A. Pilaftsis, *Phys. Rev.* **D86** (2012) 113001; Y. Zhang, X. Ji, and R. N. Mohapatra, *JHEP* **10** (2013) 104; P. Ballett, M. Hostert and S. Pascoli, *Phys. Rev.* D **99** (2019) 091701.

[137]  D. Wyler and L. Wolfenstein, *Nuc. Phys.* **B218** (1983) 205; R. N. Mohapatra and J. W. F. Valle, *Phys. Rev.* **D34** (1986) 1642; M. C. Gonzalez-Garcia and J. W. F. Valle, *Phys. Lett.* **B216** (1989) 360; G. C. Branco, W. Grimus, and L. Lavoura, *Nuc. Phys.* **B312** (1989) 492; E. K. Akhmedov, M. Lindner, E. Schnapka, and J. W. F. Valle, *Phys. Lett.* **B368** (1996) 270; *Phys. Rev.* **D53** (1996) 2752; J. Kersten and A. Yu. Smirnov, *Phys. Rev.* **D76** (2007) 073005; M. B. Gavela *et al.*, *JHEP* **09** (2009) 038; J. Barry, W. Rodejohann, and H. Zhang, *JHEP* 07 (2011) 091; H. Zhang, *Phys. Lett.* **B714** (2012) 262.

[138]  J. A. Casas and A. Ibarra, *Nucl. Phys.* **B618** (2001) 171.

[139]  P. F. Harrison, D. H. Perkins and W. G. Scott, *Phys. Lett.* B **530** (2002) 167. Also, P. F. Harrison and W. G. Scott, *Phys. Lett.* B **535** (2002) 163; Z. Z. Xing, *Phys. Lett.* B **533** (2002) 85.

[140]  M. Fukugita, M. Tanimoto and T. Yanagida, *Phys. Rev.* D **57** (1998) 4429; V. D. Barger, S. Pakvasa, T. J. Weiler and K. Whisnant, *Phys. Lett.* B **437** (1998) 107; S. Davidson and S. F. King, *Phys. Lett.* B **445** (1998) 191; G. Altarelli, F. Feruglio and L. Merlo, *JHEP* **0905** (2009) 020; D. Meloni, *JHEP* **1110** (2011) 010.

[141]  See e.g. for the golden ratio, A. Datta, F. S. Ling and P. Ramond, *Nucl. Phys.* B **671** (2003) 383; L. L. Everett and A. J. Stuart, *Phys. Rev.* D **79** (2009) 085005; F. Feruglio and A. Paris, *JHEP* **1103** (2011) 101; W. Rodejohann, *Phys. Lett.* B **671** (2009) 267; A. Adulpravitchai, A. Blum and W. Rodejohann, *New J. Phys.* **11** (2009) 063026.

[142]  L. J. Hall, H. Murayama and N. Weiner, *Phys. Rev. Lett.* **84** (2000) 2572; A. de Gouvea and

H. Murayama, *Phys. Lett.* B **747** (2015) 479.

[143]  G. Altarelli and F. Feruglio, *Rev. Mod. Phys.* **82** (2010) 2701; S. F. King and C. Luhn, *Rept. Prog. Phys.* **76** (2013) 056201.

[144]  R. Alonso *et al.*, *JHEP* **1311** (2013) 187.

[145]  S. F. King, *JHEP* **0508** (2005) 105.

[146]  C. S. Lam, *Phys. Lett.* B **656** (2007) 193; A. Blum, C. Hagedorn and M. Lindner, *Phys. Rev.* D **77** (2008) 076004; C. S. Lam, *Phys. Rev.* D **78** (2008) 073015.

[147]  See e.g. S. F. King, *JHEP* **0209** (2002) 011; P. H. Frampton, S. T. Petcov and W. Rodejohann, *Nucl. Phys.* B **687** (2004) 31; G. Altarelli, F. Feruglio and I. Masina, *Nucl. Phys.* B **689** (2004) 157; R. N. Mohapatra and W. Rodejohann, *Phys. Rev.* D **72** (2005) 053001; S. Antusch and S. F. King, *Phys. Lett.* B **631** (2005) 42; K. A. Hochmuth, S. T. Petcov and W. Rodejohann, *Phys. Lett.* B **654** (2007) 177.

[148]  See e.g. S. F. King, *Phys. Lett.* B **659** (2008) 244; Y. Shimizu, M. Tanimoto and A. Watanabe, *Prog. Theor. Phys.* **126** (2011) 81; S. F. King and C. Luhn, *JHEP* **1109** (2011) 042.

[149]  S. Antusch and S. F. King, *Phys. Lett.* B **591** (2004) 104; I. Masina, *Phys. Lett.* B **633** (2006) 134.

[150]  S. Antusch, P. Huber, S. F. King and T. Schwetz, *JHEP* **0704** (2007) 060.

[151]  G. Altarelli, P. A. N. Machado and D. Meloni, *PoS CORFU* **2014** (2015) 012.

[152]  S. T. Petcov, *Nucl. Phys.* B **892** (2015) 400; also, D. Marzocca, S. T. Petcov, A. Romanino and M. C. Sevilla, *JHEP* **1305** (2013) 073; P. Ballett *et al.*, *JHEP* **1412** (2014) 122; I. Girardi, S. T. Petcov and A. V. Titov, *Nucl. Phys.* B **894** (2015) 733. .

[153]  F. Feruglio, C. Hagedorn and R. Ziegler, *JHEP* **1307** (2013) 027; F. Feruglio, C. Hagedorn and R. Ziegler, *Eur. Phys. J.* C **74** (2014) 2753; M. Holthausen, M. Lindner and M. A. Schmidt, *JHEP* **1304** (2013) 122; C. Hagedorn, A. Meroni and E. Molinaro, *Nucl. Phys.* B **891** (2015) 499.

[154]  M. C. Chen *et al.*, *Nucl. Phys.* B **883** (2014) 267.

[155]  W. Grimus and M. N. Rebelo, *Phys. Rept.* **281** (1997) 239.

[156]  N. Aghanim *et al.* [Planck Collaboration], arXiv:1807.06209 [astro-ph.CO].

[157]  A. D. Sakharov, *JTEP Lett.* **6** (1967) 24.

[158]  M. Fukugita and T. Yanagida, *Phys. Lett.* B **174** (1986) 45.

[159]  J. A. Harvey, M. S. Turner, *Phys. Rev.* D **42** (1990) 3344.

[160]  M. A. Luty, *Phys. Rev.* D **45** (1992) 455; M. Flanz, E. A. Paschos and U. Sarkar, *Phys. Lett.* B **345** (1995) 248 [Erratum-ibid. B **382** (1996) 447 ]; M. Plumacher, *Z. Phys.* C **74** (1997) 549.

[161]  L. Covi, E. Roulet and F. Vissani, *Phys. Lett.* B **384** (1996) 169; M. Flanz, E. A. Paschos, U. Sarkar and J. Weiss, *Phys. Lett.* B **389** (1996) 693; A. Pilaftsis, *Phys. Rev.* D **56** (1997) 5431; W. Buchmuller and M. Plumacher, *Phys. Lett.* B **431** (1998) 354.

[162]  R. Barbieri, P. Creminelli, A. Strumia and N. Tetradis, *Nucl. Phys.* B **575** (2000) 61; A. Abada *et al.*, *JCAP* **0604** (2006) 004; E. Nardi, Y. Nir, E. Roulet and J. Racker, *JHEP* **0601** (2006) 164; A. Abada *et al.*, *JHEP* **0609** (2006) 010.

[163]  See e.g. S. Davidson and A. Ibarra, *Nucl. Phys. B* **648** (2003) 345; G. C. Branco *et al.*, *Nucl. Phys. B* **640** (2002) 202; T. Endoh *et al.*, *Phys. Rev. Lett.* **89** (2002) 231601 .

[164]  S. Pascoli, S. T. Petcov and A. Riotto, *Phys. Rev.* D **75** (2007) 083511; S. Pascoli, S. T. Petcov and A. Riotto, *Nucl. Phys.* B **774** (2007) 1; K. Moffat *et al.*, *Phys. Rev.* D **98** (2018) 015036; K. Moffat, S. Pascoli, S. T. Petcov and J. Turner, *JHEP* **1903** (2019) 034.

[165]  R. K. Sachs and A. M. Wolfe, *Astrophys. J.* **147** (1967) 73 [*Gen. Rel. Grav.* **39** (2007) 1929].

[166]  D. J. Eisenstein and W. Hu, *Astrophys. J.* **511** (1997) 5.

[167]  M. Tegmark *et al.* [SDSS Collaboration], *Astrophys. J.* **606** (2004) 702.

[168] F. Beutler *et al.* [BOSS Collaboration], *Mon. Not. Roy. Astron. Soc.* **444** (2014) 3501.

[169] G. J. Hill *et al.*, *ASP Conf. Ser.* **399** (2008) 115.

[170] T. Abbott *et al.* [DES Collaboration], astro-ph/0510346.

[171] N. Palanque-Delabrouille *et al.*, *JCAP* **1511** (2015) 011.

[172] S. Hannestad, H. Tu and Y. Y. Y. Wong, *JCAP* **0606** (2006) 025; M. Kaplinghat, L. Knox and Y. S. Song, *Phys. Rev. Lett.* **91** (2003) 241301; R. A. Battye and A. Moss, *Phys. Rev. Lett.* **112** (2014) 051303.

[173] S. Dodelson and L.M. Widrow, *Phys. Rev. Lett.* 72 (1994) 17.

[174] K. Abazajian, G. M. Fuller and M. Patel, *Phys. Rev.* D **64** (2001) 023501.

[175] X.-D. Shi and G.M. Fuller, *Phys. Rev. Lett.* **82** (1999) 2832; K. Petraki and A. Kusenko, *Phys. Rev.* D **77** (2008) 065014; M. Shaposhnikov, *JHEP* **08** (2008) 008; M. Drewes et al., *JCAP* **01** (2017) 025.

[176] For a review see e.g. A. Mirizzi *et al.*, *Riv. Nuovo Cim.* **39** (2016) 1.

[177] K. Scholberg, *Ann. Rev. Nucl. Part. Sci.* **62** (2012) 81.

[178] V. Trimble, *Rev. Mod. Phys.* **60** (1988) 859.

[179] K. S. Hirata *et al.*, *Phys. Rev.* D **38** (1988) 448.

[180] C. B. Bratton *et al.*, *Phys. Rev.* D **37** (1988) 3361.

# Practical Statistics for Particle Physicists

*H. B. Prosper*
Florida State University, Department of Physics, Tallahassee, USA

**Abstract**
These lectures cover the basic ideas of frequentist and Bayesian analysis and
introduce the mathematical underpinnings of supervised machine learning. In
order to focus on the essentials, we illustrate the ideas using two simple exam-
ples from particle physics.

**Keywords**
Statistics; lectures; data analysis method; statistical analysis; frequentist; Bayesian.

## 1 Introduction

Statistics and physics are similar in that each starts from sets of basic principles. They are similar also in
the fact that physicists and statisticians from time to time engage in vigorous debate about the foundations
of their respective disciplines. These disciplines, of course, also differ in significant ways. For example,
physicists are forced, at some point, to bury the hatchet. Why? Because there is an ultimate judge
of the correctness of a proposed principle, namely, *Nature*. If a principle yields results that contradict
observations then the former does not apply to Nature and is, in that sense, wrong. For statisticians,
alas, their many judges are other statisticians. Consequently, they are not compelled to reach, and for
some basic questions have not reached, a consensus. Happily, however, for the typical applications in
particle physics the debate and disagreements among statisticians can usually be ignored. But this is a
poor excuse for dismissing these disagreements as even a modicum of understanding of them can avert
hours of fruitless arguments that prove, ultimately, to be about intellectual taste and therefore cannot be
adjudicated by appealing to a third party such as Nature. Therefore, while these lectures focus on the
practical, we occasionally comment on some of these disagreements.

The remainder of the introduction, presents a birds eye view of statistical analysis. For detailed
expositions on statistical analysis aimed at physicists, we recommend the books: [1–4]. For historical
perspectives see [5, 6].

### 1.1 Samples

The result of an experiment is a sample of $N$ data $X = x_1, x_2, \cdots, x_N$, which can be characterized
with quantities called statistics[1]. A **statistic** is number that can be computed from the sample alone and
known parameters. Here are a few well-known statistics:

$$\text{the } \textbf{sample moments} \qquad x_r = \frac{1}{N} \sum_{i=1}^{N} x_i^r, \qquad (1)$$

$$\text{the } \textbf{sample average} \qquad \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i, \qquad (2)$$

$$\text{and the } \textbf{sample variance} \qquad s^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2. \qquad (3)$$

---

[1]Statisticians tend to use upper case letters to denote random variables and lower case letters to denote actual values. We do
not follow this convention.

261

The sample moments give detailed information about the sample, while the sample average and variance are measures of the center of the data and their spread. Statistics that characterize the data are called **descriptive statistics**. In these lectures, we shall encounter statistics that provide more sophisticated information about samples.

## 1.2 Populations

An infinitely large sample is called a **population**, which physicists usually refer to as an ensemble. Like other abstractions, populations can be studied mathematically and can be characterized with numbers, such as those listed below. (The symbol $E[*]$ means **ensemble average**, that is, the average over the population of the quantity within the brackets.)

| | | |
|---|---|---|
| Ensemble average | $E[x]$ | |
| Mean | $\mu$ | |
| Error | $\epsilon = x - \mu$ | |
| Bias | $b = E[x] - \mu$ | |
| Variance | $V = E[(x - E[x])^2]$ | |
| Standard deviation | $\sigma = \sqrt{V}$ | |
| Mean square error | $\text{MSE} = E[(x - \mu)^2]$ | |
| Root MSE | $\text{RMS} = \sqrt{\text{MSE}}$ | (4) |

However, unlike the statistics of a sample, the numbers that characterize a population are abstractions. After all, no one has ever amassed an infinity of anything. In practice, a population is approximated by a large sample. Such "populations" are the basis of a statistical method called the bootstrap, in which various quantities can be approximated by treating the sample as if it were a population. Large, typically simulated, samples are used in physics analyses to assess, for example, the effect of systematic uncertainties or to confirm that an analysis method performs as claimed. In a simulated "population" some quantities can be computed exactly, for example the *error* associated with each element of the "population" because $x$ and $\mu$ are known. Quantities such as bias, however, can only be approximated.

While it may not be possible to calculate a population quantity exactly, it often possible to relate one population quantity to another, which can sometimes provide useful insight. Take for example the mean square error (MSE), whose square root is called the root mean square (RMS)[2]. The MSE can be written as

$$\text{MSE} = V + b^2. \tag{5}$$

**Exercise 1:** Show this

This is an instructive result. Suppose, for example, that $\mu$ is the true Higgs boson mass and $x$ is a measurement of it. If the MSE is used as a measure of the accuracy of the mass measurements, then the result in Eq. (6) shows that correcting a measurement of the mass for bias makes sense only if, on the average, the bias-corrected results yield a smaller MSE than that of the uncorrected result. Making a bias correction may not always be the correct thing to do if the goal is to arrive at mass measurements, which, on average, are as close to the true value of the mass in the MSE sense. Using simulations to study and understand the characteristics of a population is both useful and educational. It is good practice to do lots of simple simulations (sometimes called *toy* experiments) in order to develop an intuition about statistical quantities and the behavior of statistical procedures as well as to decide whether a particular manipulation of a measurement—e.g., a bias correction—makes sense.

---

[2]The RMS and standard deviation are sometimes used interchangeably. The two quantities are identical only if the bias is zero.

Another example of the ability (and utility) of mathematical analysis with respect to a population is the calculation of the bias in the variance of a sample. When we speak of "bias in a measurement $x$", say a measurement of the Higgs boson mass, we should remember that this phrasing is a shortcut. There is very likely an *error* in $x$, which in the real world is unknown. But, strictly speaking, *bias* does not apply to $x$, but rather to the ensemble to which $x$ is presumed to belong. However, it would quickly become horribly pedantic not to use the shortcut "bias in $x$", so it is perfectly reasonable to use it so long as we remember what the phrase means. The ensemble average of the sample variance, Eq. (3), is given by

$$E[s^2] = E[\overline{x^2}] - E[\bar{x}^2],$$
$$= V - \frac{V}{N}.$$

Exercise 2: Show this

and has a bias of $b = -V/N$. The result shows that the bias can be calculated exactly only if the variance $V$ is known exactly.

## 1.3 Statistical Inference

The main goal of a theory of statistical inference is to use a sample to infer something about the associated population. We may wish to estimate (that is, measure) a parameter associated with the population, for example, the mean Higgs boson signal in the proton-proton to 4-lepton channel. Then, in order to make this estimate meaningful, we need to quantify its accuracy. Finally, we may wish to assess to what degree we can claim the signal is real and not an apparent signal caused by a fluctuation of the background. We shall consider each of these tasks using the two most commonly used theories of inference, frequentist and Bayesian. In both theories, the foundational concept is **probability**, albeit interpreted in two different ways:

- **Degree of belief** in, or assigned to, a proposition, e.g.,
  - *proposition*: it will rain in Maratea tomorrow
  - *probability*: $p = 5 \times 10^{-2}$

- **Relative frequency** of given outcomes in an infinite set of trials, e.g.,
  - *trial*: a proton-proton collision at the Large Hadron Collider (LHC)
  - *outcome*: creation of a Higgs boson
  - *probability*: $p = 5 \times 10^{-10}$

Since each theory of inference uses a different interpretation of probability, it is not surprising that the interpretation of their results differ. This can cause confusion, especially when both theories give numerically identical results. When data are plentiful, these interpretation typically do not affect how the results are subsequently used. The difficulties arise when sample sizes are small and when each approach can yield substantially different results. This is when intellectual taste becomes the main arbiter of which approach is considered the more reasonable.

The next two sections cover the application of frequentist and Bayesian theories of statistical analysis in particle physics using a simple real-word example, while the last section provides an introduction to supervised machine learning.

## 2 Frequentist Analysis

In 2014, the CMS Collaboration published its measurement of the properties of the Higgs boson in the 4-lepton final states [7]. We shall analyze the summary results of this analysis, namely, $N = 25$ observed

4-lepton events with a background estimate of $B \pm \delta B = 9.4 \pm 0.5$ events. The goal is to make statements about the mean Higgs boson event count $s$—that is, the signal, where $d = b + s$ is the mean event count and $b$ is the mean background count. Although these data are very simple, they are sufficient to illustrate the essential ideas of frequentist analysis.

Whether the data are to be analyzed using a frequentist or Bayesian approach, the starting point is the same: the first task is constructing an accurate probability model for the mechanism that generates the data.

## 2.1 The Probability Model

Given the observed count $N = 25$ events, a particle physicist would immediately model the data generation mechanism with a Poisson distribution,

$$\text{Poisson}(n, d) = \frac{e^{-d}d^n}{n!},$$

because everyone knows that is the distribution for a counting experiment. If the data comprised $M$ counts $N_m, m = 1, \cdots, M$ that are considered independent, the model would be a product of Poisson distributions. But, why is a Poisson appropriate? Let us start at the very beginning.

### 2.1.1 Bernoulli trial

A Bernoulli trial, named after the Swiss mathematician Jacob Bernoulli (1654 − 1705), is an experiment with only two possible outcomes: $S$, a success or $F$, a failure. Each collision between protons at the LHC is a Bernoulli trial in which either a Higgs boson is created ($S$) or is not ($F$). Here is a sequence of collisions results

$$F \quad F \quad S \quad F \quad F \quad F \quad F \quad S \quad F \quad \cdots$$

What is the probability of this sequence of results? There is *no* answer. Unless that is we are prepared to make assumptions, such as the following.

1. Let $p$ be the probability of a success.
2. Let $p$ be the same for every collision (trial).
3. Let $S$ and $F$ be *exhaustive* (the only possible outcomes) and *mutually exclusive* (one outcome precludes the occurrence of the other).

Assumption 3 implies that the probability of $F$ is $1 - p$. Therefore, for a given sequence $O$ of $n$ proton-proton collisions, the probability $P(k|n, p, O)$ of exactly $k$ successes and exactly $n - k$ failures is

$$P(k|n, p, O) = p^k(1 - p)^{n-k}. \tag{6}$$

The specific sequence $O$ of successes and failures is unknown at the LHC. Whenever, we have a parameter that is either irrelevant or whose value is unknown, the rules of probability theory imply that the unknown can be eliminated from the problem by summing over all possible values of the unknown, here the orders of successes and failures $O$. This rule is called **marginalization** and is one of the most important procedures in probability calculations. Applied to our problem this yields,

$$P(k|n, p) = \sum_O P(k|n, p, O) = \sum_O p^k(1 - p)^{n-k}. \tag{7}$$

Notice that every term in Eq. (7) is identical and there are $\binom{n}{k}$ of them. Therefore,

$$P(k|n, p) = \binom{n}{k} p^k(1 - p)^{n-k}, \tag{8}$$

that is, we arrive at the **binomial distribution**, Binomial(k, n, p). If $a$ is the mean number of successes in $n$ trials, then

$$a = \sum_{k=0}^{n} k \, \text{Binomial(k, n, p)},$$

$$= pn. \tag{9}$$

Exercise 4: Show this

For the Higgs boson outcomes, $p \sim 10^{-10}$ and $n \gg 10^{12}$. Therefore, it is reasonable to consider the limit $p \to 0$ and $n \to \infty$, while keeping $a$ constant. In this limit

$$\text{Binomial(k, n, p)} \to e^{-a} a^k / k!,$$

$$\equiv \text{Poisson}(k, a). \tag{10}$$

Exercise 5: Show this

We conclude that a Poisson distribution is an appropriate model when the probability of individual events is extremely small. Indeed, the distribution can be derived from a stochastic model in which that assumption is made explicit. Therefore, it is indeed reasonable to take

$$p(n|s, b) = \text{Poisson}(n, s + b) = \frac{(s+b)^n e^{-(s+b)}}{n!}, \tag{11}$$

as the probability to obtain a count $n$ given mean event count $s + b$.

We now turn to the probability model for the background data. In principle, the model should encode in detail how the background estimate was obtained. But, in order to keep matters as simple as possible, let us assume that the background estimate was obtained from an accurate Monte Carlo simulation, which yields a count $m$. The mean count in the simulation is $kb$, where $k$ is a known scale factor that relates the mean count in the simulation to that in the signal region of the experiment that yielded $N$ events. Therefore, the probability model for the background shall be taken to be

$$p(m|kb) = \text{Poisson}(m, kb). \tag{12}$$

Since the counts $n$ and $m$ are independent, the full model is

$$p(n, m|s, b) = \text{Poisson}(n, s + b)\text{Poisson}(m, kb). \tag{13}$$

## 2.2 The Likelihood Function

The **likelihood function** is the probability function—either a probability density function (pdf) if the random variables are continuous, or a probability mass function (pmf) if they are discrete—into which observations, that is, data have been inserted. Since the data are constants, the likelihood, $p(N, M|s, b)$ in our example, is a function of the parameters only. Sometimes, $p(N, M|s, b)$ is written as $L(s, b)$ to emphasize this point.

In this example, we are given $B \pm \delta B$, not $M$ and $k$. But, we can infer $M$ and $k$ from $B$ and $\delta B$ using a plausible model, namely, that $B$ and $\delta B$ are $M$ and $\sqrt{M}$ scaled down by k, that is,

$$B = M/k, \tag{14}$$

$$\delta B = \sqrt{M}/k. \tag{15}$$

Inverting these equations yields

$$M = (B/\delta B)^2 = 353.4, \tag{16}$$

$$k = B/\delta B^2 = 37.6. \tag{17}$$

Therefore, the likelihood for the count $M$ is

$$(kb)^M e^{-kb}/\Gamma(M+1), \tag{18}$$

which we have written in a form that allows for non-integral values of $M$. Writing $D = N, M$, the full likelihood can be written as

$$p(D|s,b) = \frac{(s+b)^N e^{-(s+b)}}{N!} \frac{(kb)^M e^{-kb}}{\Gamma(M+1)}. \tag{19}$$

In a more realistic analysis, a probability model for the scale factor $k$ would also be included. But, to keep things simple, we shall neglect the uncertainty in $k$.

Now that we have the likelihood function, several questions can be answered, including the following.

1. How is a parameter to be estimated?
2. How is its accuracy to be quantified?
3. How can an hypothesis be tested?
4. How is the statistical significance of a result to be quantified?

### 2.3 The Frequentist Principle

The goal of a frequentist analysis is to construct statements such that it can be guaranteed, *a priori*, that a fraction $f \geq p$ of them are true over an ensemble of similarly constructed statements. This stipulation is called the **frequentist principle** (FP) and was championed by the Polish statistician Jerzy Neyman [8]. The fraction $f$ is called the **coverage probability**, or coverage for short, and $p$ is called the **confidence level** (C.L.). An ensemble of statements that obey the frequentist principle is said to *cover*.

*Points to Note*

1. The FP applies to real ensembles [3], not just the virtual ones we simulate on a computer. Moreover, the ensembles can contain statements about different quantities. *Example*: all published measurements $x$, since the discovery of the electron in 1897, of the form $\theta \in [l(x), u(x)]$, where $\theta$ is a parameter of interest, that is, the parameter to be measured.
2. Coverage is an *objective* characteristic of ensembles of statements. However, in order to verify whether an ensemble of statements covers, we need to know which statements are true and which ones are false. Alas, since this information is generally not available in the real world there is no *operational* way to compute the coverage. The fact that we can do so in a simulation may give us confidence that the actual coverage of published statements is as the simulation reports, but does not prove that it is so.
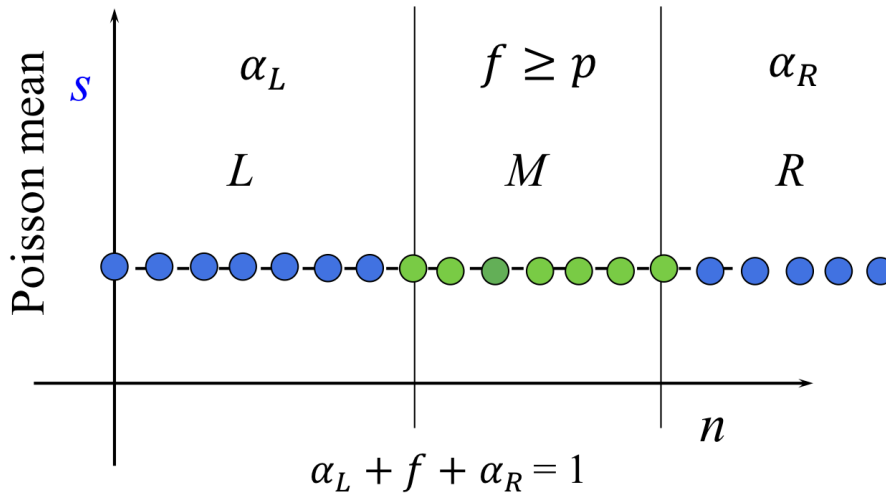
*Example*

Consider an ensemble of different experiments, each with a different mean count $\theta$, and each yielding a count $N$. Each experiment makes a single statement of the form

$$N + \sqrt{N} > \theta,$$

---

[3]Strictly speaking, we mean real samples because, as we have defined it, an ensemble is a synonym for a population, which by definition contains infinitely many elements

**Fig. 1:** Plotted is the tensor product of the parameter space, with parameter $s$, and the space of observations with potential observations $n$. For a given value of $s$, the observation space is partitioned into three disjoint intervals, labeled $L$, $M$, and $R$, such that the probability to observe a count $n$ in $M$ is $f \geq p$, where $p =$ is the desired confidence level.

which is either true or false. If these were real experiments, we would not be able to determine which statements are true and which are false and, therefore, determine the coverage. Suppose that each mean count $\theta$ is randomly sampled from `uniform(0, 10)`, with range [0, 10], and suppose that these means are known as would be the case in a simulation. Since the numbers are known, we can compute the coverage probability $f$.

**Exercise 7:** Compute the coverage of these statements; repeat the exercise using `uniform(0, 1000)`

In the next section, we discuss the important concept of the confidence interval, which is the classic exemplar of the frequentist principle.

## 2.4 Confidence Intervals

In 1937, Neyman [8] introduced the concept of the **confidence interval**, a way to quantify uncertainty in estimates that respects the frequentist principle. Confidence intervals are a concept best explained through an example. Consider an experiment that observes $n = N$ events with mean signal count $s$ and no background. A confidence interval $[l(N), u(N)]$, with confidence level CL $= p$, permits a statement of the form

$$s \in [l(N), u(N)], \tag{20}$$

with the *a priori* guarantee that a fraction $f \geq p$ of statements will be true over an ensemble of such statements, not necessarily about the same quantity or the same kind of experiment. For simplicity, however, we shall consider experiments of the same kind, but which differ by their mean signal count $s$.

Consider Fig. 1, which shows the tensor product of the parameter space $\{s\}$ and the space of potential observations $\{N\}$ as well as the potential observations, represented by the dots, of an experiment with mean count $s$. The two vertical lines divide the space of observations into the three regions labeled $L$, $M$, and $R$. The region $M$ is chosen so that the probability to obtain a count in that region is $f \geq p$, where $p$ is the desired confidence level (CL). The probabilities to obtain a count in region $L$ or region $R$ are $\alpha_L$ and $\alpha_R$, respectively. Since the three regions span the space of observations, $\alpha_L + f + \alpha_R = 1$.
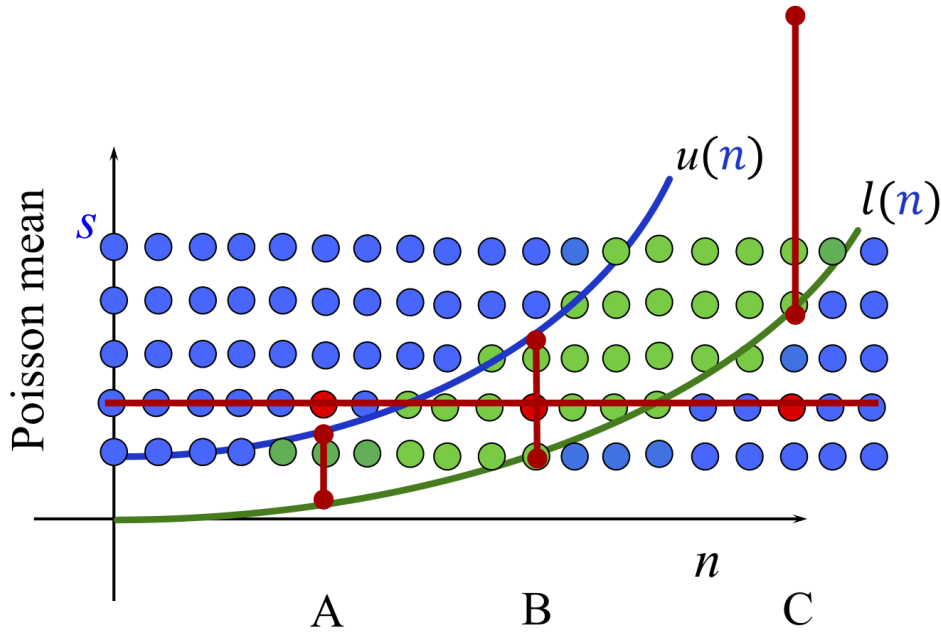
**Fig. 2:** The algorithm for defining region $M$ (see Fig. 1), must be repeated for every value of $s$ that is possible *a priori*. For the experiment whose mean $s$ is represented by the thick horizontal line, the figure shows three possible outcomes, labeled A, B, and C, and their associated confidence intervals $[l(n), u(n)]$. Only outcomes, such as B, which lie within the region $M$ of the experiment will yield intervals that bracket $s$. The probability to obtain such an interval is $f \geq p$, by construction.

For a given coverage $f$, the choice of region $M$ is not unique and different methods have been suggested to define it. The first method was devised by Neyman [8], which we shall consider shortly. Another method was suggested by Feldman and Cousins [9]. We shall use that method to explain the general construction of confidence intervals.

*Feldman-Cousins Method*

In the Feldman-Cousins method, every potential count $n$ is associated with a pair of numbers: a weight $p(n|s) \, / \, p(n|\hat{s})$, where $\hat{s} = n$ is the maximum likelihood estimate of $s$, together with the probability $p(n|s)$ to obtain the count $n$. The counts are placed in *descending* order of their weights. Starting with the first count in the ordered list, a set of counts $(n_{(1)}, n_{(2)}, \cdots)$ is accumulated one by one until their summed probabilities $f = \sum_{(i)} p(n_{(i)}|s) \geq p$. The symbol $(i)$ denotes the ordinal value of a count in the ordered list. The set of counts $(n_{(1)}, n_{(2)}, \cdots)$ defines an interval in the space of observations whose lowest (leftmost) and highest (rightmost) counts $n_L$ and $n_R$ are given by $n_L = \min(n_{(1)}, n_{(2)}, \cdots)$ and $n_R = \max(n_{(1)}, n_{(2)}, \cdots)$, respectively. This construction (for this single parameter problem) guarantees that the probability to obtain a count within region $M$ is $f \geq p$ [4].

There is, however, a snag with any algorithm to define $M$. The latter can only be defined if the mean count associated with an experiment is known. This may well be true within a simulation, but it is not so in the real world. Therefore, any algorithm for defining the region $M$ must be repeated for every value of $s$ that is considered possible *a priori*, as illustrated in Fig. 2. The repetition produces regions $M_s$, labeled by the mean count $s$, that define two curves, labeled $l(n)$ and $u(n)$, in the product space $\{s\} \otimes \{n\}$. For a given $n$, these curves define the confidence intervals $[l(n), u(n)]$. Over an ensemble

---

[4]We write $f \geq p$ rather than $f = p$ because, in general, for a discrete distribution it is not possible to satisfy the equality except at specific values of $s$.
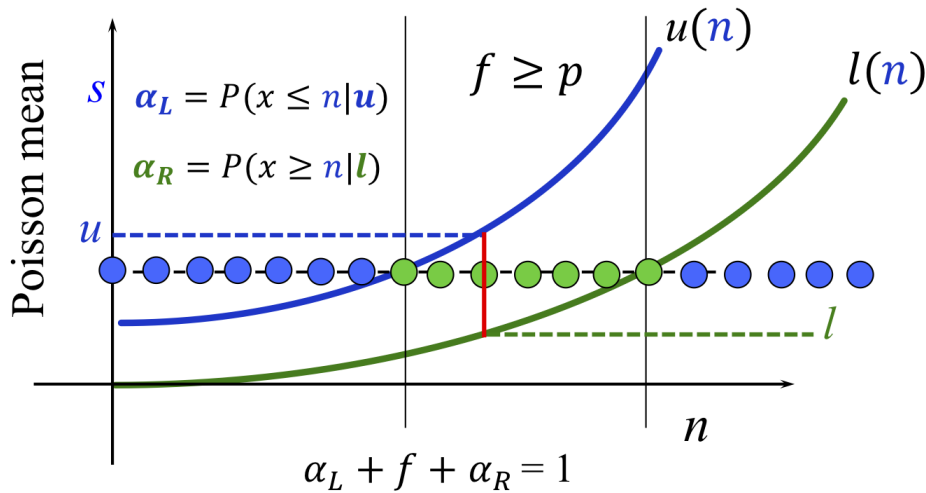
**Fig. 3:** The Neyman method. For every $n$, an interval $[l(n), u(n)]$ is computed by solving the equations in the plot. See text for details.

of experiments—and irrespective of their associated mean count $s$, the fraction of statements of the form $s \in [l(n), u(n)]$ that are true is $f \geq p$, by construction. To see this, consider again Fig. 2. It shows three possible outcomes for the experiment defined by the thick horizontal line together with the three possible confidence intervals (the vertical lines terminated with dots). If an observation lands in the region $M$ for that experiment, the interval $[l(n), u(n)]$ will bracket the mean count $s$, as shown in the figure. If a count lands in region $L$, then the upper limit $u(n)$ will lie below $s$ and, consequently, the interval $[l(n), u(n)]$ will exclude $s$. If $n$ lands in region $R$, then the lower limit $l(n)$ will lie above $s$ and the interval will exclude $s$. Therefore, the interval $[l(n), u(n)]$ will include $s$ only if $n$ lies in $M$, for which the probability is $f \geq p$. A procedure for constructing confidence intervals in this manner is called a **Neyman construction**.

*Neyman Method*

The algorithm described above requires that a region $M$ be constructed for each value of $s$. An alternative algorithm was devised by Neyman in his 1937 paper and is illustrated in Fig. 3. For every $n$, the upper and lower limits are found by solving

$$P(x \leq n|u) = \alpha_L, \tag{21}$$
$$P(x \geq n|l) = \alpha_R. \tag{22}$$

Equation (21) yields a curve $u(n)$ for which the probability to obtain a count $x \leq n$, for a given $s$, is $\alpha_L$, while Eq. (22) yields a curve $l(n)$ for which the probability to obtain a count $x \geq n$, for a given $s$, is $\alpha_R$. These curves can also be made using the Neyman construction described above for the Feldman-Cousins method, but the solution using Eqs. (21) and (22) is computationally more efficient. Figure 4 shows the coverage probability over the parameter space for the Neyman intervals, in which we have chosen $\alpha_L = \alpha_R = (1 - p)/2$. This choice, the one made by Neyman, define **central confidence intervals**. As advertised, these confidence intervals satisfy the frequentist principle. Also shown is the coverage for intervals of the form $[N - \sqrt{N}, N + \sqrt{N}]$ and $[N - \sqrt{N}, N + \sqrt{N} + \exp(-N)]$. These intervals are *approximate* confidence intervals in that they do not satisfy the frequentist principle exactly. Notice, however, that for $s > 2.5$ the coverage of these intervals bounces around the $p = 0.683$ line. Therefore, over a large sample of experiments, with a distribution of Poisson means, it is plausible that the coverage could turn out to be close to the desired confidence level.
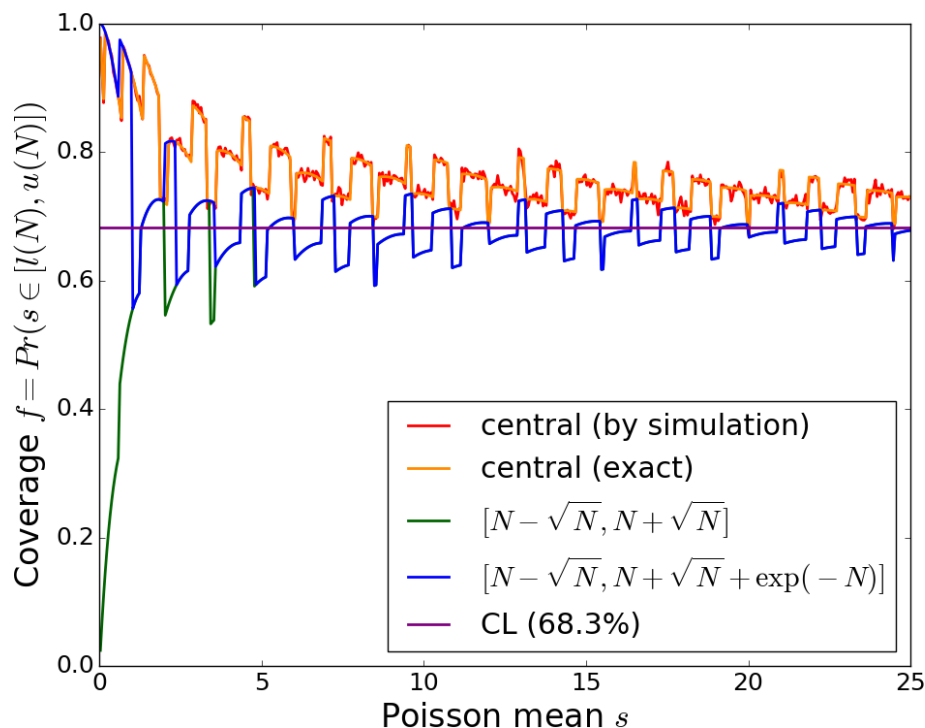
**Fig. 4:** Coverage probability $f$ as a function of the Poisson mean $s$. As expected, the central intervals satisfy the frequentist principle, namely, $f \geq p$, where $p = 0.683$ is the confidence level. The coverage for two other sets of intervals are shown for which the frequentist principle is not satisfied.

A notable feature of Fig. 4 is the jaggedness of the coverage probabilities over the parameter space. The jaggedness is caused by the discreteness of the Poisson distribution. For a discrete distribution, coverage equal to the desired confidence level is possible only at specific values of $s$. Therefore, if we insist on the frequentist principle, $f \geq p$, the price to be paid is *over-coverage* in subsets of the parameter space.

## 2.5   The Profile Likelihood

The likelihood function,

$$p(D|s,b) = \frac{(s+b)^N e^{-(s+b)}}{N!} \frac{(kb)^M e^{-kb}}{\Gamma(M+1)}, \tag{23}$$

contains *two* parameters, the mean signal count $s$ and mean background count $b$. However, the **parameter of interest** is the mean signal. The mean background count is needed to define the probability model, but inferences about it are not of interest. The parameter $b$ is an example of a **nuisance parameter**. One way or another, we must rid a probability model of *all* nuisance parameters if we wish to make inferences about the parameter(s) of interest, here the mean Higgs boson signal count $s$. A widely accepted method for doing so is to convert the likelihood function into a function called the profile likelihood. But, before discussing this, we briefly describe the most common frequentist method to arrive at estimates of parameters.

Given the likelihood function $L(s,b) \equiv p(D|s,b)$, its parameters can be estimated by maximizing $L(s,b)$, or, equivalently, maximizing $\ln L(s,b)$, with respect to $s$ and $b$,

$$\frac{\partial \ln p(D|s,b)}{\partial s} = 0 \quad \text{leading to } \hat{s} = N - B,$$

$$\frac{\partial \ln p(D|s,b)}{\partial b} = 0 \quad \text{leading to } \hat{b} = B,$$

as expected. Estimates found this way (first done by Karl Gauss and systematically developed by Fisher [10]) are called **maximum likelihood estimates** (MLE). This method generally leads to reasonable estimates, but, as is true of other procedures in statistical analysis, the method has its good and bad features, as noted below.

– *The Good*

– Maximum likelihood estimates are *consistent*, that is, the RMS of estimates goes to zero as more and more data are included in the likelihood. This basically says that acquiring more data makes sense because the accuracy of results is expected to improve.

– If an *unbiased* estimate of a parameter exists, the maximum likelihood procedure will find it.

– Given the MLE for $s$, the MLE for any function $y = g(s)$ of $s$ is $\hat{y} = g(\hat{s})$. This useful feature means that it possible to maximize the likelihood using any parameterization of it, say $s$, because, at the end, we can transform to the parameter of interest using $\hat{y} = g(\hat{s})$.

– *The Bad*

– In general, MLEs are biased.

> **Exercise 7:** Show this
> Hint: Taylor expand $\hat{y} = g(s + \hat{s} - s)$ about $s$ and consider its ensemble average.

– *The Ugly*

– Most MLEs are biased, which, unfortunately, encourages the routine application of bias correction. But correcting for bias only makes sense if the RMS of an unbiased result is less than or equal to the RMS of a biased result. Recall that the RMS $= \sqrt{V + b^2}$, where $V$ is the variance and $b$ is the bias.

We now return to the profile likelihood. In order to make an inference about the signal, $s$, the 2-parameter model $L(s,b)$ must be reduced to one involving $s$ only. In principle, this must be done while respecting the frequentist principle, that is, $f \geq p$, where $f$ is the coverage probability of an ensemble of statements and $p$ is the desired confidence level. In practice, all nuisance parameters are replaced by their MLEs conditional on given values of the parameters of interest. For the Higgs boson example, an estimate of $b$ is found as a function of $s$, $\hat{b} = f(s)$, and $b$ is replaced by $\hat{s}$ in $L(s,b)$. This leads to a function $L_p(s) = L(s, f(s))$ called the **profile likelihood**. For the likelihood in Eq. (23),

$$\hat{b} = f(s) = \frac{g + \sqrt{g^2 + 4(1+k)Ms}}{2(1+k)},$$

$$\text{where } g = N + M - (1+k)s. \tag{24}$$

Figure 5 shows a density plot of the likelihood $L(s,b)$ with the function $\hat{b} = f(s)$ superimposed. Notice that $\hat{b} = f(s)$ goes through the mode of $L(s,b)$, which occurs at $s = \hat{s} = N - B = 15.6$ events. Figure 6 shows the profile likelihood.

Replacing the (unknown) true value of $b$ with an estimate thereof is clearly an approximation. Therefore, it should come as no surprise that inferences based on the profile likelihood are not guaranteed to be satisfy the frequentist principle exactly. However, it is found that for the typical applications in particle physics (as will be evident below), the procedures based on the profile likelihood work surprisingly well. Moreover, the use of the profile likelihood has a sound theoretical justification.
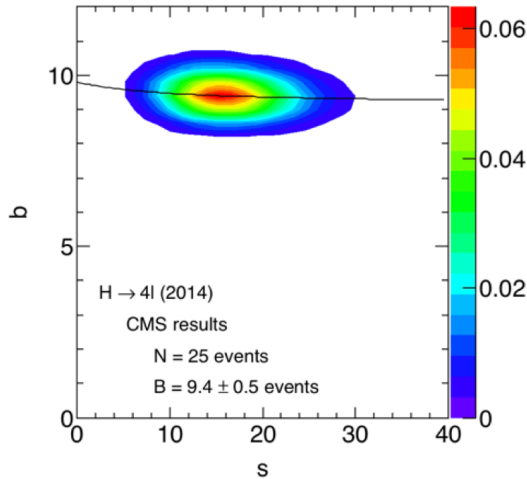
**Fig. 5:** The likelihood $L(s, b)$ and the graph of the function $\hat{b} = f(s)$.
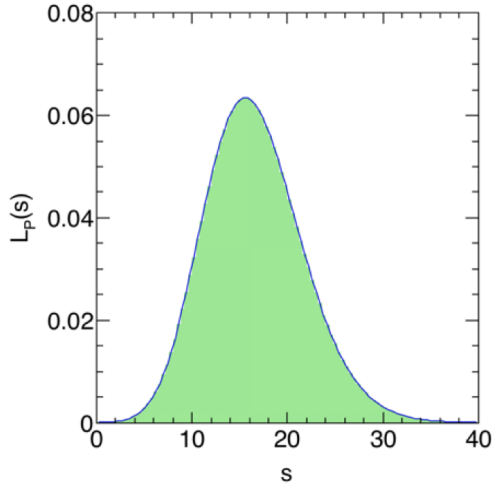


**Fig. 6:** The profile likelihood $L_p(s) \equiv L(s, f(s))$.

Consider the profile likelihood ratio

$$\lambda(s) = L_p(s)/L_p(\hat{s}), \qquad (25)$$

where $\hat{s}$ is the MLE of $s$. Taylor expand the associated quantity

$$t(s) = -2\ln\lambda(s) \qquad (26)$$

about $\hat{s}$,

$$t(\hat{s} + s - \hat{s}) = t(\hat{s}) + t'(\hat{s})(s - \hat{s})$$
$$+ t''(\hat{s})(s - \hat{s})^2/2 + \cdots$$
$$\approx (s - \hat{s})^2/2/\sigma^2 + \cdots,$$
$$\text{where } \sigma^2 \approx 2/t''(\hat{s}). \qquad (27)$$

The quadratic approximation is called the Wald approximation (1943) (see Cowan et al. [11]). If $\hat{s}$ does not occur on the boundary of the parameter space (in which case the derivative of $t$ at $\hat{s}$ is zero), the sample is large enough (that is, when the density of $\hat{s}$ is approximately Gaussian$(\hat{s}, s, \sigma)$), and if $s$ is the true value of the signal, then the density of $t(s)$ converges to a $\chi^2$ density of one degree of freedom. The result, which is important because of its generality, is a special case of Wilks' theorem (1938) (Cowan et al. [11]).

Since $t(s) \approx \chi^2$, we can compute an approximate 68% confidence interval by solving

$$t(s) = -2\ln\lambda(s) = 1, \qquad (28)$$

for the lower and upper limits of the interval. Given $N = 25$ observed 4-lepton events, a background estimate of $B \pm \delta B = 9.4 \pm 0.5$, we can state that

$$s \in [10.9, \ 21.0] \quad @ \ 68\% \ \text{C.L.} \qquad (29)$$

**Exercise 8:** Verify this interval.

As noted, intervals constructed using the profile likelihood are not guaranteed to satisfy the frequentist principle. However, for applications in particle physics the coverage of these intervals is usually very good even for small amounts of data.

## 2.6 Hypothesis Tests

In the previous section, we concluded that $s \in [10.9, 21.0]$ @ 68% C.L. This result strongly suggests that a signal exists in the $N = 25$ 4-lepton events observed by CMS. But, a qualitative statement such as this is generally considered insufficient. The accepted practice is to perform an hypothesis test. Indeed, in particle physics, a discovery is declared only if a certain quantitative threshold has been reached in an hypothesis test.

An hypothesis test, in the frequentist approach, is a procedure for *rejecting* an hypothesis, which adheres to the following protocol.

1. Decide which hypothesis is to be *rejected*. This is called the **null hypothesis**. At the LHC, this is usually the background-only hypothesis.

2. Construct a function of the data called a **test statistic** with the property that large values of it would cast doubt on the veracity of the null hypothesis.

3. Choose a test statistic threshold above which we are inclined to reject the null. Do the experiment, compute the statistic, and reject the null if the threshold is breached.

We consider two related variants of this protocol, one by Fisher [10] and the other by Neyman, both developed in the 1930s. Fisher and Neyman disagreed strenuously about hypothesis testing, which suggests that the topic is rather more subtle than it seems. Fisher held that an hypothesis test required consideration of the null hypothesis only, while Neyman argued that a proper test required consideration of both a null as well as an alternative hypothesis. Physicists ignore these disagreements and see utility in an amalgam of the approaches of Fisher and Neyman. The is eminently sensible and pragmatic, whereas our quasi-religious adherence to a $5\sigma$ threshold before declaring a discovery is not always sensible.

We first illustrate Fisher's theory of hypothesis testing and follow with a description of Neyman's theory.

*Fisher's Approach*

We take the null hypothesis, which is denoted by $H_0$, to be the background-only model, that is, the Standard Model without a Higgs boson and compute a measure of the incompatibility of $H_0$ with the observations, called a **p-value**, defined by

$$\text{p-value}(x_0) = P(x > x_0|H_0), \quad (30)$$

where $x$ is a test statistic, designed so that large values indicate departure from the null hypothesis, and $x_0$ is the observed value of the statistic. Figure 7 shows the location of $x_0$. The p-value is the probability that $x$ could have been higher than the $x_0$. Fisher argued that a sufficiently small p-value implies that either the null hypothesis is false or something rare has occurred. If the p-value is extremely small, say $\sim 3 \times 10^{-7}$, then of the two possibilities the response of the particle physicist is to reject the null hypothesis and declare that a discovery has been made. The p-value for our example, neglecting the uncertainty in the background estimate, is

$$\text{p-value} = \sum_{k=N}^{\infty} \text{Poisson}(k, 9.4) = 1.76 \times 10^{-5}, \text{ with } N = 25.$$

Since the p-value is a bit non-intuitive, it is conventional to map it to a $Z$-**value**, that is, the number of standard deviations the observation is *away from the null* if the distribution were a Gaussian. The $Z$-value can be computed using [5].
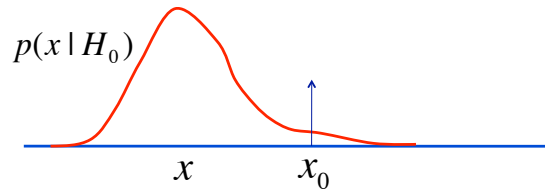
$$Z = \sqrt{2}\,\text{erf}^{-1}(1 - 2\text{p-value}). \quad (31)$$

A p-value of $1.76 \times 10^{-5}$ corresponds to a $Z$ of $4.14\sigma$. The $Z$-value can be calculated using the `Root` function

$$Z = \texttt{TMath::NormQuantile(1-p-value)}.$$

If the p-value is judged to be small enough, or the $Z$-value is large enough, the background-only hypothesis is rejected.

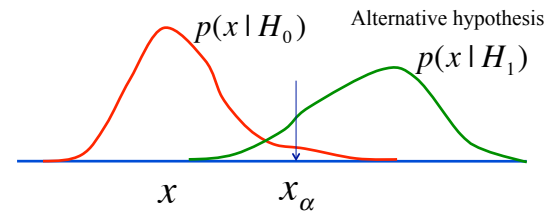**Fig. 7:** The p-value is the tail-probability, $P(x > x_0|H_0)$, calculated from the probability density under the null hypothesis, $H_0$. Consequently, the probability density of the p-value under the null hypothesis is uniform$(0, 1)$.

---

[5] $\text{erf}(x) = \frac{1}{\sqrt{\pi}}\int_{-x}^{x}\exp(-t^2)\,dt$ is the error funtion.
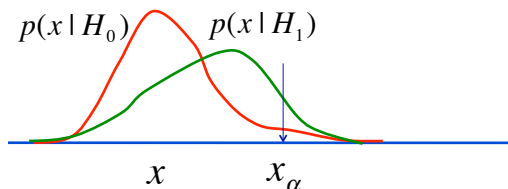
*Neyman's Approach*

As noted, Neyman insisted that a correct hypothesis test required *two* hypotheses to be considered, the null hypothesis $H_0$ and an alternative hypothesis $H_1$. This is illustrated in Fig. 8. The null is the same as before but the alternative hypothesis is the Standard Model with a Higgs boson, that is, the background plus signal hypothesis. Again, the statistic $x$ is constructed so that large values would cast doubt on the validity of $H_0$. However, the Neyman test is specifically designed to respect the frequentist principle. A *fixed* probability $\alpha$ called the **significance (or size) of the test** is chosen, which corresponds to some threshold value $x_\alpha$ defined by



**Fig. 8:** Distribution of a test statistic $x$ for two hypotheses, the null $H_0$ and the alternative $H_1$. In Neyman's approach to testing, $\alpha = P(x > x_\alpha | H_0)$ is a *fixed* probability called the significance of the test, which for a given class of experiments corresponds the threshold $x_\alpha$. The hypothesis $H_0$ is rejected if $x > x_\alpha$.

$$\alpha = P(x > x_\alpha | H_0). \tag{32}$$

Should the observed value $x_0 > x_\alpha$, or equivalently, p-value($x_0$) $< \alpha$, the hypothesis $H_0$ is rejected in favor of the alternative. By construction, a repeated application of this test will reject a fraction $\alpha$ of true null hypotheses. Since these are false rejections, we say that these are **Type I errors**. Neyman's test discards the p-value and reports only $\alpha$ and whether or not the null was rejected. However, in particle physics, in addition to reporting the results of the test, perhaps announcing a discovery, we also report the observed p-value. This makes sense because there is a more information in the p-value than merely reporting the fact that a null hypothesis was rejected at a significance level of $\alpha$.



**Fig. 9:** See Fig. 8 for details. Unlike the case in Fig. 8, the two hypotheses $H_0$ and $H_1$ are not that different. It is then not clear whether it makes practical sense to reject $H_0$ when $x > x_\alpha$ only to replace it with an hypothesis $H_1$ that is not much better.

Given that Neyman's test requires an alternative hypothesis there is more that can be said than simply reporting the result of the test and the observed p-value. Figure 8 shows that we can also calculate

$$\beta = P(x \leq x_\alpha | H_1), \tag{33}$$

which is the relative frequency with which we reject true alternative hypotheses $H_1$. This mistake is called a **Type II error**. The quantity $1 - \beta$ is called the **power** of the test and is the relative frequency with which we would accept the true alternative hypotheses. The defining feature of the Neyman test is that, in accordance with the Neyman-Pearson lemma (see for example Ref. [2]), the power is maximized subject to the constraint that $\alpha$ is fixed. The Neyman-Pearson lemma asserts that given two simple hypotheses—that is, hypotheses in which all parameters have specified values—the optimal test statistic $t$ for conducting an hypothesis test is the likelihood ratio $t = p(x|H_1)/p(x|H_0)$.

Maximizing the power seems like a reasonable procedure. Consider Fig. 9, which shows that the significance of the test in this figure is the same as that in Fig. 8. Therefore, the Type I error rates are identical. However, the Type II error rate is much greater in Fig. 9 than in Fig. 8 because the power of the test is considerably weaker in the former. Consequently, it is debatable whether rejecting the null is a wise course of action since the alternative hypothesis is not that much better. This insight was one source of Neyman's disagreement with Fisher. Neyman objected to the possibility that one might reject a null hypothesis regardless of whether it made sense to do so. He argued that the goal of hypothesis

testing is always one of deciding between competing hypotheses. Fisher's counter argument was that an alternative hypothesis may not be available, in which case we either give up or we have a method to test the only hypothesis that is available in order to decide whether it is worth keeping. In a Bayesian analysis an alternative hypothesis is also needed, in agreement with Neyman viewpoint, but is used in a way that neither he nor Fisher agreed with.

So far we have assumed that the hypotheses $H_0$ and $H_1$ are simple, that is, fully specified. Alas, most of the hypotheses that arise in realistic particle physics analyses are not of this kind. In the Higgs boson example, the probability models depend on a nuisance parameter for which only an estimate is available. Consequently, neither the background-only nor the background plus signal hypotheses are fully specified. Such hypotheses are examples of **compound hypotheses**. In the following, we illustrate how hypothesis testing proceeds in this case using the 4-lepton example.

*Compound Hypotheses*

In Sec. 2.5, we reviewed the standard way nuisance parameters are handled in a frequentist analysis, namely, their replacement by their conditional MLEs, thereby converting the likelihood function to the profile likelihood. In the 4-lepton example, this yielded the function $L_p(s) = L(s, f(s))$. The justification for this is that the statistic $t(s) = \ln \lambda(s)$, where $\lambda(s) = L_p(s)/L_p(\hat{s})$ and $\hat{s}$ is the MLE of $s$ can be used to compute (approximate) confidence intervals in light of Wilks' theorem, which essentially states that $t(s) \approx \chi^2$. Therefore, the same statistic can also be used as a test statistic with the associated p-values calculated using the $\chi^2$ density. Moreover, since, by definition, $Z = \sqrt{\chi^2}$, the p-value calculation can be sidestepped altogether. Using $N = 25$ and $s = 0$, we find $\sqrt{t(0)} = 4.13$, which is to be compared with $Z = 4.14$, the value found neglecting the $\pm 0.5$ event uncertainty in the background.

In summary, the statistic $t(s)$ can be used to test null hypotheses as well as compute confidence intervals and, therefore, provides a unified way to deal with both tasks. If $s$ is the true value of the mean signal, then the distribution of $t(s)$ under that hypothesis is a $\chi^2$ density with one degree of freedom, $p(\chi^2|ndf = 1)$. Sometimes, however, it is necessary to consider $t(s)$ when the value of $s$ in the argument differs from the value $s$, say $s_0$, which determines the density of $t(s)$. For example, suppose that a model of new physics predicts a mean count $s_0$ and an analysis is planned to test this model. We may be interested to know, for example, what value of $t(s)$ we might expect for a given amount of data. If $s = 0$, the goal may be to determine the average or median significance with which we may be able to reject the background-only hypothesis. Since the predicted signal $s_0$ differs from $s = 0$, the density of $t(s, \hat{s})$—where for clarity, the dependence on the estimate $\hat{s}$ is made explicit—will no longer be $\chi^2$, but rather a non-central $\chi^2$ density, $p(\chi^2|ndf = 1, nc)$ with non-centrality parameter $nc$, an approximate value for which is $nc = t(s, s_0)$; that is, it is the test statistic computed using an **Asimov**[6] data set [11] in which the "observed" count $N$ is set equal to the true mean signal count, $s_0 + b$.

## 3   Bayesian Analysis

Bayesian analysis is merely applied probability theory with the following significant twist: a method is Bayesian if

- it is based on the degree of belief interpretation of probability and
- it uses Bayes' theorem

$$p(\theta, \omega|D) = \frac{p(D|\theta, \omega)\, \pi(\theta, \omega)}{p(D)},$$ (34)

---

[6]The name of this special data set is inspired by the short story *Franchise* by Isaac Asimov describing a futuristic United States in which, rather than having everyone vote in a general election, a single (presumably representative) person is chosen to answer a series of questions whose answers are analyzed by an AI system. The AI system then decides the outcome of the election by determining what would have been the outcome had the general election been held!

where

$$D = \text{observed data},$$
$$\theta = \text{parameters of interest},$$
$$\omega = \text{nuisance parameters},$$
$$p(D|\theta, \omega) = \text{likelihood},$$
$$p(\theta, \omega|D) = \text{posterior density},$$
$$\pi(\theta, \omega) = \text{prior density},$$

for *all* inferences. The posterior density is the final result of a Bayesian analysis from which, if desired, various summaries can be extracted. The posterior density assigns a weight to every hypothesis about the values of the parameters of the probability model, which, in addition to the likelihood, also includes a function called the prior density or **prior** for short. The parameters can be discrete, continuous, or both, and nuisance parameters are eliminated by marginalization,

$$p(\theta|D) = \int p(\theta, \omega|D) \, d\omega, \tag{35}$$
$$\propto \int p(D|\theta, \omega) \, \pi(\theta, \omega) \, d\omega.$$

The prior $\pi(\theta, \omega)$ encodes whatever assumptions we make and information we have about the parameters $\theta$ and $\omega$ independently of the data $D$. A key feature of the Bayesian approach is recursion: the use of the posterior density $p(\theta, \omega|D)$ as the prior in a subsequent analysis.

These rules are simple, yet they yield an extremely powerful and general inference algorithm. However, particle physicists remain wedded to the frequentist approach because of the still widespread perception that the Bayesian algorithm is too subjective to be useful for scientific work. However, there is considerable published evidence to contrary, including in particle physics, witness the successful use of Bayesian analysis in the discovery of single top quark production at the Tevatron [16, 17] and searches for new physics at the LHC [18–20].

So, why do particle physicists, for the most part, remain skeptical about Bayesian analysis? For many, the Achilles heel of the Bayesian approach is the difficulty of specifying a believable prior over the parameter space of the likelihood function. In our example, in order to make an inference about the mean event count $s$ using the data $N = 25$ events with a background of $B \pm \delta B = 9.4 \pm 0.5$ events, a prior density $\pi(s, b)$ must be constructed. Even after more than two centuries of effort, discussion, and argument, however, statisticians have failed to reach a consensus about how to do this in the general case. Nevertheless, Bayesian analysis is widely and successfully used, and used even within particle physics. This strongly suggests that we should refrain from overstating the difficulties. After all, physics is replete with approximations, both of a technical and conceptual nature. The same is true of statistical analysis. But, of course, this is no excuse for sloppiness. Rather it is a reminder not to make perfection the enemy of the good.

The particle physicists who have given this topic some thought seem to agree with the statisticians who argue that the following invariance property should hold for any prior, at least ideally,

$$\pi_\phi(\phi)d\phi = \pi_\theta(\theta)d\theta, \tag{36}$$

where $\phi = f(\theta)$ is a one-to-one mapping of the parameter vector $\theta$, e.g., $\theta = (s, b)$, to the new parameter vector $\phi$ and $\pi_\phi$ and $\pi_\theta$ are, in general, different functions of their arguments. If the above invariance holds, then the posterior density will likewise be reparametrization invariant in the same sense as the prior. Suppose we have a rule for creating a prior $\pi(*)$ and we apply this rule to create the density $\pi_\phi$. The same rule is now used to create $\pi_\theta$ after which we transform from $\pi_\theta(\theta)d\theta$ to $\pi(\phi)d\phi$. Invariance with respect to the choice of parametrization demands that $\pi = \pi_\phi$. It surely ought not to matter whether

we parametrize the likelihood $p(D|s,b)$ in terms of $s$ and $b$ or in terms of $s$ and $u = \sqrt{b}$. After all, the likelihood hasn't really changed, therefore, it would be odd if this "non-change" changed the posterior density. But, whether or not a change occurs depends on the nature of the prior, as the following example shows.

Consider the probability function $p(D|s) = \text{Poisson}(D|s)$, written in two different ways: $p(D|s) = \exp(-s)s^D/D!$ and $p(D|\sigma) = \exp(-\sigma^2)\sigma^{2D}/D!$, where $\sigma = \sqrt{s}$. In order to compute the posterior densities $p(s|D)$ and $p(\sigma|D)$ priors must be specified. The most widely used rule for doing so is: choose the prior to be flat, that is, uniform, e.g., $\pi(s) = 1$ and $\pi(\sigma) = 1$ in the parameter space. Notice that for an unbounded parameter space $\int \pi(s)\,ds = \int \pi(\sigma)\,d\sigma = \infty$. Yes, this has a bad look, but it is not necessarily a problem [12]! The posterior density in the $s$ parametrization is $p(s|D) = \exp(-s)s^D/D!$, while it is $p(\sigma|D) = \exp(-\sigma^2)\sigma^{2D}/\Gamma(D + 1/2)$ in the $\sigma$ parametrization.

Now, if we transform $p(\sigma|D)d\sigma$ to $p'(s|D)ds$ the result is $p'(s|D) = \exp(-s)s^{D-1/2}/\Gamma(D + 1/2)$, which clearly differs from $p(s|D)$. But, this is not surprising given that the flat prior is not reparametrization invariant. Some regard this as a serious problem, one that worsens as the dimensionality of the parameter space increases. Others point to the numerous successful uses of the uniform prior, even in problems with high dimensional parameter spaces, and accept the lack of invariance as a price worth paying in order to avoid the not inconsiderable effort of constructing an invariant prior.

A general method to create invariant priors was suggested by Jeffreys in the 1930s [15], which in the intervening years has received considerable mathematical validation through many different lines of reasoning (see, for example, [22]). The Jeffreys prior is given by

$$\pi(\theta) = \sqrt{\det I(\theta)}, \tag{37}$$

where $I_{ij} = -E\left[\dfrac{\partial^2 \ln p(x|\theta)}{\partial\theta_i\,\partial\theta_j}\right]$ is the Fisher information matrix,

and where the average is with respect to potential observations $x$ sampled from the density $p(x|\theta)$. When the Jeffreys rule is applied to $p(x|\mu,\sigma) = \text{Gaussian}(x,\mu,\sigma)$ it yields

$$\pi(\mu,\sigma)d\mu\,d\sigma = \frac{d\mu\,d\sigma}{\sigma^2}. \tag{38}$$

**Exercise 9:** Show this

Ironically, the resulting posterior density was rejected by Jeffreys, and subsequently by statisticians because it yielded unsatisfactory inferences! The preferred prior for the Gaussian is

$$\pi(\mu,\sigma)d\mu\,d\sigma = \frac{d\mu\,d\sigma}{\sigma}, \tag{39}$$

because it leads to excellent results.

So, what is a confused physicist to make of this? One possibility is to reject the whole Bayesian omelette and stick to the frequentist gruel. It may be a tad thin for some, but it is at least relatively easy to make. The other is to dismiss the arguments that yield Eq. (37) in favor of reasoning that yields Eq. (39) (see, for example, [21]). Yet another way forward is to take seriously the many persuasive arguments that lead to Eq. (37) and try to understand what the reported failures of the Jeffreys prior for problems involving more than one parameter is telling us. Here is a hint of some understanding. Note that Eq. (39) can be written as

$$\pi(\mu,\sigma)d\mu\,d\sigma = \sigma\left[\frac{d\mu\,d\sigma}{\sigma^2}\right],$$
$$= \sigma_0\exp(\ln\sigma/\sigma_0)\left[\frac{d\mu\,d\sigma}{\sigma^2}\right]. \tag{40}$$

This suggests, in the spirit of [22], that it is better to interpret the Jeffreys prior as simply an invariant measure on the parameter space of the associated likelihood function, one that assigns equal weight to every *probability density* labeled by $\theta$. Assigning equal weight to every probability density is a reparametrization invariant procedure, while, as we saw above, assigning equal weight to every *parameter* is not. If this interpretation is accepted, then the prior density is actually given by

$$\pi(\theta) = g(\theta)\,\sqrt{\det I(\theta)}, \tag{41}$$

where $g(\theta)$ is a function that could assign non-equal weights to the probability densities, such as the term before the brackets in Eq. (40). That term is essentially the exponential of the entropy of the Gaussian density, which assigns a weight $\propto \sigma$ to every density indexed by $\mu, \sigma$. What is missing is a convincing theoretical framework for choosing $g(\theta)$, a challenge that we leave to the reader.

For our example, we shall forego invariance in order to keep things simple and use a flat prior in both $s$ and $b$. But, before delving back into the example, we review hypothesis testing in Bayesian analysis.

### 3.1 Model Selection

Hypothesis testing (also known as model selection) in Bayesian analysis requires the calculation of an appropriate posterior density or probability, as is true of all fully Bayesian calculations,

$$p(\theta, \omega, H|D) = \frac{p(D|\theta, \omega, H)\,\pi(\theta, \omega, H)}{p(D)}, \tag{42}$$

where we have explicitly included the index $H$ to identify the different hypotheses. By marginalizing $p(\theta, \omega, H|D)$ with respect to all parameters except the ones that label the hypotheses or models, $H$, we arrive at

$$p(H|D) = \int p(\theta, \omega, H|D)\, d\theta\, d\omega, \tag{43}$$

that is, the probability of hypothesis $H$ given observed data $D$. In principle, the parameters $\omega$ could also depend on $H$. For example, suppose that $H$ labels different parton distribution function (PDF) models, say CT14, MMHT, and NNPDF, then $\omega$ would depend on the PDF model and should be written as $\omega_H$. Like a Ph.D., it is usually convenient to arrive at the end-point, here the probability $p(H|D)$, in stages.

1. Factorize the prior, e.g.,

$$\begin{aligned}\pi(\theta, \omega_H, H) &= \pi(\theta, \omega_H|H)\,\pi(H),\\ &= \pi(\theta|\omega_H, H)\,\pi(\omega_H|H)\,\pi(H).\end{aligned} \tag{44}$$

In many cases, we can assume that the parameters of interest $\theta$ are independent, *a priori*, of both the nuisance parameters $\omega_H$ as well as the model label $H$, in which case we can write, $\pi(\theta, \omega_H, H) = \pi(\theta)\,\pi(\omega_H|H)\,\pi(H)$.

2. Then, for each hypothesis, $H$, compute the function

$$p(D|H) = \int p(D|\theta, \omega_H, H)\,\pi(\theta, \omega_H|H)\, d\theta\, d\omega_H. \tag{45}$$

3. Then, compute the probability of each hypothesis,

$$p(H|D) = \frac{p(D|H)\,\pi(H)}{\sum_H p(D|H)\,\pi(H)}. \tag{46}$$

Clearly, in order to calculate the probabilities $p(H|D)$ it is necessary to specify the priors $\pi(\theta, \omega|H)$ and $\pi(H)$. With some effort, it is possible to arrive at an acceptable form for $\pi(\theta, \omega|H)$, however, it is highly unlikely that consensus could ever be reached on the prior $\pi(H)$. At best, we would have to make do with a convention. For example we could, by convention, assign equal probabilities to the two hypotheses $H_0$ and $H_1$, *a priori*, that is, $\pi(H_0) = \pi(H_1) = 0.5$. But, do we really believe that the Standard Model and the MSSM are equally probable models?

One way to sidestep the polemics of assigning $\pi(H)$ is to compare probabilities,

$$\frac{p(H_1|D)}{p(H_0|D)} = \left[\frac{p(D|H_1)}{p(D|H_0)}\right] \frac{\pi(H_1)}{\pi(H_0)}, \tag{47}$$

but use only the term in brackets, called the global **Bayes factor**, $B_{10}$, as a way to compare hypotheses. The Bayes factor is the factor by which the relative probabilities of two hypotheses *changes* as a result of incorporating the data, $D$. The word global indicates that we have marginalized over all the parameters of the two models. The *local* Bayes factor, $B_{10}(\theta)$ is defined by

$$B_{10}(\theta) = \frac{p(D|\theta, H_1)}{p(D|H_0)}, \tag{48}$$

where,

$$p(D|\theta, H_1) \equiv \int p(D|\theta, \omega_{H_1}, H_1) \, \pi(\omega_{H_1}|H_1) \, d\omega_{H_1}, \tag{49}$$

are the **marginal** or integrated likelihoods in which we have assumed the *a priori* independence of $\theta$ and $\omega_{H_1}$. We have further assumed that the marginal likelihood $H_0$ is independent of $\theta$, which is a very common situation. For example, $\theta$ could be the expected signal count $s$, while $\omega_{H_1} = \omega$ could be the expected background $b$. In this case, the hypothesis $H_0$ is a special case of $H_1$, namely, it is the same as $H_1$ with $s = 0$. An hypothesis that is a special case of another is said to be **nested** within the more general hypothesis. All this will become clearer when we work through the Bayesian analysis of the 4-lepton data.

There is a notational subtlety that may be missed: because of the way we have defined $p(D|\theta, H)$, we need to multiply $p(D|\theta, H)$ by the prior $\pi(\theta)$ and then integrate with respect to $\theta$ in order to calculate $p(D|H)$.

### 3.2 Bayesian Analysis of 4-lepton Data

In this section, we shall

1. compute the posterior density $p(s|D)$,
2. compute a 68% credible interval $[l(D), u(D)]$, and
3. compute the global Bayes factor $B_{10} = p(D|H_1)/p(D|H_0)$,

as a way to illustrate a Bayesian analysis of the 4-lepton data.

*Probability model*

The likelihood is the same as that used in the frequentist analysis, namely, Eq. (23). However, the likelihood is only part of the model; we also need a prior $\pi(s, b)$ that encodes what we *know*, or *assume*, about the mean background and signal independently of the observations $D$. How exactly that should be done remains an active area of debate and research. Below, we shall take the easy way out!

One point that should be noted is that the prior $\pi(s, b)$ can be factorized in two ways,

$$\pi(s, b) = \pi(s|b) \, \pi(b),$$

$$= \pi(b|s)\,\pi(s). \tag{50}$$

It is worth noting because $\pi(s,b)$ is routinely written as $\pi(s,b) = \pi(s)\pi(b)$, which is not true, in general. The *a priori* independence of $s$ and $b$ is an assumption, one that we shall make. What do we know about $s$ and $b$? We know that $s$ and $b$ are $\geq 0$. We also know the probability function and how $s$ and $b$ enter it. Given this information, there are well founded methods to construct $\pi(s,b)$. However, for simplicity, for $b$, we shall use the improper prior $\pi(b) = k$, where $k$ is the scale factor in the likelihood $p(D|s,b)$, and either the improper prior $\pi(s) = 1$, or the proper prior $\pi(s) = \delta(s - 15.6)$. An improper prior is one that integrates to infinity, which as noted above is not necessarily problematic [12].

*Marginal likelihood*

Having completed the probability model, the rest of the Bayesian analysis proceeds in a routine manner. First, it is convenient to eliminate the nuisance parameter $b$, using the improper prior $\pi(b) = k$,

$$p(D|s, H_1) = \int_0^\infty p(D|s, b)\,\pi(b)\,db,$$

$$= \frac{1}{M}(1-x)^2 \sum_{r=0}^N \mathrm{Beta}(x, r+1, M)\,\mathrm{Poisson}(N-r|, s), \tag{51}$$

where $x = 1/(1+k)$,

**Exercise 10:** Show this

and thereby arrive at the marginal likelihood $p(D|s, H_1)$. The symbol $H_1$ has been introduced to represent the hypothesis that the signal is non-zero.

*Posterior density*

Given the marginal likelihood $p(D|s, H_1)$ and $\pi(s)$ we can compute the posterior density,

$$p(s|D, H_1) = p(D|s, H_1)\,\pi(s)/p(D|H_1), \tag{52}$$

where,

$$p(D|H_1) = \int_0^\infty p(D|s, H_1)\,\pi(s)\,ds.$$

Setting $\pi(s) = 1$ yields,

$$p(s|D, H_1) = \frac{\sum_{r=0}^N \mathrm{Beta}(x, r+1, M)\,\mathrm{Poisson}(N-r|s)}{\sum_{r=0}^N \mathrm{Beta}(x, r+1, M)}. \tag{53}$$

**Exercise 11:** Derive an expression for $p(s|D, H_1)$ assuming $\pi(s) = \mathrm{Gamma}(qs, 1, U+1)$ where $q$ and $U$ are known constants.

The posterior density $p(s|D, H_1)$ completes the inference about the mean signal $s$. In principle, we could stop there, but, in practice, summaries of the posterior density are furnished, such as a **credible interval**, the analog of a confidence interval. But, like confidence intervals, credible intervals, $[l(D), u(D)]$ with credible level $p$, defined by

$$\int_{l(D)}^{u(D)} p(s|D, H_1)\,ds = p \tag{54}$$

are not unique. The analog of Neyman's central interval is the central credible interval defined by

$$\int_0^{l(D)} p(s|D, H_1)\,ds = (1-p)/2,$$
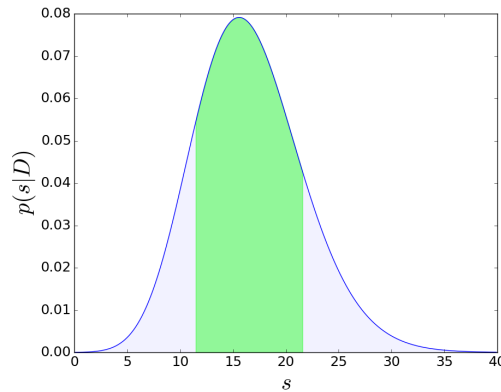
**Fig. 10:** Posterior density for 4-lepton data. The shaded area is the 68% central credible interval.

$$\int_{u(D)}^{\infty} p(s|D, H_1)\, ds = (1 - p)/2. \tag{55}$$

For the 4-lepton data this leads to the central credible interval $[11.5, 21.7]$ for $s$ with $p = 0.683$, which is shown in Fig. 10. The statement $s \in [11.5, 21.7]$ at 68% C.L. means there is a 68% probability that $s$ lies in the specified interval. Unlike the analogous frequentist statement, this one is about this particular interval and the 68% is a degree of belief, not a relative frequency. Statements of this form do, of course, have a coverage probability. However, *a priori*, there is no reason why the coverage probability of credible intervals should satisfy the frequentist principle. In practice, it is found that credible intervals with appropriately chosen priors can moonlight as approximate confidence intervals. But when this happens if does not mean that their interpretations somehow merge, it simply means that a misinterpretation of the intervals is likely to be benign.

*Bayes factor*

We noted above that

$$p(D|H_1) = \int_0^{\infty} p(D|s, H_1)\, \pi(s)\, ds.$$

Furthermore, $p(D|H_1) < \infty$ even with the improper prior $\pi(s) = 1$. However, another *arbitrary* constant besides unity could have been chosen, for example, $\pi(s) = C$. That constant would not have altered the posterior density $p(s|D, H_1)$ and therefore choosing $C = 1$ as a matter of convenience was fine. However, here we wish to compute the global Bayes factor $B_{10} = p(D|H_1) \,/\, p(D|H_0)$. The background-only hypothesis, $H_0$, is nested in $H_1$ and has marginal likelihood $p(D|H_0) \equiv p(D|0, H_1)$. Since the constant $k$ in the background prior $\pi(b) = k$ scales both $p(D|H_1)$ and $p(D|H_0)$ the constant cancels and no issue arises from using an improper background prior. However, since for $H_1$ $\pi(s) = C$ and the parameter $s$ appears only in the calculation of $p(D|H_1)$, the Bayes factor is scaled by the arbitrary constant $C$. Consequently, the Bayes factor can be assigned any value merely by choosing an appropriate value for $C$. This is clearly unsatisfactory. The upshot is that while improper priors may yield reasonable results for the posterior density $p(s|D, H_1)$, albeit ones that are not reparametrization invariant, that is not the case for Bayes factors. To arrive at a satisfactory Bayes factor, a proper prior must be used. The simplest such prior is, for example, $\pi(s) = \delta(s - \hat{s})$, where $\hat{s} = N - B = 15.6$ events. With this prior, the Bayes factor is

$$B_{10} = \frac{p(D|H_1)}{p(D|H_0)} = 4967.$$

We conclude that the 4-lepton observations increase the probability of hypothesis $s = 15.6$ events relative to the probability of the hypothesis $s = 0$ by $\approx 5000$. In order to avoid large numbers, the Bayes factor can be mapped into a measure akin to the frequentist "$n$-sigma",

$$Z = \sqrt{2 \ln B_{10}}, \tag{56}$$

which gives $Z = 4.13$.

The Bayesian and frequentist results are approximately the same, which is typically the case when the data are sufficient. This is because the influence of the prior is smaller than when the data are sparse.

## 4  Supervised Machine Learning

The project of creating artificial beings that mimicked some characteristics of humans has been a dream of visionaries for millennia. But, during the Second World War, dreams gave way to a desperate focus on matters of life and death when latter day visionaries sought to create algorithms that could solve difficult problems such as cracking military codes in real-time. After the war, the pursuit of artificially intelligent agents was revived. In 1950, the great English mathematician Alan Turing, whose genius helped save millions of lives and shortened the most calamitous war in history, proposed an operational definition of such an agent, a test now known as the *Turing test* [23]. The test cuts to the chase regarding what it means to be intelligent: if it is impossible to tell whether one is conversing with a person or a machine and you are in fact conversing with a machine then the latter is intelligent. In the decades following the publication of the Turing test, progress towards creating such agents was slow, in part because the required conceptual breakthroughs were lacking and in part because the available computing power was severely limited.

However, an enormous change has occurred during the last decade or so, driven in part by algorithmic breakthroughs, but mostly by the exponential growth in the size of data sets and the available computing power. In just a few years, the field of *machine learning*, that is, the use of computer-based algorithms to construct useful models of data, has gone from research lab to everyday commercial applications. To be sure, there are many things humans do that seem far beyond current machine learning capabilities. It is still the case that we are unable to replicate a young child's ability to intuit the fact that the noises she hears from the people around her have meaning. Nor can we replicate the extraordinary human ability to be "trained" on a relatively small number of instances of, say, pictures of the Golden Gate bridge, and yet be able to identify the Golden Gate in other pictures of the bridge taken from perspectives that may never have been seen before. Nevertheless, impressive progress has been made recently. Arguably, the most notable is the breakthrough by the Google subsidiary *DeepMind* in creating an agent that taught itself to play to superhuman levels the ancient Chinese game of Go, as well as Chess and Shogi (Japanese chess) *tabula rasa*. These self-teaching feats were achieved in a mere 24 hours [24]!

Our purpose here is considerably more modest; it is to emphasize something that can easily get lost in the hype, namely, that these systems are, for the most part, "simply" highly non-linear high-dimensional parameter space functions that provide mappings from one space to another. The breakthrough has been the ability to fit these enormously complicated functions on practical timescales. In order to avoid complications that merely obfuscate, we consider a simplified version of the following problem: separating Higgs boson events in which the Higgs boson is produced via vector boson fusion (VBF) from events in which the Higgs boson is created via gluon gluon fusion (ggF). But, first, we give an overview of a few key ideas of machine learning.

Most machine learning algorithms fall into five broad categories:

1. supervised learning,
2. semi-supervised learning,

3. unsupervised learning (i.e., pattern detection)

4. reinforcement learning, and

5. generative learning.

The simplest category of algorithm is supervised learning in which the data for fitting models, i.e., training them, consist of labeled objects. If the labels define the class to which objects belong, for example, $-1$, or $0$, for gluon gluon fusion events and $+1$ for vector boson fusion events, then, as shown below, the resulting function will be a *classifier*. If the labels form a continuous set, then the resulting function will be a *regression function* (sometimes called a "regressor"). For example, suppose the objects are jets characterized by their transverse momentum $p_T$ and pseudo-rapidity $\eta$ and possibly other detailed characteristics, such as the electromagnetic fraction, while the labels are the true jet transverse momenta. The regressor will be a correction function that maps the jet characteristics to an approximation of the true jet $p_T$. Our example will be a simple VBF/ggF classifier.

## 4.1 A Bird's Eye View of Supervised Learning

Supervised machine learning can be construed as a game in which winning means picking the best function (or functions) from a function space. The game includes three elements:

1. a function space $\mathcal{F} = \{f(x, w)\}$ containing parametrized functions $f(x, w)$, where $x$ are object characteristics—*features* in machine learning jargon—and $w$ are the parameters;

2. a loss function $L(y, f)$, which measures the cost of making a bad function choice, and where $y$ are labels associated with the features $x$, and

3. a constraint $C(w)$ that places some restriction on the choices.

The best function $f(x, w*)$ is found by minimizing the constrained *empirical risk*,

$$R(f) = \sum_{i=1}^{K} L(y_i, f_i) + C(w), \text{ where } f_i = f(x_i, w), \tag{57}$$

with respect to the choice of function $f$, which in practice means with respect to the parameters $w$.

*Minimization via Gradient Descent*

A loss function, through the empirical risk, defines a "landscape" in the space of parameters, or equivalently in the space of functions. The goal is to find the lowest point in that landscape, usually by moving in the direction of the local negative gradient,

$$w_j \leftarrow w_j - \rho \frac{\partial R}{\partial w_j}, \quad j = 1, \cdots J, \tag{58}$$

where $\rho$ is called the learning rate and $J$ is the dimensionality of the parameter space, which, in some recent commercial applications can be in the millions. As is, the algorithm in Eq. (58) would fail miserably because of the complexity of the landscape and the possibility that the minimizer could get stuck in a local minimum or diverge away from the minimum because of the instability caused by a saddle point. To alleviate this problem, the standard approach is to replace the exact derivatives $\partial R / \partial w_j$ by *noisy* estimates thereof. This is usually achieved by replacing $R$ by an approximation that uses a small subset—that is, *batch*—of the training data in the sum that defines $R$. Typically, a new batch is used at every step of the minimization algorithm. This minimization algorithm is called *stochastic gradient descent*, of which there are many variations. The addition of noise increases the chance that the minimizer will escape from an unfavorable location in the parameter space.

*To Infinity and Beyond*

It is intuitively clear that a successful minimization of the empirical risk, Eq. (57), will yield a solution $f(x, w*)$ that is as close as possible to the labels, or **targets**, $y$. But, in mathematics, as in physics, we can gain a a clearer understanding of a construct by taking a suitable limit of it. To that end, consider the limit of $R(f)/K$, that is, the *average loss*, as $K \rightarrow \infty$. Writing the average loss in that limit as $E$, and assuming that the effect of the constraint goes to zero in that limit, we can write

$$
\begin{aligned}
E[f] &= \int dx \int dy \, L(y, f) \, p(y, x), \\
&= \int dx \, p(x) \left[ \int dy \, L(y, f) \, p(y|x) \right],
\end{aligned}
\tag{59}
$$

where we have used $p(y|x) = p(y, x)/p(x)$. The function $p(y, x)$ is the (typically unknown) joint probability density of the targets and features $(y, x)$. Whether the features $x$ represent an event, a jet, an image, or piece of writing, and $y$ represents useful known data about each instance of $x$, all the information about the mapping from $x$ to $y$ is contained in the joint probability density $p(y, x)$. This is an important point because the failures of machine learning are almost always due to an object with known characteristics $x'$, but unknown label $y'$, not being a member of the population $\{(y, x)\}$ that defines $p(y, x)$. If an agent is trained on a million images of dogs and cats, it is not surprising that it will classify a horse as either a dog or a cat because the probability density $p(y, x)$ does not encompass images of horses. The point is that the function $f(x, w)$ will do what it is designed to do. But, what exactly is $f(x, w)$ designed to do? To answer this question concretely, let us consider the minimization of Eq. (59) with the widely used **quadratic loss**,

$$
L(y, f) = (y - f)^2.
\tag{60}
$$

If we change the function $f$ by an arbitrary amount $\delta f$ this induces a change

$$
\delta E = 2 \int dx \, p(x) \, \delta f(x, w) \left[ \int dy \, (y - f) \, p(y|x) \right],
\tag{61}
$$

in the mean loss $E$, which, in general, is not zero. If, however, the function $f(x, w)$ is sufficiently flexible, it will be possible to reach the minimum of $E$, where $\delta E = 0$. But, we want this to hold for all variations $\delta f$—because these variations are, after all, arbitrary—and for all values of $x$ in order that the function $f$ not fail—that is, perform poorly—for some subset of the space of features. This can be assured provided that the quantity in brackets in Eq. (61) is zero, that is, if

$$
\boxed{f(x, w*) = \int y \, p(y|x) \, dy.}
\tag{62}
$$

Equation (62) is an important result because it tells us precisely what the function $f(x, w*)$ approximates. If one uses the quadratic loss, then the function $f(x, w*)$ approximates the conditional average of the targets. This result was first derived in the context of neural networks [25–27], however, the result holds irrespective of the details of the function $f(x, w)$. In particular, the function does not have to be a neural network. The result holds provided that

1. we use sufficient training data $T = \{(y, x)\}$,
2. we use a sufficiently flexible function $f(x, w)$ and
3. we use an appropriate loss function.

Moreover, if we choose targets in the discrete set $y \in \{0, 1\}$, the general result reduces to

$$
\boxed{f(x, w*) = p(y = 1|x)}
\tag{63}
$$

We conclude that if we minimize the average quadratic loss using training data in which one class of objects is labeled with $y = 0$ and the other with $y = 1$, the function $f(x, w*)$ approximates the probability that the object with features $x$ belongs to the class labeled with $y = 1$; that is, $f(x, w*)$ is a classifier that approximates the class probability. From Bayes theorem, this class probability, $p(1|x)$, can be written as

$$p(1|x) = \frac{p(x|1)\, p(1)}{p(x|1)\, p(1) + p(x|0)\, p(0)}, \tag{64}$$

where $p(1)$ and $p(0)$ are the prior probabilities associated with the two classes. Typically, one trains with $p(1) = p(0)$, in which case $p(1|x)$ is referred to as a discriminant, $D(x)$, and is given by

$$D(x) = \frac{p(x|1)}{p(x|1) + p(x|0)}. \tag{65}$$

*Boosted Decision Trees*

Boosted decision trees (BDT) [28] are, currently, the most popular machine learning method in particle physics; and for good reason. They perform well, they are faster to train than neural networks, they are insensitive to poorly performing variables, and they are resistant to overfitting. In view of their widespread use, it is worth taking the time to understand exactly what this machine learning model entails. We shall highlight key features of BDTs using a simple example in which we seek to separate Higgs boson events produced via vector boson fusion (VBF) from gluon gluon fusion (ggF) produced events. In this section, we first discuss decision trees (DT) and then the notion of boosting, that is, enhancing the performance of a machine learning model by averaging over many models.

A decision tree is a nested sequence of *if then else* statements, which can also be viewed as a histogram whose bins are created recursively through *binary partitioning*. The VBF/ggF example uses two discriminating variables (features) $|\Delta\eta|_{jj}$ and $m_{jj}$, the absolute pseudo-rapidity difference between the two most forward (i.e., largest rapidity) jets in the event and the associated di-jet mass, respectively. Figure 11 shows two representations of a decision tree for our VBF/ggF discrimination example.

At face value, decision trees do not seem to fit into the mathematical ideas about loss functions discussed above. In particular, it is far from clear what loss function, if any, is being minimized when a decision tree is grown. However, all successful uses of decision trees entail averaging over many of them. As we shall see, it is the averaging that provides the connection to a loss function. Averaging also mitigates a serious problem with decision trees, namely, their instability. Even minor changes to the training data can radically alter the structure of a tree.

The first successful averaging algorithm, called `AdaBoost`, was published by AT&T researchers Freund and Schapire in 1997 [29] who showed that it was possible to create high performance classifiers by averaging ones (called **weak learners**) that perform only marginally better than classification via a coin toss. The algorithm builds a classifier using training data labeled by the discrete labels $y = -1$ or $y = +1$. In the VBF/ggF example below, $y = -1$ is assigned to ggF events and $+1$ is assigned to VBF events. The algorithm, for $N$ training events and $K$ decision trees, proceeds as follows:

1. **initialize** event weights $\omega_{1,n} = 1/N$, $n = 1, \cdots, N$
2. **repeat for** $k \in 1, \cdots K$
    (a) fit a tree $f_k(x)$ that returns either $-1$ or $+1$, using the current event weights $\{w_{k,n}\}$
    (b) compute error rate $\epsilon_k = \sum_{n=1}^{N} \omega_{k,n} I[-y_n f_k(x_n)]$, $I(z) = 1$ if $z > 1$, 0 otherwise
    (c) compute coefficient $\alpha_k = \frac{1}{2}\ln[(1 - \epsilon_k)/\epsilon_k]$
    (d) update weights $w_{k+1,n} = w_{k,n}\exp(-\alpha_k y_n f_k(x_n))/Z_k$,
        where $Z_k = \sum_{n=1}^{N} \omega_{k,n}\exp(-\alpha_k y_n f_k(x_n))$
3. classifier $f(x) = \sum_{k=1}^{K} \alpha_k f_k(x)$
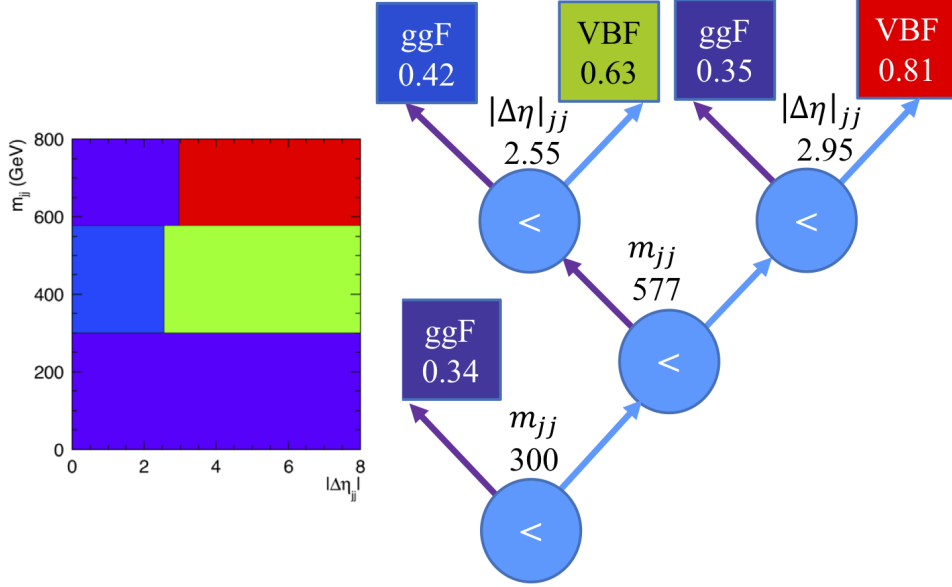
**Fig. 11:** Two representations of a decision tree to separate VBF from ggF events based on the variables $|\Delta\eta|_{jj}$ and $m_{jj}$. On the right, the decision tree is represented as a branching structure in which the circles, called *nodes*, represent *if then else* decisions, that is, *binary* decisions. The boxes terminate the tree and are referred to, appropriately, as *leaves*. On the left, the decision tree is represented as a 2D histogram in which the bins, which correspond to the leaves, have been defined by recursive binary partitioning. The bin boundaries, that is, the binary partitions, correspond to the decisions. At a given node, the left branch is taken if $x < x_{\text{cut}}$ otherwise the right branch is taken; $x_{\text{cut}}$ is an optimal cut on the variable $x \in \{|\Delta\eta|_{jj}, m_{jj}\}$. The numbers within the leaves are the VBF purity $p = S/(S + B)$, where $S$ and $B$ are the VBF and ggF event counts in a given bin, that is, leaf.

In step 2(d), the weight of incorrectly classified events, for which $y_n f_k(x_n) = -1$, is *increased*, while that of correctly classified events, for which $y_n f_k(x_n) = +1$, is *decreased*.

AdaBoost is a rather cryptic algorithm, which, like decision tree classifiers, does not seem to fit into the general discussion about average loss given above. However, subsequent to the publication of the AdaBoost algorithm, Friedman, Hastie, and Tibshirani [30] showed that this algorithm can be viewed as a way to minimize the average loss function

$$E[f] = \int dx \int dy \exp[-yf(x)] \, p(y, x), \tag{66}$$

whose minimum occurs at

$$f(x) = \frac{1}{2} \ln \frac{p(y = +1|x)}{p(y = -1|x)}. \tag{67}$$

To see this consider the problem of minimizing the error rate on the training sample,

$$\epsilon = \frac{1}{N} \sum_{n=1}^{N} I[y_n f(x_n)].$$

Minimizing the error rate directly in a reasonable amount of time is extremely difficult, therefore, in practice, a proxy for the error rate is minimized instead. Noting that $\exp(-y_n f(x_n)) > 1$ when $y_n f((x_n) < 0$, one such proxy is the righthand side of the following

$$\epsilon = \frac{1}{N} \sum_{n=1}^{N} I[y_n f(x_n)],$$

$$\leq \frac{1}{N} \sum_{n=1}^{N} e^{-y_n f(x_n)} I[y_n f(x_n)],$$

$$\leq \frac{1}{N} \sum_{n=1}^{N} e^{-y_n f(x_n)}. \tag{68}$$

Note that in the limit $N \to \infty$, the righthand side of the above equation, which is an upper bound on the error rate, converges to Eq. (66). Furthermore, from the recursive definition of the normalized event weights $w_{k+1,n} = w_{k,n} \exp(-\alpha_k y_n f_k(x_n))/Z_k$ in the `AdaBoost` algorithm, we conclude that

$$\epsilon \leq \frac{1}{N} \sum_{n=1}^{N} e^{-y_n f(x_n)} = \prod_{k=1}^{K} Z_k,$$

$$= \prod_{k=1}^{K} \sum_{n=1}^{N} \omega_{k,n} \exp(-\alpha_k y_n f_k(x_n)) \equiv \epsilon'. \tag{69}$$

If we regard the coefficients $\alpha_k$ as free parameters (by neglecting the dependence of the event weights on $\alpha_k$), we can minimize $\epsilon'$ with respect to $\alpha_k$ by solving

$$\frac{\partial \epsilon'}{\partial \alpha_k} = - \left( \prod_{j \neq k} Z_j \right) \sum_{n=1}^{N} \omega_{k,n} y_n f_k(x_n) \exp(-\alpha_k y_n f_k(x_n)) = 0,$$

$$\text{that is, } \sum_{n=1}^{N} \omega_{k,n} y_n f_k(x_n) \exp(-\alpha_k y_n f_k(x_n)) = 0. \tag{70}$$

Since $y \in \{-1, +1\}$, we can write

$$e^{-\alpha_k} \sum_{n=1}^{N} \omega_{k,n} I[y_n f_k(x_n)] - e^{\alpha_k} \sum_{n=1}^{N} \omega_{k,n} I[-y_n f_k(x_n)] = 0,$$

$$e^{-\alpha_k}(1 - \epsilon_k) - e^{\alpha_k} \epsilon_k = 0,$$

$$\text{where, recall, } \epsilon_k = \sum_{n=1}^{N} \omega_{k,n} I[-y_n f_k(x_n)], \tag{71}$$

is the weighted error rate. We therefore conclude that the upper bound on the error rate $\epsilon'$ is minimized if we choose the coefficients to be $\alpha_k = \frac{1}{2} \ln[(1 - \epsilon_k)/\epsilon_k]$, which is indeed the choice made in `AdaBoost`.

Therefore, in spite of appearances, boosted decision trees fit into the mathematical framework sketched above. In particular, `AdaBoost` can be viewed as a clever way to minimize the average exponential loss given in Eq. (66). Moreover, while the boosted decision tree $f(x)$ cannot be interpreted as a probability, it can be mapped to a probability by inverting Eq. (67),

$$p(y = +1|x) = \frac{1}{1 + \exp(-2f(x))}. \tag{72}$$

Below, we illustrate the use of the `AdaBoost` algorithm using the Toolkit for Multivariate Analysis `TMVA` [31], which is released with the `ROOT` [32] package from CERN. Note, in the `TMVA` implementation, $\alpha_k$ is defined omitting the factor of 1/2, therefore, in order to convert the unnormalized BDT, $f(x)$, in `TMVA` to a probability, the appropriate mapping is

$$p(y = +1|x) = \frac{1}{1 + \exp(-f(x))} \quad (\text{TMVA}). \tag{73}$$

287
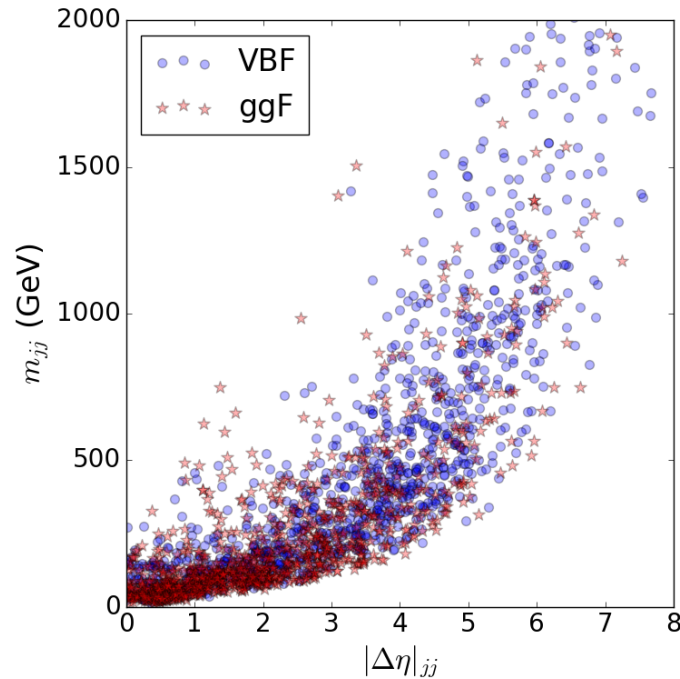
**Fig. 12:** Simulated distributions of the discriminating variables $(|\Delta\eta_{jj}|, m_{jj})$ for VBF and ggF events. As expected, there is a larger rapidity gap between the jets inVBF events than those in ggF, which arise from gluon radiation.

### VBF/ggF discrimination

In this example, a BDT is trained using the `AdaBoost` algorithm in `TMVA` to discriminate between events in which the Higgs boson is created via vector boson fusion (VBF) and events in which the Higgs boson is created via gluon gluon fusion (ggF). The key difference between VBF events and ggF events is that the former features a pair of forward (i.e., large rapidity) jets that is absent from the latter. It is found that the two most discriminating variables between these two classes of events are the absolute pseudo-rapidity difference $|\Delta\eta|_{jj}$ between the two jets and the associated di-jet mass $m_{jj}$ and. The predicted distributions of the two variables is shown in Fig. 12.

We use a training sample size of $N = 20,000$ events, split equally between VBF and ggF events with assigned targets of $y = +1$ and $y = -1$, respectively. The `TMVA` training parameters are `BoostType=AdaBoost`, `NTrees=800`—the number of trees $K$, `nEventsMin=100`—the minimum number of events per bin, and `nCuts=50`—the number of binary partitions per variable to search for the optimal partition, i.e., cut. The optimal cut is the one which gives the greatest *decrease* in impurity as measured by the Gini index[7], defined by $p(1-p)$ where $p = S/(S+B)$ is the purity and $S$ and $B$ are the signal and background counts, respectively, in a given bin. A bin is maximally pure, either pure signal or pure background, when the Gini index is zero.

Figure 13 shows the first six decision trees as histograms, each with its associated coefficient $\alpha_k = \ln[(1 - \epsilon_k)/\epsilon_k]^8$ printed on the histogram. A decision tree is a piecewise constant function in which each bin (i.e., leaf) is assigned a value. In the `AdaBoost` algorithm, the values are $y = \pm1$; in our example, $y = -1$ for bins in which $B > S$ (i.e., ggF bins) and $+1$ for bins in which $S > B$ (i.e., VBF bins). A given feature vector $x = |\Delta\eta|_{jj}, m_{jj}$, characterizing an event, will fall in a bin in each of the six decision trees of Fig. (13) and the BDT is equal to the average $\sum_{k=1}^{6} f_k(x)$ where each tree $f_k(x)$

---

[7]After Italian statistician Corrado Gini, 1884-1965.

[8]As noted, `TMVA` omits the factor of 1/2.
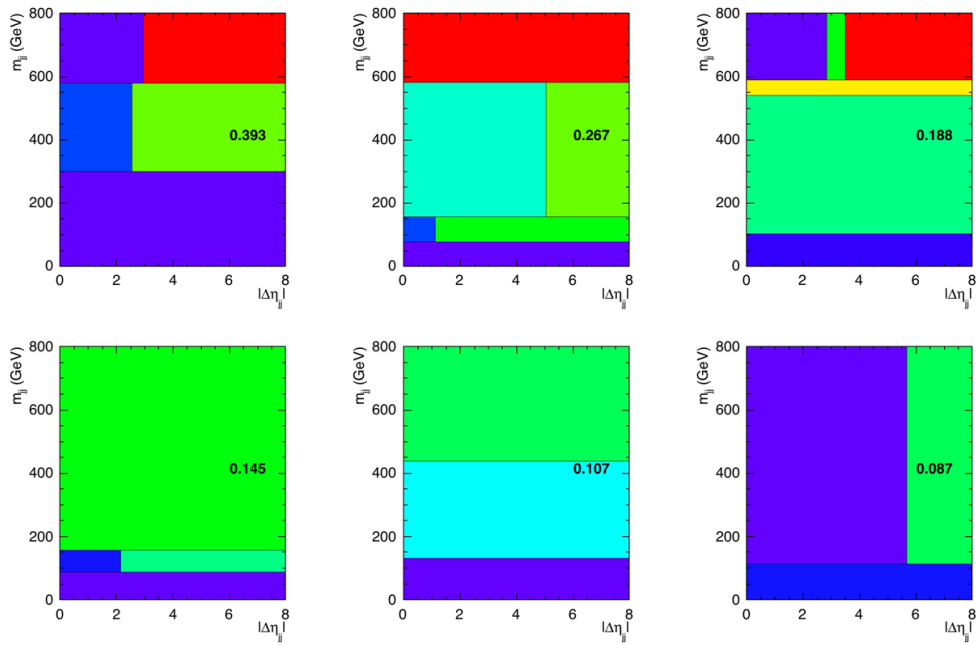
**Fig. 13:** The first six of the 800 decision trees, displayed as 2D histograms, showing the coefficients $\alpha_1, \cdots, \alpha_6$ associated with the threes.
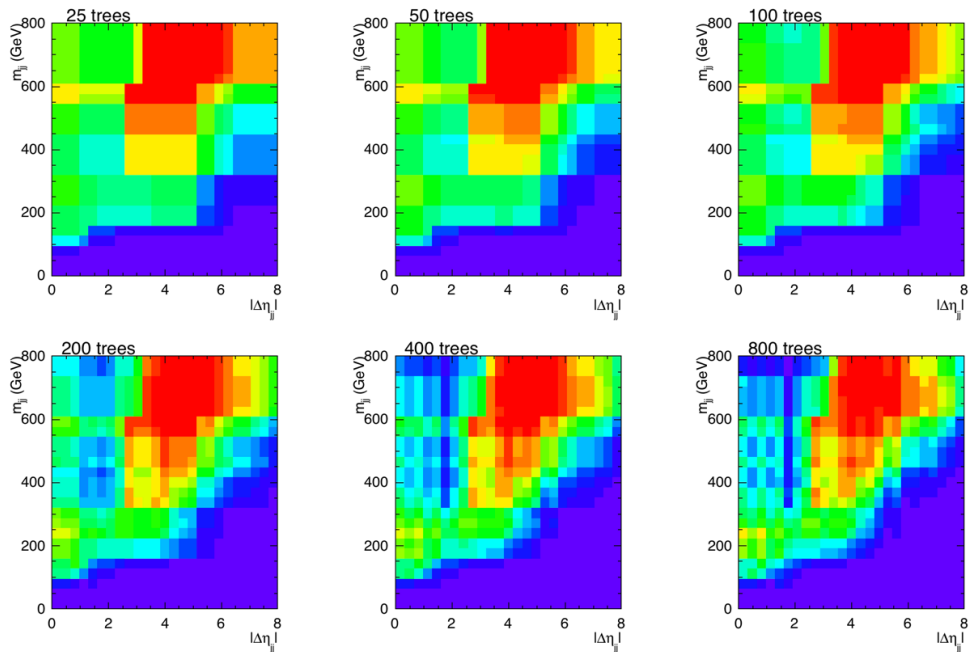


**Fig. 14:** The outputs of boosted decision trees averaged over differing numbers of decision trees, 25, 50,...,800. Each $BDT(x)$, with $x = |\Delta\eta|_{jj}, m_{jj}$, is mapped to the probability $p(y = +1 \,|\, x) = 1/[1 + \exp(-BDT(x))]$.
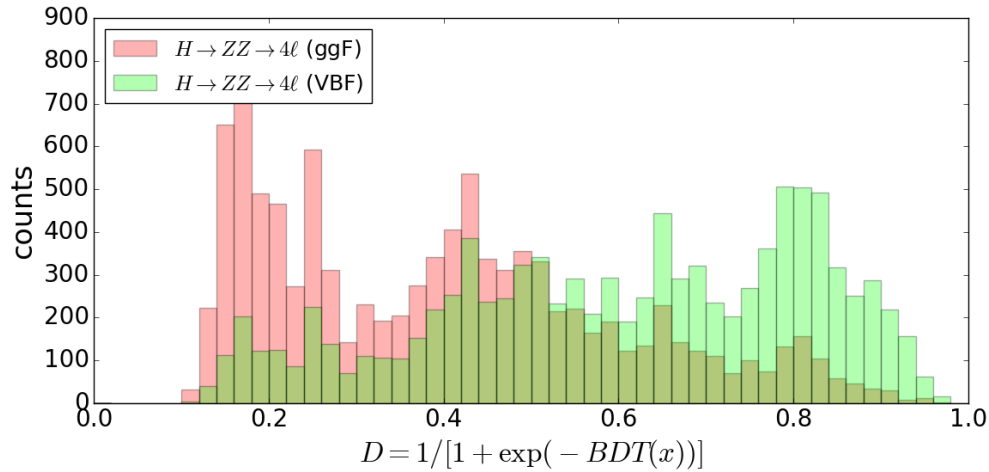
**Fig. 15:** The distributions of the discriminant $D(x) = 1/[1 + \exp(-BDT(x))]$, where $BDT(x)$ is a boosted decision tree with $K = 800$ trees.
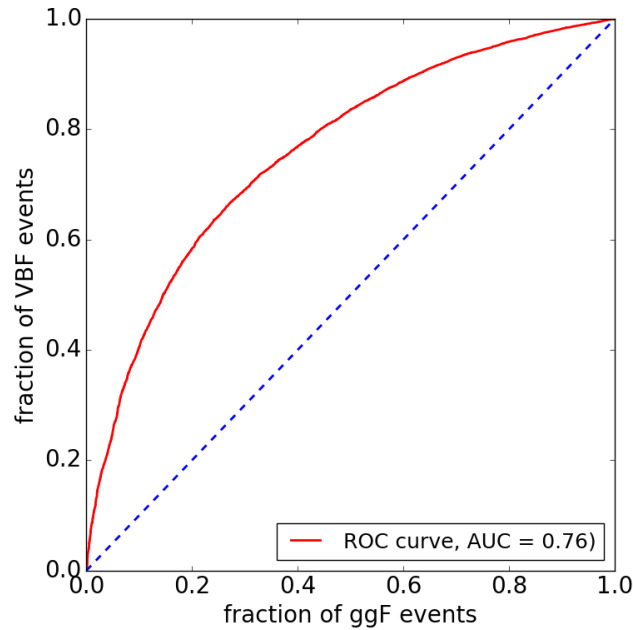


**Fig. 16:** Receiver operating characteristic (ROC) curve. The area under the curve (AUC) is a commonly used global measure of the discrimination power of a classifier.

returns either $+1$ or $-1$ depending on the bin in which $x$ falls. In other words, a BDT is an average over histograms, each with different set of bins. While the piecewise constant nature remains, the more histograms (that is, trees) are averaged, the smoother one expects the BDT output to become. This is illustrated in Fig. 14, which shows the effect of averaging over an increasing number of trees. Finally, Figs. 15 and 16 show the distribution of the BDT, in which the output has been mapped to the probability $p(\text{VBF} \,|\, x) \equiv p(y = +1 \,|\, x) = 1/[1 + \exp(-BDT(x))]$, and the receiver operating characteristic (ROC) curve of the BDT.

The ROC curve, and the area underneath it (AUC), are often used as simple measures of the performance of a binary classifier. The larger the AUC the better the performance of the classifier.

## Summary

We have given an overview of the frequentist and Bayesian approaches to statistical inference and a brief survey of the main mathematical ideas that underpin supervised machine learning. Frequentist analysis is based on the relative frequency interpretation of probability and, ideally, adheres to the frequentist principle: repeated application of a statistical procedure will yield statements a fraction $f \geq p$ of which are guaranteed to be true, where $p$ is the desired confidence level. The Bayesian approach uses the degree of belief interpretation of probability and Bayes theorem as the primary inference algorithm. In both approaches, the key task is building an accurate probability model.

A brief introduction to supervised machine learning was given in which the emphasis was clarifying the critical role of the loss function. We noted the mathematical fact that the quantity approximated by a machine learning model is determined by the loss function and not by the particulars of the model provided that sufficient training data are used, the model is sufficiently flexible, and a good approximation to the minimum of the average loss can be found.

## References

[1] L. Lyons, *Statistics for Nuclear and Particle Physicists*, Cambridge University Press, Cambridge (1989).

[2] F. James, *Statistical Methods in Experimental Physics*, 2nd Edition, World Scientific, Singapore (2006).

[3] G. Cowan, *Statistical Data Analysis*, Oxford University Press, Oxford (1998).

[4] R. J. Barlow, *Statistics: A Guide To The Use Of Statistical Methods In The Physical Sciences*, The Manchester Physics Series, John Wiley and Sons, New York (1989).

[5] S. K. Chatterjee, *Statistical Thought: A Perspective and History*, Oxford University Press, Oxford (2003).

[6] L. Daston, "How Probability Came To Be Objective And Subjective," Hist. Math. 21, 330 (1994).

[7] S. Chatrchyan *et al.* [CMS Collaboration], "Measurement of the properties of a Higgs boson in the four-lepton final state," Phys. Rev. D **89**, no. 9, 092007 (2014) doi:10.1103/PhysRevD.89.092007 [arXiv:1312.5353 [hep-ex]].

[8] J. Neyman, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability," Phil. Trans. R. Soc. London A236, 333 (1937).

[9] G. J. Feldman and R. D. Cousins, "Unified approach to the classical statistical analysis of small signals," Phys. Rev. D57, 3873 (1998).

[10] S. E. Fienberg and D. V. Hinkley, eds., *R.A. Fisher: An Appreciation*, Lecture Notes on Statistics, Volume 1, Springer Verlag (1990).

[11] G. Cowan, K. Cranmer, E. Gross, O. Vitells "Asymptotic formulae for likelihood-based tests of new physics," Eur. Phys. J. C71, 1554 (2011).

[12] G. Taraldsen and B.H. Lindqvist, "Improper Priors Are Not Improper," The American Statistician, Vol. 64, Issue 2, 154 (2010).

[13] G. Aad *et al.* [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**, 1 (2012) [arXiv:1207.7214 [hep-ex]].

[14] S. Chatrchyan *et al.* [CMS Collaboration], "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," Phys. Lett. B **716**, 30 (2012) [arXiv:1207.7235 [hep-ex]].

[15] H. Jeffreys, *Theory of Probability*, 3rd Edition, Clarendon Press, Oxford (1961).

[16] V. M. Abazov *et al.* [D0 Collaboration], "Observation of Single Top Quark Production," Phys. Rev. Lett. **103**, 092001 (2009) [arXiv:0903.0850 [hep-ex]].

[17] T. Aaltonen *et al.* [CDF Collaboration], "First Observation of Electroweak Single Top Quark Production," Phys. Rev. Lett. **103**, 092002 (2009) [arXiv:0903.0885 [hep-ex]].

[18] S. Sekmen *et al.*, "Interpreting LHC SUSY searches in the phenomenological MSSM," JHEP **1202**, 075 (2012) doi:10.1007/JHEP02(2012)075 [arXiv:1109.5119 [hep-ph]].

[19] V. Khachatryan *et al.* [CMS Collaboration], "Phenomenological MSSM interpretation of CMS searches in pp collisions at sqrt(s) = 7 and 8 TeV," JHEP **1610**, 129 (2016) doi:10.1007/JHEP10(2016)129 [arXiv:1606.03577 [hep-ex]].

[20] V. Khachatryan *et al.* [CMS Collaboration], "Search for supersymmetry in pp collisions at sqrt(s) = 8 TeV in final states with boosted W bosons and b jets using razor variables," Phys. Rev. D **93**, no. 9, 092009 (2016) doi:10.1103/PhysRevD.93.092009 [arXiv:1602.02917 [hep-ex]].

[21] L. Demortier, S. Jain and H. B. Prosper, "Reference priors for high energy physics," Phys. Rev. D **82**, 034002 (2010) [arXiv:1002.1111 [stat.AP]].

[22] I. J. Myung, V. Balasubramanian, and M. A. Pitt, "Counting probability distributions: Differential geometry and model selection," PNAS, **97** 11170-11175 (2000); doi: 10.1073/pnas.170283897.

[23] A. Turing, "Computing Machinery and Intelligence," Mind **59** 433-460 (1950).

[24] D. Silver et al., "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm," Science **362** 1140-1144 (2018); DOI: 10.1126/science.aar6404.

[25] Ruck et al., IEEE Trans. Neural Networks 4, 296-298 (1990).

[26] Wan, IEEE Trans. Neural Networks 4, 303-305 (1990).

[27] Richard and Lippmann, Neural Computation. 3, 461-483 (1991).

[28] H. J. Yang, B. P. Roe and J. Zhu, "Studies of boosted decision trees for MiniBooNE particle identification," Nucl. Instrum. Meth. A **555**, 370 (2005) doi:10.1016/j.nima.2005.09.022 [physics/0508045].

[29] Y. Freund and R.E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and Sys. Sci. **55** (1), 119 (1997).

[30] J. Friedman, T. Hastie and R. Tibshirani, ï£¡Additive logistic regression: a statistical view of boosting,ï£¡ The Annals of Statistics, **28** (2), 377-386 (2000).

[31] P. Speckmayer, A. Hocker, J. Stelzer and H. Voss, "The toolkit for multivariate data analysis, TMVA 4," J. Phys. Conf. Ser. **219**, 032057 (2010). doi:10.1088/1742-6596/219/3/032057

[32] Rene Brun and Fons Rademakers, "ROOT - An Object Oriented Data Analysis Framework," Proceedings AIHENP 96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A **389**, 81-86 (1997). See also https://root.cern.ch/.

# Organizing Committee

T. Donskova (Schools Administrator, JINR)
C. Duhr (CERN & UCLouvain)
N. Ellis (CERN)
M. Mulders (CERN)
A. Olchevsky (JINR)
K. Ross (Schools Administrator, CERN)
G.. Zanderighi (CERN & Oxford Univ.)

# Local Organizing Committee

M.G. Alviggi (Univ. Federico II & INFN Naples)
F. Bianchi (Univ. Turin & INFN Turin)
C. Bozzi (INFN Ferrara.)
G. Carlino (INFN Naples)
N. Cavallo (Univ. Basilicata & INFN Naples)
F. Cirotto (INFN Naples)
A. Di Crescenzo (Univ. Federico II & INFN Naples)
F. Fabozzi (Univ. Basilicata & INFN Naples)
A.O.M. Iorio (Univ. Federico II & INFN Naples)
L. Lista (INFN Naples)
M. Merola (Univ. Federico II & INFN Naples)
C. Satriano (Univ. Basilicata & INFN Rome I)
S. Spagnolo (Univ. Salento & INFN Lecce)

# International Advisors

F. Gianotti (CERN)
V. Matveev (JINR)
N. Pastrone (INFN Turin)
A. Skrinsky (BINP)
N. Tyurin (IHEP)

# Lecturers

B. Allanach (Cambridge Univ.)
A. Bednyakov (JINR)
A. De Simone (SISSA)
E. Fraga (UFRJ)
F. Maltoni (UCLouvain)
S. Pascoli (Durham Univ.)
N. Pastrone (INFN Turin)
H. Prosper (Florida State Univ.)
F. Tramontano (Univ. Federico II & INFN Naples)
J. van den Brand (Nikhef)
J. Zupan (Univ. Cincinnati)

# Discussion Leaders

S. Borowka (CERN)
D. Pagani (TUM)
A. Pikelner (Hamburg Univ.)
R. Sadykov (JINR)
N. Vignaroli (INFN Padua)
E. Vryonidou (CERN)

# Students

Christina AGAPOPOULOU

Ece AKILLI

Alejandro ALFONSO ALBERO

Konie AL KHOURY

Afiq ANUAR

Cristian BALDENEGRO BARRERA

Anjishnu BANDYOPADHYAY

Samuel BELIN

Violaine BELLEE

Anna BENECKE

Alessandra BETTI

Federico BETTI

Nihal BRAHIMI

Milene CALVETTI

Sonia CARRA

Cheng CHEN

Thibault CHEVALERIAS

Hadar COHEN

Sandra CONSUEGRA RODRIGUEZ

Ana Rosario CUETO GOMEZ

Daniel DABROWSKI

Mohamed DARWISH

Matteo DEFRANCHIS

Yasiel DELABAT DIAZ

Francesco Armando DI BELLO

Aleksandr DOLMATOV

Etienne DREYER

Archil DURGLISHVILI

Filippo ERRICO

Peter Johannes FALKE

Blake Christopher FORLAND

Simone FRANCESCATO

Dylan FRIZZELL

Ilnur GABDRAKHMANOV

Christoph Andreas GARBERS

Simona GARGIULO

Lukas GERRITZEN

Aishik GHOSH

Antonio GIANNINI

Ekaterina GOVORKOVA

Francesco Giuseppe GRAVILI

Ralf GUGEL

Ziyu GUO

Andreas HOENLE

Rosanna IGNAZZI

Ruth Magdalena JACOBS

Adomas JELINSKAS

Hessamoddin KAVEH

Joscha KNOLLE

Olesya KUCHINSKAYA

Alessandro LAPERTOSA

Francesco LA RUFFA

Marco LAVORGNA

Vasilisa LENIVENKO

You-Ying LI

Ivan LILIENBERG

Jared LITTLE

Roman LITVINOV

Alena LOESLE

Daniele MARANGOTTO

Nis MEINERT

Elmira MELYAN

Titus MOMBACHER

Filip NECHANSKY

Hoang Dai Nghia NGUYEN

Daniel Stefaniak NIELSEN

Tadej NOVAK

Aleksey PAVLOV

DanyerPEREZ ADAN

Anton PETROV

Thu Le Ha PHAM

Anum QURESHI

Diego RAMIREZ GARCIA

Baptiste RAVINA

Daniele RUINI

Alessia SAGGIO

Merve SAHINSOY

Daniel SAVOIU

Daniela SCHAEFER

Korbinian SCHWEIGER

Jonathan SHLOMI

Lais SOARES LAVRA

Daniel SPITZBART

Nicolas TONON

Lukas TROPP

Fabrizio TROVATO

Fang-Ying TSAI

Eftychia TZOVARA

Antonio VAGNERINI

Valentina VECCHIO

Basile VERMASSEN

Emmanouil VOURLIOTIS

Michael WASSMER

Thomas WESTON

Julian WOLLRATH

Vincent Wai Sum WONG

# Posters

| Poster title | Presenter |
| --- | --- |
| A high granularity timing detector (HGTD) in ATLAS: sensors and electronics testing | AGAPOPOULOU, C. |
| ATLAS LAr Calo electronics upgrade: Tests and Performance of prototypes | AL KHOURY, K. |
| Probing the hard color-singlet exchange at 13 TeV | BALDENEGRO-BARRERA, C. |
| Heavy Ion Physics at LHCb with EXPLORINGMATTER | BELIN, S. |
| CMS Pixel Detector - Investigating the DCDC issue | BENECKE, A. |
| Search for Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ fully hadronic final state with the ATLAS detector | BETTI, A. |
| $R(D^*)$ with hadronic $\tau$ decays at LHCb | BETTI, F. |
| Strategies for new measurement in VH. channel, H→bb | CALVETTI, M. |
| Search for direct chargino pair production with W-boson mediated decays in final states with two leptons at Run II with the ATLAS detector | CARRA, S. |
| Search for $t\bar{t}t\bar{t}$ production with the ATLAS detector at the LHC | CHEVALEÉRIAS, TH. |
| Search for a very light NMSSM Higgs Boson produced in decays of the 125 GeV scalar boson, decaying into a pair of muons and a pair of $\tau$ leptons at 13 TeV | CONSUEGRA RODRIGUEZ, S. |
| Design of the MPD / NICA Time-of-Flight System | DABROWSKI, D. |
| Search for Heavy Neutral Lepton with CMS detector non-prompt signature | DARWISH, M. |
| Impact of statistical fluctuations in the Monte Carlo on template fits | DEFRANCHIS, M. |
| NRC Kurchatov Institute Scintillation Fibre Mat Production Centre | DOLMATOV, A. |
| Search for flavour-changing neutral current top-quark decays $t \to qZ$ in proton-proton collisions at 13 TeV with the ATLAS detector | DURGLISHVILI, A. |

| Poster title | Presenter |
|---|---|
| Search for new massive resonances in dilepton mass spectra in pp collisions with full 2016 Run II dataset at CMS | ERRICO, F. |
| Search for $Z' \to ll$ resonances and combination with $W' \to l\nu$ | FALKE, P. |
| Online monitoring system for the BM@N experiment | GABDRAKHMANOV, I.R. |
| Jet residual correction with 2017 Dijet data at CMS | GARBERS, CH. |
| Modelling of the W+jets background in the 1-lepton channel for the VH, H$\to b\bar{b}$ analysis | GARGIULO, S. |
| The Mu3e Experiment | GERRITZEN, L. |
| Electromagnetic Shower Simulation with Adversarial Neural Nets in ATLAS | GHOSH, A. |
| Search for anomalous electroweak production of WW/WZ in association with a high-mass dijet system in pp collisions at 8 TeV with the ATLAS detector | GIANNINI, A. |
| Measurement of the $H \to WW* \to e\nu\mu\nu$ process with the ATLAS experiment | GUGEL, R. |
| Preparing the search for Vh resonances for the full LHC Run-2 dataset | HÖNLE, A. |
| Estimating the top mass | IGNAZZI, R. |
| Higgs tagging validation in g$\to$bb evens with the ATLAS detector | JACOBS, R. |
| Measurement of the W boson's helicity fractions using $t\bar{t}$ events in the dileptonic final states at 13 TeV | KAVEH, H. |
| Luminosity measurement with the CMS experiment | KNOLLE, J. |
| My transition to High Energy Physics | KUCHINSKAYA, O. |
| High energy and intensity proton beam on ATLAS IBL pixel modules (3D and planar) | LAPERTOSA, A. |
| Search for heavy diboson resonances in the llqq final state in pp collisions at 13 TeV with the ATLAS detector at the LHC | LAVORGNA, M. |
| The BM@N Experiment at JINR: Tracking System | LENIVENKO, V. |
| Test of CP invariance in vector-boson-fusion production of the Higgs boson in $H \to \tau_{\text{lep}}$ decays with the ATLAS detector | LÖSLE, A. |

| Poster title | Presenter |
|---|---|
| Search for type-III Seesaw heavy leptons in pp collisions at 13 TeV with the ATLAS detector | NOVAK, T. |
| The effect of the seam on the collection of primary ionization | PAVLOV, A. |
| W charge asymmetry measurement and the EMC effect study from the heavy ion collisions at the ATLAS detector | PHAM, T.L.H. |
| Search for squarks and gluinos in final states with jets and missing transverse momentum at 13 TeV with the ATLAS detector using the recursive jigsaw technique | QURESHI, A. |
| Towards the XENONnT Dark Matter Experiment | RAMIREZ GARCIA, D. |
| Beam energy measurement of VEPP2000 collider by Compton back-scattering method | ROMAN, L. |
| Serial Powering for the CMS Pixel Detector | RUINI, D. |
| Search for 2HDM neutral Higgs bosons through the process $H \to ZA \to l^+l^- b\bar{b}$ with the CMS detector | SAGGIO, A. |
| Search for heavy resonances decaying into WW/WZ/ZZ at 13 TeV | SCHÄFER, D. |
| Search for $t\bar{t}H(b\bar{b})$ in the all hadronic channel at 13 TeV with the CMS detector | SCHWEIGER, K. |
| Search for new elementary scalars in opposite-charge dilepton final states with the CMS experiment | SPITZBART, D. |
| Multiplicity dependence of strangeness production in pp collisions at 5.02 TeV with ALICE at the LHC | TROPP, L. |
| ATLAS Hardware-based Tracking for the Trigger for the HL-LHC upgrade | TROVATO, F. |
| Z(ll)+MET Analysis: ZZ background estimate | TSAI, F-Y. |
| Search for Supersymmetry in soft dilepton+MET final states | VOURLIOTIS, E. |
| Search for Higgs-boson production in association with a $t\bar{t}$ pair in the H $\to b\bar{b}$ channel | WASSMER, M. |
| Search for Top-squark pair production at ATLAS | WOLLRATH, J. |
| Search for heavy resonances decaying to a Z/W/H boson and a photon with ATLAS at 13 TeV | WONG, V. |